

**An investigation into the replicon of a  
broad host range mobilizable plasmid  
from the moderately thermophilic  
bacterium *Acidithiobacillus caldus***

The crest of the University of Stellenbosch, featuring a shield with a red and white design, topped with a crown and a banner. The Latin motto 'Festera tubercant cultus recti' is inscribed on a scroll at the base.  
**By**  
**Murray Newell Gardner**

**Dissertation presented for the Degree of Doctor of  
Philosophy at the University of Stellenbosch.**

**Supervisor: Prof DE Rawlings**

**December 2003**

## **Declaration**

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Date: 31 July 2003



## Abstract

The moderately thermophilic (45 to 50°C), highly acidophilic (pH 1.5 to 2.5), chemolithoautotrophic *Acidithiobacillus caldus* strain “f” was isolated from a biooxidation process used to treat nickel ore concentrates. Trans-Alternating Field Electrophoresis (TAFE) analysis of total DNA from the *At. caldus* cells revealed two plasmids of approximately 14 and 45-kb. The 14-kb plasmid, designated pTC-F14, was cloned and shown by replacement of the cloning vector with a kanamycin resistance gene to be capable of autonomous replication in *Escherichia coli*. Autonomous replication was also demonstrated in *Pseudomonas putida* and *Agrobacterium tumefaciens* LBA 4404 which suggested that pTC-F14 was a broad host-range plasmid. Sequence analysis of the pTC-F14 replicon region revealed five open-reading frames, and a replicon organization like that of the broad host-range IncQ plasmids. Three of the open-reading frames encoded replication proteins with amino acid sequence identities similar to that of the IncQ-like plasmid pTF-FC2 (RepA, 81%; RepB, 78%; RepC, 74%). This high level of relatedness suggested that the two replicons had evolved from a common ancestor. Since closely related replicons are usually incompatible, the compatible replicons of pTC-F14 and pTF-FC2 raised the question of how the replicons of the two sister plasmids had evolved such that they can now co-exist in the same host cell line. Further incompatibility testing with the IncQ-like plasmid pIE1108 and the IncQ prototype plasmid RSF1010/R1162/R300B determined that pTC-F14 was compatible with pIE1108, but incompatible with the IncQ prototype plasmid. It was found that the RepB and RepA replication proteins of pTF-FC2 and pIE1108 were able to complement the pTC-F14 orthologs only if pTC-F14 RepC was present in *trans*. The RepC protein of pTC-F14 was thus plasmid-template specific, while the RepA and RepB proteins were less plasmid-template specific. A five nucleotide possible iteron-discriminating region in the direct repeats of IncQ-like plasmid *oriV* regions has been identified (Tietze, E. (1998) *Plasmid* 39: 165-181). The iteron sequence of pTC-F14 differs from pTF-FC2 and pIE1108 by three nucleotides in this iteron-discriminating region. It was therefore proposed that co-evolution of the iterons and the RepC protein to a point where the RepC protein no longer recognizes the iteron sequence of a closely related sister plasmid is the mechanism by which replicons evolve to become compatible in



the same host cell. The incompatibility determinant of the IncQ prototype plasmid RSF1010/R1162/R300B was also sought, and subsequently localized to the region encoding the IncQ prototype plasmid's *repAC* genes. Interference with the initiation of pTC-F14 replication by the IncQ prototype plasmid was demonstrated by growth inhibition of a replication-deficient M13 bacteriophage into which *oriV*<sub>pTC-F14</sub> had been cloned. Secondly, the IncQ prototype derivative pKE462 displaced a ColE1-*oriV*<sub>pTC-F14</sub> construct in complementation assays, and a construct containing only the pTC-F14 *repBAC* genes similarly displaced the pKE462 plasmid. As the *oriV*<sub>RSF1010</sub> region was not incompatible with a pTC-F14 replicon, this suggested that it was not the *oriV* region which was expressing incompatibility, but the products of the IncQ prototype plasmid *repAC* genes. It is proposed that incompatibility between pTC-F14 and the IncQ prototype plasmid was the consequence of the *repAC* gene products binding to the iterons of the related replicon, and that these products are unable to initiate replication. The compatible phenotypes expressed by members of the IncQ plasmid family indicates the inadequacy of using plasmid incompatibility as a classification system. Alignment of the amino acid sequences of the three replication protein orthologs clearly showed that the IncQ plasmid family was divided into two groups. To account for replication protein relatedness and the incompatibility phenotype expressed, it is now proposed that that members of the IncQ family be classified into subdivisions that reflect the different IncQ-like replicons identified in this study. Investigation of pTC-F14 replicon regulation identified a putative promoter sequence which is believed to regulate the initiation of a 5.1-5.7-kb polycistronic transcript that encodes all the replication proteins of the pTC-F14 replicon and the MobB and MobA proteins of the IncP-type mobilization module. The large polycistronic transcript appears to be regulated by the RepB protein of the pTC-F14 replicon, and is not subject to cross-regulation by related IncQ plasmids. This suggested that the RepB primase function was not plasmid specific, but that its regulatory function was replicon specific. A second putative promoter sequence identified upstream of the pTC-F14 *pasAB* operon was, however, cross-regulated by the closely related pTF-FC2 plasmid. The pTC-F14 *pas* operon encodes two proteins with high amino acid sequence identity (PasA, 81 %; PasB, 72 %) to the plasmid addiction system of pTF-FC2. This is the second time a plasmid addiction system of this type has been found on an IncQ-like plasmid.



## Opsomming

Die matig termofiliese (45 to 50°C), hoogs asidofiliese (pH 1.5 to 2.5), chemolito-outotrofiese *Acidithiobacillus caldus* ras "f" is geïsoleer vanaf 'n biooksiderende proses wat gebruik word om gekonsentreerde nikkel-erts te behandel. Trans-Afwisselende Veld Elektroforese (TAVE) analise van totale DNA vanaf die *At. caldus* selle, het twee plasmiede van ongeveer 14 en 45-kb. onthul. Die 14-kb plasmied, genaamd pTC-F14, is gekloneer en deur vervanging van die kloneringsvektor met 'n kanamisien weerstandsgeen is daar gewys dat hierdie plasmied in staat is tot outonome replikasie in *Escherichia coli*. Outonome replikasie is ook gedemonstreer in *Pseudomonas putida* en *Agrobacterium tumefaciens* LBA 4404 wat suggereer dat pTC-F14 'n wye gasheer-reeks plasmied is. Volgorde analise van die pTC-F14 replikon area het vyf oop leesrame onthul, en 'n replikon organisasie soortgelyk aan dié van die wye gasheer-reeks IncQ plasmiede. Drie van die oop leesrame kodeer vir replikasie proteïene met aminosuur volgordes ooreenstemmend met dié van die IncQ-tipe plasmied pTF-FC2 (RepA, 81%; RepB, 78%; RepC, 74%). Hierdie hoë vlak van verwantskap stel voor dat die twee replikons vanaf 'n gemeenskaplike voorouer ontwikkel het. Aangesien naby-verwante replikons gewoonlik onverenigbaar is, het die verenigbaarheid van die replikons van pTC-F14 en pTF-FC2 die vraag laat ontstaan van hoe die replikons van twee susterplasmiede ontwikkel het, sodat hulle nou gelyktydig in dieselfde gasheer sellyn kan voortbestaan. Verdere onverenigbaarheid toetsing van die IncQ-tipe plasmied pIE1108 en die IncQ prototipe plasmied RSF1010/R1162/R300B, het bepaal dat pTC-F14 verenigbaar is met pIE1108, maar onverenigbaar met die IncQ prototipe plasmied. Daar is gevind dat die RepB en RepA replikasie proteïene van pTF-FC2 en pIE1108 in staat was om die pTC-F14 ortoloë te komplementeer, slegs as pTC-F14 RepC in *trans* teenwoordig was. Die RepC proteïen van pTC-F14 is dus plasmied-templaatspesifiek, terwyl die RepA en RepB proteïene minder plasmied-templaatspesifiek is. 'n Moontlike iteron-onderskeidende vyf-nukleotied area in die direkte herhalings van die IncQ-tipe plasmied *oriV* areas, is geïdentifiseer (Tietze, E. (1998) *Plasmid* 39: 165-181). Die iteron volgorde van pTC-F14 verskil van pTF-FC2 en pIE1108 met drie nukleotiedes in hierdie iteron-onderskeidende area. Om hierdie rede is daar voorgestel dat ko-evolusie van iterons en die RepC proteïen, tot by 'n



punt waar die RepC proteïen nie meer die iteron volgorde van 'n naby-verwante susterplasmied herken nie, die meganisme is waardeur replikons ontwikkel om verenigbaar te word in dieselfde gasheersel. Die onverenigbaarheidsbepaler van die IncQ prototipe plasmied RSF1010/R1162/R300B is ook ondersoek en gelokaliseer tot die area wat kodeer vir die IncQ prototipe plasmied se *repAC* gene. Inmenging met die inisiasie van pTC-F14 replikasie deur die IncQ prototipe plasmied is gedemonstreer deur groei vertraging van 'n replikasie-gebreklike M13 bakteriofaag waarin die *oriV<sub>pTC-F14</sub>* gekloneer is. Tweedens is die ColE1-*oriV<sub>pTC-F14</sub>* konstruk vervang deur die IncQ prototipe-afgeleide pKE462 in komplementasie proewe, en is die pKE462 plasmied op soortgelyke wyse vervang deur 'n konstruk wat slegs die pTC-F14 *repBAC* gene bevat. Aangesien die *oriV<sub>RSF1010</sub>* area nie verenigbaar was met 'n pTC-F14 replikon nie, stel dit voor dat dit nie die *oriV* area is wat onverenigbaarheid uitdruk nie, maar die produkte van die IncQ prototipe plasmied se *repAC* gene. Dit is voorgestel dat onverenigbaarheid tussen pTC-F14 en die IncQ prototipe plasmied die gevolg is van die *repAC* geenprodukte wat bind aan die iterons van die verwante replikon en dat hierdie produkte nie in staat is om replikasie te inisieer nie. Die verenigbare fenotipes wat deur die lede van die IncQ plasmied familie uitgedruk word, dui aan op die ontoereikendheid van die gebruik van plasmied onverenigbaarheid as 'n klassifikasie sisteem. Vergelyking van die aminosuur volgordes van die drie replikasie proteïen ortoloë wys duidelik daarop dat die IncQ plasmied familie in twee groepe verdeel is. Om verantwoording te doen vir die replikasie proteïen verwantskap en die onverenigbare fenotipe wat uitgedruk is, word daar nou voorgestel dat die lede van die IncQ familie geklassifiseer word in subafdelings wat die verskillende IncQ-tipe replikons geïdentifiseer in hierdie studie, reflekteer. Ondersoek na die pTC-F14 replikon regulering het 'n moontlike promotor volgorde geïdentifiseer. Daar word gemeen dat hierdie promotor die inisiasie van 'n 5.1-5.7-kb polisistroniese transkrip reguleer, wat kodeer vir al die replikasie proteïene van die pTC-F14 replikon en die MobB en MobA proteïene van die IncP-tipe mobilisasie module. Die groot polisistroniese transkrip blyk om gereguleer te word deur die RepB proteïen van die pTC-F14 replikon, en word nie gekruis-reguleer deur die IncQ plasmiede nie. Dit stel voor dat die RepB primase se funksie nie plasmied-spesifiek is nie, maar dat die reguleerbare funksie replikon-spesifiek is. 'n Tweede moontlike promotor volgorde wat stroom-op van die pTC-F14 *pasAB* operon geïdentifiseer is, is egter gekruis-reguleer deur die pTF-FC2 plasmied. Die pTC-F14

*pas* operon kodeer vir twee proteïene met hoë aminosuur volgorde verwantskappe (PasA, 81 %; PasB, 72 %) aan die plasmied-verslaafde sisteem van pTF-FC2. Dit is die tweede keer dat hierdie tipe plasmied-verslaafde sisteem in 'n IncQ-tipe plasmied gevind is.



## Acknowledgements

My most sincere gratitude goes to my supervisor, Professor D.E. Rawlings, for his constant support and enthusiastic guidance throughout this study. His enthusiasm for his subject and willingness to give freely of his time served as a source of inspiration.

I am deeply indebted to Dr Shelly Deane for her invaluable assistance in the laboratory, and for our numerous discussions as I tried to put the puzzle together. Special thanks are due to my colleagues in the Biomining Research Unit, both past and present, who have provided the entertainment and assistance that has made my time so much easier. Of special mention is Bronwyn Butcher whose encouragement and friendship I have always appreciated. I would also like to thank all the members of the Department of Microbiology for assistance in their various capacities, but mostly the 'tea club' who provided a much lighter side to each day.

I would like to thank the National Research Foundation, BHP-Billiton, and the University of Stellenbosch for financial support.

I have to thank Graham Glover, Mark Robertson, Grant Surtees and Lawrence Schäfer, who not only brought Colin to life, but have been a constant source of encouragement since the days of College House. I could not have asked for a finer group of friends.

My deepest gratitude and love goes to my mother and my brother Sean, who through their unwavering confidence and telephonic support gave me the determination to achieve my ambition. I also want to thank my wife's parents who have been so understanding of my needs, and who provided us with a home in our hour of need.

Finally, to my wife Helena, who began this with me, I cannot express how much your support, understanding, patience and love has meant to me over the years.



## Table of Contents

<b>Abbreviations</b>	x
<b>Chapter One:</b> Literature review	1
<b>Chapter Two:</b> Replicon analysis of the IncQ-like plasmid pTC-F14 isolated from <i>Acidithiobacillus caldus</i> strain “F”	71
<b>Chapter Three:</b> Incompatibility and complementation of the pTC-F14 replicon	101
<b>Chapter Four:</b> Regulation of the pTC-F14 replicon, and cross- regulation by IncQ and IncQ-like plasmids	124
<b>Chapter Five:</b> General Discussion	147
<b>Appendix A:</b> Growth media, additives and solutions	159
<b>Appendix B:</b> Strains, plasmids and bacteriophage used in this study	161
<b>Appendix C:</b> Primers used in this study	164
<b>Appendix D:</b> Plasmid maps	165
<b>Appendix E:</b> Annotation of pTC-F14 replicon DNA sequence	169
<b>Literature cited</b>	175

## Abbreviations

A	adenosine
A	alanine
Å	Ångström units
A+T-rich	adenosine and thymidine rich sequence
Ala	alanine
ADP	adenosine 5'-diphosphate
Amp	ampicillin
AMP	adenosine 5'-phosphate
Arg	arginine
Asp	aspartic acid
ATP	adenosine-5'-triphosphate
bp	base pairs
°C	degrees Celsius
C	cysteine
C	cytosine
C-terminal	carboxyl-terminus
cDNA	complementary DNA
CSPD®	disodium 3-(4-methoxy Spiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.1 <sup>3,7</sup> ]decan}-4-yl) phenyl phosphate
CTP	cytidine-5'-triphosphate
ctRNA	countertranscribed RNA
D	aspartic acid
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytidine-5'-triphosphate
DEPC	diethyl pyrocarbonate
dGTP	deoxyguanosine-5'-triphosphate
DIG	digoxigenin-11-dUTP
DNA	deoxyribonucleic acid
DNA Pol I	DNA polymerase I
DNA Pol II	DNA polymerase II
DNA Pol III	DNA polymerase III
DNA Pol III HE	DNA polymerase III holoenzyme
DNA Pol IV	DNA polymerase IV
DNA Pol V	DNA polymerase V
dNMP	deoxyribonucleotide monophosphate
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded deoxyribonucleic acid
<i>dso</i>	double strand origin
dUTP	deoxyuridine triphosphate
E	glutamic acid
EDTA	ethylenediaminetetraacetic acid
F	phenylalanine
FIS	factor for inversion stimulation

g	gram
G	guanine
G	glycine
G+C-rich	guanosine and cytidine rich sequence
Gln	Glutamine
Glu	glutamic acid
GTP	guanosine-5'-triphosphate
H	histidine
HE	holoenzyme
His	histidine
HTH	helix-turn-helix
I	isoleucine
IHF	integration host factor
Ile	isoleucine
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
K	lysine
kb	kilobase pairs or 1000-bp
kDa	kilodaltons
kJ	kilojoules
kV	kilovolts
L	leucine
l	liter
LIC	leading-strand initiation and control
Lys	lysine
LZ	leucine zipper
M	methionine
M	molar
mA	milliampere
mb	megabases
Met	methionine
mg	milligram
ml	milliliters
mM	millimolar
MOPS	3-[N-Morpholino]propanesulfonic acid
mRNA	messenger RNA
N	asparagines
N	normal
N-terminal	amino-terminus
NAD <sup>+</sup>	nicotinamide-adenine dinucleotide
NADH	reduced form of NAD
NCBI	National Center for Biotechnology Information
NMN	nicotinamide mononucleotide
NTP	nucleoside-5'-triphosphate



nm	nanometers
OH-terminus	hydroxyl-terminus
ONPG	<i>o</i> -nitrophenyl- $\beta$ -D-galactoside
ORF	open reading frame
<i>oriC</i>	chromosomal origin of replication
<i>oriT</i>	origin of transfer
<i>oriV</i>	origin of vegetative replication
p	phosphate
P	proline
PAS	primosome assembly site
PCR	polymerase chain reaction
p.f.u	plaque forming units
Phe	phenylalanine
P <sub>i</sub>	inorganic phosphate
PP <sub>i</sub>	pyrophosphate
pRNA	RNA oligonucleotide primer
Pu	purine
Py	pyrimidine
Q	glutamine
R	arginine
RBS	ribosome binding site (Shine-Dalgarno)
Rep	replication proteins
RNA	ribonucleic acid
rRNA	ribosomal RNA
RNAP	RNA polymerase
RT-PCR	reverse-transcriptase polymerase chain reaction
S	sedimentation coefficient
S	serine
s	second(s)
SDS	sodium dodecyl sulfate
Sm	streptomycin
SSB	single-stranded DNA binding protein
ssDNA	single-stranded deoxyribonucleic acid
<i>ssi</i>	single-stranded initiation site
<i>sso</i>	single strand origin
Su	sulfonamide
T	threonine
T	thymine
TE	Tris EDTA buffer
Tris	Tris (hydroxymethyl) aminomethane
Trp	tryptophan
Tyr	tyrosine
$\mu$ F	micofarads

$\mu\text{g}$	microgram
$\mu\text{l}$	microliters
$\mu\text{m}$	micrometer
V	valine
v/v	volume/volume
Val	valine
W	tryptophan
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -galactoside
Y	tyrosine
ZF	zinc finger

# Chapter One

## Literature review

### Contents

1.1)	Introduction	2
1.2)	Prokaryote chromosome replication	3
	1.2.1) DnaA-mediated initiation of replication	4
	1.2.2) DnaB helicase	8
	1.2.3) DnaG primase	14
	1.2.4) DNA Polymerase III holoenzyme	18
	1.2.5) DNA Polymerase I and DNA ligase	22
	1.2.6) Tus-mediated replication termination	25
	1.2.7) Topoisomerases	27
1.3)	Modes of plasmid replication	29
	1.3.1) Rolling circle replication	31
	1.3.1.1) General replication mechanism	32
	1.3.1.2) Origin comparisons	33
	1.3.1.3) Replication initiator protein comparisons	34
	1.3.1.4) Replication and control mechanism comparisons	35
	1.3.2) Theta replication	39
	1.3.2.1) General replication mechanism	39
	1.3.2.2) Origin comparisons	41
	1.3.2.3) Replication initiator protein comparisons	45
	1.3.2.4) Replication mechanism comparisons	47
	1.3.2.5) Copy number control	54
1.4)	Plasmid host range constraints	62
	1.4.1) Dependence on host replication proteins	62
	1.4.2) Origin structure and topology	63
	1.4.3) Rep protein interactions	65
	1.4.4) Efficient plasmid-host communication	66
1.5)	Conclusion	66
1.6)	Aims of this study	69



## 1.1) Introduction

The biomining environment is a strongly selective ecological niche in which the acidic (pH 1.5-4.0), low nutrient, inorganic conditions favour bacteria such as the acidophilic, obligately autotrophic chemolithotrophs that acquire their energy by oxidation of reduced iron and/or sulfur compounds. The biomining environment is thus markedly different from the heterotrophic environment inhabited by microbes such as the enteric bacteria and Pseudomonads.

The contribution plasmids make to the dissemination and maintenance of information in the biomining environment has not been as extensively investigated as plasmid-mediated genetic exchange in the ecological niche occupied by neutral heterotrophic bacteria. However, genetic exchange in the biomining environment almost certainly takes place, and by means of plasmids (Rawlings and Tietze, 2001).

The focus on plasmid-mediated genetic exchange in the heterotrophic environment has led to the characterization of a diverse range of plasmid replicons from this ecological niche. The ecological isolation of the bacterial community in the biomining environment poses questions as to whether plasmid replicons found in this niche are unique, or whether the replication strategies employed by these plasmid replicons show similarities with those characterized from heterotrophic bacteria.

Therefore, it is of interest to investigate plasmids isolated from the biomining environment so as to determine what plasmid replication mechanisms are successful in this ecological niche, and what contribution these replication mechanisms may make to the dissemination and maintenance of information in this environment.

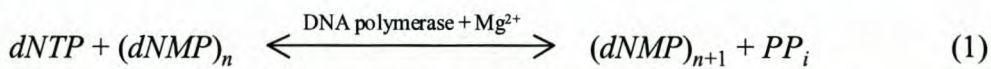
The two general replication strategies used by plasmids to load and assemble the replisome multiprotein complex at the origin of replication are: i) rolling circle replication, or ii) theta replication; of which the single-stranded displacement replication mechanism practiced by Incompatibility group Q (IncQ) plasmids is a variation of theta replication (Espinosa *et al.*, 2000). Plasmids which utilize either of these replication mechanisms all have a varying dependence on host replication factors.



Therefore, a general reference needs to be made to prokaryote chromosome replication so as to provide an understanding for the basis of replication mechanisms employed by plasmids, and how this dependence may restrict the host range of a plasmid. As specific reference will be made to the chromosome replication mechanism of *Escherichia coli*, protein designations are for this species, and homologs in other species are inferred.

## 1.2) Prokaryote chromosome replication

Arthur Kornberg's search for an enzyme which incorporated nucleotides into a growing DNA chain (Eqn 1) led to his discovery of DNA polymerase (deoxynucleotidyltransferase); later to be shown to be one of five different DNA polymerases in prokaryote cells, and now called DNA Pol I (Table 1.1).



Reaction equation for DNA chain elongation where dNMP is any deoxynucleoside monophosphate and dNTP is any triphosphate. Reproduced from Mathews and van Holde (1990).

**Table 1.1:** Biochemically documented template-dependent DNA polymerases found in *E. coli*. Adapted from Mathews and van Holde (1990); Sutton and Walker (2001).

Name	Alternate designations	Structural gene	Function(s)
Pol I		<i>polA</i>	Maturation of Okazaki fragments; nucleotide excision repair
Pol II	DinA	<i>polB</i>	Induced replisome reactivation; translesion DNA synthesis
Pol III		<i>polC</i>	DNA replication; nucleotide excision repair
Pol IV	DinB	<i>umuC</i>	Adaptive mutagenesis; translesion DNA synthesis
Pol V	UmuC or UmuD' <sub>2</sub> C	<i>umuC</i>	translesion DNA synthesis ("SOS mutagenesis")

The state of DNA dictates which DNA polymerase enzyme(s) will participate in a particular polymerase reaction (Eqn 1), as DNA polymerases are template dependent. There are a variety of templates available for DNA polymerases to act upon. Whereas single-stranded DNA (ssDNA) molecules provide a template for DNA replication through the concerted action of DNA Pol I and DNA Pol III, nicks or single-stranded gaps in double-stranded DNA (dsDNA) molecules provide substrates for DNA Pol II, DNA Pol IV, and DNA Pol V. DNA Pol II, DNA Pol IV, and DNA Pol V are active



in processes such as base excision repair, translesion DNA synthesis, and DNA recombination (Sutton and Walker, 2001).

There is a strong suggestion that an important level of control rests with other proteins associated with a particular DNA template substrate. This protein-DNA template association assists in determining which DNA polymerase is necessary to catalyze the polymerase reaction, as DNA polymerases cannot distinguish between a 3'-hydroxyl (OH) terminus that is part of a replication fork, or that which is the result of nicked DNA (Sutton and Walker, 2001). However, the defining mechanism of a DNA polymerase in all the above processes is that of deoxyribonucleotide (dNTP) addition to a 3'-OH terminus of a primer resulting in a growing polynucleotide chain in a 5'→3' direction. Watson-Crick base-pairing interactions between the incoming dNTP and template DNA selects for the correct complementary dNTP to be added, while the 3'-OH group of the primer terminus covalently binds the added nucleotide to the growing chain. This covalent bond is achieved by nucleophilic attack on the inner ( $\alpha$ ) phosphate of the dNTP substrate by the 3'-OH group of the primer terminus. The need for a 3' primer terminus and a ssDNA template limits the kinds of reactions that a DNA polymerase can catalyze (Mathews and van Holde, 1990).

Prokaryote chromosomal replication is somewhat less complicated than in eukaryotes as replication is initiated at a single origin of replication (*oriC*), and under normal replication conditions the two DNA polymerases DNA Pol III and DNA Pol I replicate the DNA molecule from the *oriC* to the terminus in a bidirectional manner.

### **1.2.1) DnaA-mediated initiation of replication**

A new round of chromosomal DNA replication in *E. coli* occurs when approximately 15-25 DnaA proteins bind to the *oriC* resulting in a DnaA-dependent initiation of replication (Messer *et al.*, 2001; Zheng *et al.*, 2001).

Replication initiation appears to be regulated by the binding of adenosine nucleotides ATP and ADP to DnaA, where the ATP-DnaA protein is active for initiation replication, while ADP-DnaA is not. Hydrolysis of ATP, resulting in ADP-DnaA, is



enhanced by the factor *IdaB* and the  $\beta$ -subunit sliding clamp component of DNA Pol III holoenzyme (DNA Pol III HE) when assembled on DNA (Katayama *et al.*, 1998; Zheng *et al.*, 2001). This action, referred to as regulatory inactivation of DnaA (RIDA), can be reversed in the presence of *oriC* DNA and acidic phospholipids in a fluid bilayer (Zheng *et al.*, 2001).

DnaA consists of four domains: domain I is involved in oligomerization, II acts as a flexible linker, III binds ATP and phospholipids, and IV is involved in binding to specific sequences in the *oriC* referred to as DnaA-boxes (Sharma *et al.*, 2001). The binding of DnaA domain III to acidic phospholipids is thought to occur by means of an amphipathic helix structure that acts as a membrane surface-seeking domain of peripherally associated membrane proteins (Skarstad and Boye, 1994).

As the inner membrane of *E. coli* is composed of zwitterionic phosphatidylethanolamine (~70%), and the anionic phospholipids phosphatidylglycerol (~25%) and cardiolipin (~4%) (Zheng *et al.*, 2001), this suggests that there is a direct link between lipid composition of the cellular membrane and the ability of DnaA to function in the initiation of replication. Immunogold and immunofluorescent analysis has confirmed cytoplasmic membrane-associated DnaA (Newman and Crooke, 2000).

A requirement for the initiation of replication is that the *oriC* of the *E. coli* chromosome is on a supercoiled template (Messer *et al.*, 2001). The consequence of DNA condensation, DNA supercoiling, and the interplay of chromosomal proteins with DNA topological features results in the organization of the bacterial chromosome into a higher-order structure referred to as the nucleoid (Bahloul *et al.*, 2001). A number of small, yet abundant, DNA-binding proteins are believed to play an important role in the structuring of the bacterial nucleoid. These histone-like proteins include the site-specific DNA-binding proteins IHF (integration host factor) and FIS (Factor for Inversion Stimulation), and the general DNA-binding proteins HU and H-NS. These nucleoid proteins not only define the architecture of the bacterial chromosome, but also play an important role in the initiation of replication from the chromosomal *oriC* (Atlung and Hansen, 2002).



The *E. coli oriC* contains five DnaA boxes, designated R1-R5(M), the additional DnaA binding sites I1-I3, and an A+T-rich region (Fig 1.1).

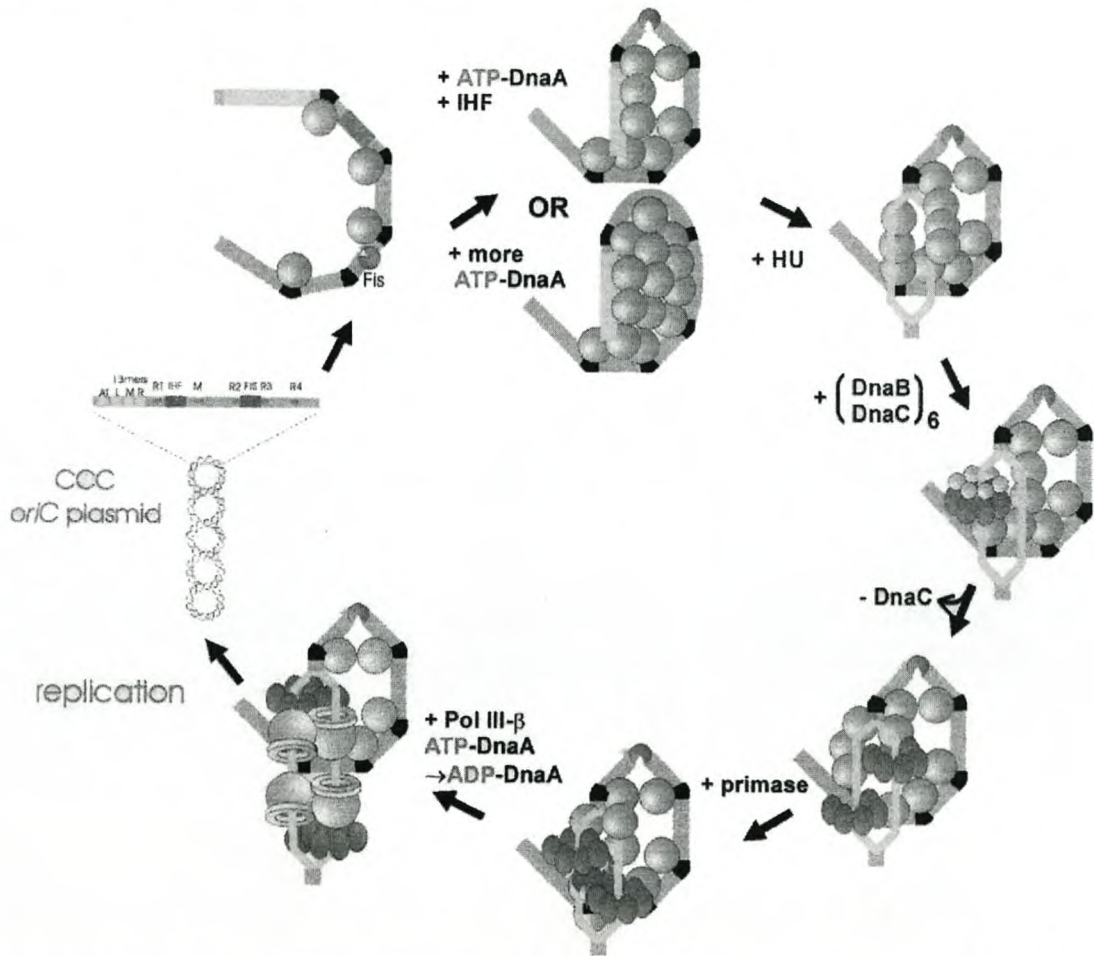


**Figure 1.1:** Structural organization of the *E. coli oriC*. ‘AT’ designates the A+T-rich region with the three 13-mer repeats alongside. The R1-R5(M) DnaA boxes are shaded black, while the additional DnaA binding sites I1-I3 are shaded grey. The specific binding sites for the nucleoproteins IHF and FIS are shown in their relative positions. Adapted from Ryan *et al.* (2002).

The A+T-rich region of *oriC* is organized as the AT-cluster, and three A+T-rich 13-mer repeats. Each of the repeats have the Dam methyltransferase GATC recognition sequence (Messer *et al.*, 2001). The R1-R5(M) DnaA boxes consist of an asymmetric 9-mer consensus sequence (5'-TT<sup>A</sup>/<sub>T</sub>TNCACA-3'), with high affinity boxes R1, R2 and R4 binding the DnaA monomer throughout the cell cycle, while low affinity boxes I1, I2, I3, R3 and R5(M) require binding cooperativity with the nucleoproteins (Schaper and Messer, 1995; Atlung and Hansen, 2002; Ryan *et al.*, 2002).

The assembly of the ‘initiation complex’ occurs in a synchronous manner (Fig 1.2). With the DnaA high affinity boxes containing the DnaA monomer, the FIS protein binds to its specific site between DnaA-boxes R2 and R3. The binding of FIS inhibits the next stage of the initiation phase until, presumably, the correct complement of ATP-DnaA accumulates, whereupon FIS leaves the complex (Grimwade *et al.*, 2000). IHF then binds to its specific binding site between DnaA-box R1 and R5(M), introducing a strong bend. Binding of IHF is transient, and it appears that IHF redistributes pre-bound DnaA from R1 and R4 to I1, I2, I3 and R5(M) (Grimwade *et al.*, 2000). It is not clear whether IHF induces the removal of DnaA from R1 and R4 so that DnaA is available to the low affinity boxes, or if IHF promotes sharing of DnaA between R1 and R4, and the I1-I3 sites (Grimwade *et al.*, 2000). This binding of IHF could be the molecular switch that controls initiation of replication, since when IHF is not present in the complex, replication initiation does not occur (Ryan *et al.*, 2002). HU protein then binds to I3, but its exact function is unclear. HU protein may mediate local *oriC* negative supercoiling (Bahloul *et al.*, 2001). Ryan *et al.* (2002) speculates that if HU mediates supercoiling then the threshold level of DnaA necessary to produce unwinding would be a function of the amount of HU present in

the complex as well as the presence of IHF. The binding of IHF, together with HU protein, results in the formation of the ‘open complex’ by causing the partial unwinding of the A+T-rich region.



**Figure 1.2:** Proposed model for the replication initiation cycle of the *E. coli* chromosome. Reproduced from Messer *et al.* (2001).

Both ATP-DnaA and ADP-DnaA bind with the same affinity to the 9-mer consensus sequences, but ATP-DnaA exhibits an additional affinity to a 6-mer consensus sequence (5'-AGatct-3'), recently named the ATP-DnaA box. The A+T-rich region of *E. coli oriC* contains 6 ATP-DnaA boxes. Binding of ATP-DnaA to these boxes in the A+T-rich region follows only after ATP-DnaA has bound to DnaA-box R1 (Messer *et al.*, 2001). This results in further destabilization of the helix and unwinding of the A+T-rich region. The unwound region spans 28-base pairs (bp) without, and 44- to 46-bp with SSB (single-stranded DNA binding) protein bound (Messer *et al.*, 2001).



SSB protein is a tetramer that acts as a helix destabilizing protein. It preferentially and cooperatively binds ssDNA preventing dsDNA from re-forming prematurely. Apart from its helix destabilizing ability, SSB protein appears to be essential for primosome and replisome assembly, yet has no defined enzymatic activity (Kowalczykowski *et al.*, 1994).

Once the open complex is formed, interaction of DnaA with the DnaBC helicase complex results in the formation of the pre-priming complex required for the assembly of the primosome.

### 1.2.2) DnaB helicase

The DnaC helicase-loading protein is required to load the replicative helicase DnaB into the 44-bp exposed region of the open complex, expanding it to about 65-bp (Messer *et al.*, 2001). The DnaC protein forms a DnaC<sub>6</sub>:DnaB<sub>6</sub> complex with DnaB, and in doing so the DnaC is activated to interact with ssDNA, thereby recruiting the DnaB helicase to *oriC* (Learn *et al.*, 1997).

The DnaBC helicase complex now interacts with the *oriC*:DnaA nucleoprotein structure to form the second-stage pre-priming complex. The interaction requires ATP and specific protein-protein interactions between either DnaA and DnaC, or DnaA and DnaB, or both (Learn *et al.*, 1997). DnaA facilitates the loading of DnaB onto ssDNA, since SSB protein bound ssDNA is not a substrate for the helicase. DnaA interacts directly with DnaB as the amino (N)-terminal domain I of DnaA interacts with the central  $\beta$  domain of DnaB, and residues from the DnaA central domain III interact with the N-terminus of DnaB (Messer *et al.*, 2001). The binding of DnaB helicase to ssDNA occurs in a sequence-independent manner (Patel and Picha, 2000). It is speculated that it is the loading of DnaB, together with the hydrolysis of ATP, that displaces DnaC from the *oriC*:DnaA:DnaB:DnaC complex (Learn *et al.*, 1997).

The DnaB helicase is an enzyme that controls the formation and translocation of the replisome in *E. coli* (Biswas *et al.*, 2002). It is the interaction of the DnaB helicase



with replication proteins in the replication fork that allows it to form the replisome. DnaB couples the primosome complex to the DNA Pol III HE's on both strands so that the leading and lagging strands are replicated simultaneously (Biswas and Biswas, 1999[a]). Hall and Matson (1999) define a helicase as an enzyme in which conserved motifs identify an engine that powers the unwinding of double-stranded nucleic acids using energy derived from nucleotide hydrolysis. The enzyme also undergoes conformational changes that allow for the transduction of energy between nucleotide and nucleic acid binding sites.

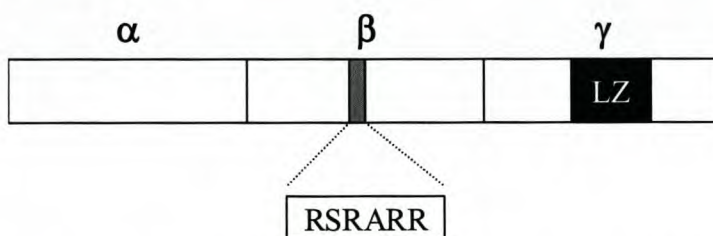
Helicase primary structure comparisons have led to the identification of several conserved sequence patterns. These conserved sequence patterns define a helicase family or superfamily. The two largest superfamilies, SF-1 and SF-2, are characterized by seven conserved regions of sequence homology, whereas SF-3 has only three conserved regions. The SF-4 superfamily, which includes the *E. coli* DnaB, T7 gene 4, and T4 gene 41 helicases, is smaller and defined by five (H1, H1a, H2, H3 and H4) conserved regions (Lohman and Bjornson, 1996; Hall and Matson, 1999). The only regions of sequence similarity shared uniformly among all the helicase superfamilies are the "A" and "B" motifs of the "Walker Box" (Walker *et al.*, 1982). The Walker A and B motifs have been shown to be useful predictors of a nucleoside-5'-triphosphate (NTP) binding site. The H1 motif of the SF-4 helicases contains the Walker A consensus sequence that forms the "P-loop" within the NTP-binding site, while the H2 motif contains the Walker B sequence. Although the Walker A and B motif sequence pattern is being continually revised as more SF-4 helicase sequences become available, there does appear to be a highly conserved and essential lysine in the Walker A motif (H1), and a conserved aspartate in the Walker B motif (H2) that interacts with NTP via  $Mg^{2+}$  (Lohman and Bjornson, 1996; Hall and Matson, 1999; Patel and Picha 2000).

The SF-4 structural class of helicases to which the *E. coli* DnaB helicase belongs, assemble into ring-shaped hexamers comprising six 52 kDa monomers (Biswas *et al.*, 2002). Assembly of the hexameric helicase requires  $Mg^{2+}$ , nucleotide binding, or both. Helicase assembly proceeds through leucine zipper (LZ)-mediated dimerization of monomers, followed by the association of the dimers to form the hexamer. The binding to DNA greatly stabilizes the hexameric structure (Patel and Picha, 2000;



Biswas *et al.*, 2002). Electron microscopic studies are suggestive of a doughnut-shaped rosette structure with an internal diameter of  $\sim 40$  Å (Biswas and Biswas, 1999[b]).

Biswas and Biswas (1999[a]) were able to assign specific functions to three domains they identified on the 471 amino acid *E. coli* DnaB helicase (Fig 1.3). Using a series of deletion constructs these authors showed that domain  $\alpha$  was found to be indispensable for helicase activity, despite having no enzymatic activity, nor ATPase or DNA-binding activities. Domain  $\beta$  was shown to contain the functional ATPase active site, and that addition of domain  $\gamma$  restored the DNA-dependent ATPase activity to wild-type level. Domain  $\gamma$  was also implicated in hexamer formation as the polypeptides of domains  $\alpha$  or  $\beta$  could only form dimers. This suggested that a protein-protein interface required for hexamer formation was located within domain  $\gamma$  (Biswas and Biswas, 1999[a]).



**Fig 1.3:** Schematic diagram of the *E. coli* DnaB protein domains. Domain  $\beta$  contains the functional ATPase, as well as the RSRARR DNA-binding motif (indicated by grey box). The leucine zipper motif in domain  $\gamma$  is indicated by the black box. Reproduced from Biswas and Biswas (1999[a]).

The consensus DNA-binding LZ-motif of classic LZ-proteins is bipartite. One half of the DNA-binding motif, contains the BBXBXB consensus sequence (where B represents Arg or Lys, and X is any amino acid), which is separated by four amino acids from the other half containing the BXBxBB inverted repeat. The LZ dimerization motif has a consensus sequence of LX<sub>6</sub>LX<sub>6</sub>LX<sub>6</sub>LX<sub>6</sub>L, and the residues adopt an  $\alpha$ -helical secondary structure in which the leucines are arranged on the same face creating a hydrophobic spine. The parallel contact of two hydrophobic spines results in dimerization (García de Viedma *et al.*, 1996). The LZ domain of classic LZ proteins has been shown to form a rigid Y-shaped structure that binds the DNA helix through its BBXBXB and BXBxBB peptide domains (Biswas and Biswas, 1999[b]).

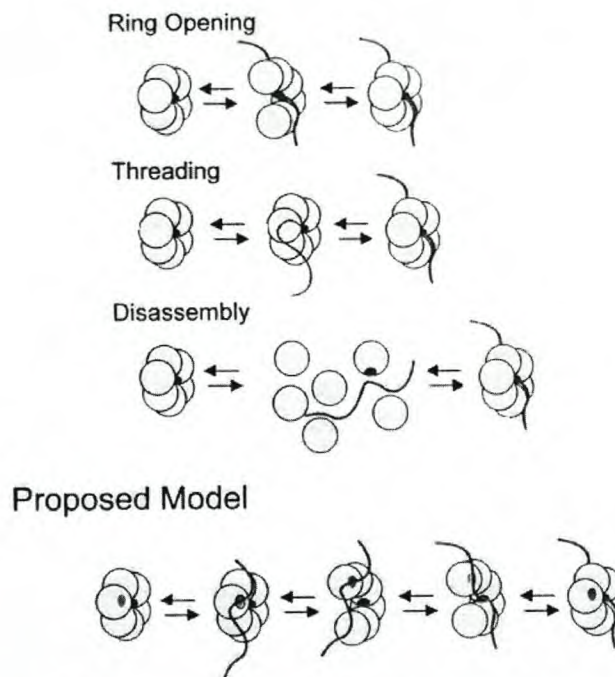


Only one half of the LZ consensus DNA-binding motif, the RSRARR sequence, is present in the  $\beta$  domain of DnaB. It is separated from the LZ dimerization domain by 31 residues, and has the N-terminal Leu of its consensus sequence substituted for Ile. The uncharacteristically large 31 residue spacer region may act as a hinge allowing for flexible movement of the DNA-binding domain; a feature not required by classic LZ proteins (Biswas and Biswas, 1999[b]).

The conclusion that DNA binding by DnaB is mediated through the formation of a dimeric LZ motif was reached from deletion studies showing that removal of only the DnaB LZ-dimerization motif, but not the RSRARR sequence, resulted in complete loss of DNA binding, helicase activity, and hexamer formation (Biswas and Biswas, 1999[b]). Each dimeric unit, formed by the association of two monomers through the LZ-dimerization motif contains only one ssDNA binding site and one active ATP hydrolysis site. Thus, the hexameric DnaB contains only three DNA binding sites orientated at  $120^\circ$  to each other, and three active ATP hydrolysis sites. Hydrolysis of one ATP molecule will result in the rotation about the plane of ssDNA by  $120^\circ$  (Biswas and Biswas, 1999[b]).

The ring-like structure of *E. coli* DnaB is thought to favour processivity by topologically tethering the helicase to the DNA. By encircling the DNA strand the DnaB helicase decreases the probability of complete dissociation from DNA during translocation. However, protein-DNA photo-crosslinking studies show that only one subunit of DnaB efficiently cross-links to ssDNA. This suggests that ssDNA does not interact simultaneously with all six surfaces of the central channel of the helicase hexamer (Patel and Picha, 2000). The mechanism for assembly of the DnaB hexamer around ssDNA is not clearly defined. There are three general models, descriptively termed 'ring-opening', 'threading' and 'disassembly'. Patel and Picha (2000) propose a fourth model incorporating elements of the above three models (Fig 1.4). These authors suggest that ssDNA binds to loading sites on the outer surface of the hexamer. Subsequently DNA binds to the central channel of the hexamer after ring-opening, followed by closure of the ring around the ssDNA.





**Figure 1.4:** Current models which attempt to explain how DNA binds to the central channel of a ring-shaped helicase. Reproduced from Patel and Picha (2000).

The release of DnaC from the open complex is concomitant with ATP hydrolysis, thereby activating the DnaB helicase. It has been shown that in the presence of non-hydrolyzable ATP analogs, DnaB preferentially binds to the 5' strand of unwound DNA. This is consistent with its 5'→3' unwinding activity when powered by ATP hydrolysis (Matson and Kaiser-Rogers, 1990; Benkovic *et al.*, 2001).

However, there is no current model that is able to incorporate ATPase-coupled translocation on ssDNA with that of base-pair separation. There are three, two, and one-site translocation models which attempt to explain helicase translocation on ssDNA. Patel and Picha (2000) suggest that the NTPase-driven movement of helicases occurs by a similar mechanism to that used by F<sub>1</sub>-ATPase. The F<sub>1</sub> complex of ATPase has three active sites which hydrolyze ATP by an alternating site mechanism. This means that only one site can bind ATP, the second can only hydrolyze ATP, and the third can only release the hydrolyzed products. Each active site of the F<sub>1</sub>-ATPase sequentially switches between the different conformations, and the changes at the three sites are linked (Patel and Picha, 2000). Applying these mechanistic coupling rules to DnaB, the three-site sequential translocation model was proposed in which the helicase cycles sequentially through a series of conformational states; driven by the binding and/or hydrolysis of ATP (Lohman and Bjornson, 1996;



Patel and Picha, 2000). This mechanism suggests that three active sites of the hexameric helicase interact with DNA in a sequential manner that leads to unidirectional movement. One conformational state of the active site tightly binds DNA, the second conformational state begins to release DNA, and the third conformational state begins to bind DNA at a different lattice position upon binding ATP. The repetition of these cycles results in processive unidirectional movement (Patel and Picha, 2000).

As the DnaB helicase unwinds duplex DNA at the replication fork, the ssDNA that is formed is bound by SSB protein. Primarily SSB protein is thought to coat ssDNA to remove hairpin-blocks to DNA Pol III HE progression (Yuzhakov *et al.*, 1999). It is, however, thought that SSB protein may also play a significant part in regulation of the helicase ATPase activity. DnaB has a strong DNA-dependent ATPase activity, and a high basal ATPase activity. It has been proposed that SSB protein modulates this activity by inhibiting DnaB ATPase activity, thereby preventing futile and wasteful ATP hydrolysis when the helicase is not acting on a replication fork (Biswas *et al.*, 2002).

During translocation dsDNA is unwound by an as yet undefined process. Three models propose mechanisms by which dsDNA can be unwound. The 'wedge' model suggests that NTPase-coupled movement provides enough force to enable the helicase to destabilize and separate the base-pairs at the junction. The 'torsional' model suggests that one strand of DNA is bound in the central channel while the other is bound to the outer surface of the hexamer. By rotating, the helicase generates torque that destabilizes dsDNA. The 'helicase-destabilizing' model suggests that protein-DNA interactions destabilize dsDNA. This model proposes that protein conformational changes brought about by NTPase activity leads to base-pair separation on the helicase active site (Patel and Picha, 2000).

With two DnaB helicases now bound to each strand (Fig 1.2), helicase activity causes 5'→3' translocation such that each helicase now moves past the other, thereby providing enough template for the DnaG primase to initiate two opposed and overlapping leading strands (Messer *et al.*, 2001). Biswas *et al.* (2002) showed that a



five nucleotide fork length is adequate for maximal stimulation of DnaB helicase activity, and that there was no observable increase in helicase rate when DnaB contacted forks larger than this. The *in vitro* rate at which *E. coli* DnaB has been estimated to unwind dsDNA is 30-60 bp.s<sup>-1</sup> in the presence of DNA gyrase. However, this increases substantially to ~700 bp.s<sup>-1</sup> in the presence of DNA Pol III (Lohman and Bjornson, 1996; Patel and Picha, 2000).

### 1.2.3) DnaG primase

The separation of dsDNA by the DnaB helicase results in two strands with opposite polarity. As DNA Pol III HE moves only in a 5'→3' direction, the unidirectional growth of both DNA strands can only be explained by the semidiscontinuous model of DNA replication. According to this model, the leading strand (3'-terminated daughter chain) is synthesized continually, while the lagging strand (5'-terminated daughter strand) is synthesized in small units (Okazaki fragments) that are subsequently join by DNA ligase to yield a complete dsDNA molecule (Frick and Richardson, 2001).

During DNA replication in *E. coli*, the 60 kDa DnaG primase synthesizes an RNA oligonucleotide primer (pRNA) that provides a 3'-OH terminus from which chain elongation can be initiated by the DNA Pol III HE. The DnaG primase is essential for functional replication priming in *E. coli* and cannot be replaced by RNA polymerase (RNAP) (Versalovic and Lupski, 1993). The DnaG primase synthesizes a small complementary pRNA at the *oriC* for the initiation of leading strand synthesis, and pRNA's on the lagging strand during semidiscontinuous DNA replication. The catalytic abilities of primases and RNAP are functionally identical in that they both synthesize RNA by nucleophilic attack between the 3'-OH group of the terminal ribonucleotide and the  $\alpha$ -phosphate of substrate ribonucleotide triphosphates in a DNA-dependent manner (Versalovic and Lupski, 1993). At the 5'-CTG-3' DnaG target recognition sequence an 8-12 ribonucleotide primer having a pppAG(N)<sub>8-10</sub> sequence is synthesized (Benkovic *et al.*, 2001; Frick and Richardson, 2001). The DnaG primase preferentially synthesizes primers of 11 ( $\pm$ 1) nucleotides in length when part of the replisome, although longer primers of up to 60 nucleotides can be



synthesized, but this appears to be only possible when SSB protein is not present (Yoda and Okazaki, 1991).

The crystal structure of the core region of DnaG revealed a cashew-shaped molecule characterized by a shallow wedge-shaped cleft on the concave side of the protein (Keck *et al.*, 2000). Frick and Richardson (2001) have proposed a mechanism for primer synthesis that involves two NTP binding sites. The site at which the NTP is to be incorporated at the 5' end of the primer is referred to as the initiation site. The second site which binds the NTP added to the 3' end of the primer is called the elongation site. These authors suggest that pRNA synthesis is catalyzed in a minimum of five steps: (1) template binding, (2) NTP binding, (3) initiation, (4) extension to a functional primer, and (5) primer transfer to DNA polymerase.

The DnaG primase is targeted to the replication fork by direct interaction with the replicative helicase DnaB as it moves through the dsDNA molecule creating the replication fork. It is believed that SSB protein assists in directing DnaG to ssDNA, and in the assembly of DnaG in the primosome complex. The ssDNA is thought to thread through the primase in a 3'→5' direction. As DnaG and DnaB bind together to form a physical complex, the movement of the DnaB helicase is believed to transport the primase to initiation sites (Benkovic *et al.*, 2001; Frick and Richardson, 2001). The interaction of DnaG with DnaB is mediated by the carboxy (C)-terminal 16 amino acids of DnaG, and results in a >10<sup>3</sup>-fold activation of DnaG activity. The interaction with DnaB also appears to broaden the primase recognition sequence to favour 5'-CAG-3' (Benkovic *et al.*, 2001). Yoda and Okazaki (1991) suggest that the interaction of DnaG with DnaB may completely release the specificity of *E. coli* DnaG for its triplet recognition sequence such that the DnaG consensus recognition sequence may be as non-specific as 5'-PyPyPu-3' (Py, pyrimidine; Pu, purine). This would increase the number of potential targets for DnaG, thereby increasing the frequency of priming.

In the mechanism proposed by Frick and Richardson (2001), the first NTP binds to the elongation site, while the second NTP binds to the initiation site and is incorporated at the 5' end of the primer. This NTP, bound to the initiation site, retains



its 5'-triphosphate moiety after dinucleotide synthesis (Frick and Richardson, 2001). The newly incorporated NTP of the growing oligoribonucleotide primer is transferred during each cycle of primer extension to the initiation site, clearing the elongation site of the primase for the binding of the next NTP (Frick and Richardson, 2001).

Primer transfer to DNA Pol III is mediated through the interaction of the DNA Pol III  $\chi$  subunit, SSB protein and DnaG. Yuzhakov *et al.* (1999) have proposed a three-point switch to explain this transfer mechanism. These authors have shown that DNA Pol III cannot assemble onto the primer template while DnaG is still bound. Since DnaG is tightly associated with its pRNA, it must be displaced for DNA Pol III to assemble. The underlying mechanism of this switch is the competition between DnaG and the  $\gamma$ -complex of DNA Pol III (see section 1.2.4) for SSB protein. The strong affinity of DnaG for the pRNA is possibly to protect the primer from nuclease attack, but appears mainly to be a means by which the DnaG specifies which DNA polymerase enzyme acts on the primer (Yuzhakov *et al.*, 1999). The switch begins with the DNA Pol III  $\chi$  subunit, part of the  $\gamma$ -complex that loads the  $\beta$ -clamp of the Pol III HE, contacting SSB protein bound to the ssDNA of the lagging strand. This disrupts the DnaG-SSB protein contact causing the SSB protein to switch contact between the DnaG primase and the DNA Pol III  $\chi$  subunit, displacing the DnaG primase. Consistent with this proposed mechanism was the finding that DNA Pol III HE lacking the  $\chi$  subunit was unable to displace DnaG bound to its primer (Yuzhakov *et al.*, 1999).

All prokaryote primase proteins are composed of three regions: an N-terminal zinc binding region, an oligoribonucleotide synthesis domain, and a C-terminal domain that interacts with a helicase. Three cysteinyl sulfhydryls and one histidine nitrogen near the N-terminus of the DnaG protein tightly bind one  $Zn^{2+}$  ion (Benkovic *et al.*, 2001). Due to the configuration of these residues, they could potentially form a zinc-finger (ZF) motif that may be involved in the recognition of priming sites on the DNA template, but is, however, thought to have only a purely structural role by preventing disulfide formation (Ilyina *et al.*, 1992; Benkovic *et al.*, 2001, Frick and Richardson, 2001). This ZF-like motif is only present in bacterial primases and has the consensus sequence CPFHXEKTPSF(T/S/A)VX<sub>3</sub>KQX(F/Y)HCFGC (Versalovic and Lupski,



1993). Generally, the primase oligoribonucleotide synthesis domain contains five motifs, labeled motif 2-6. No functional or structural role has been attributed to motif 2, but motifs 3-6 appear to be critical for NTP binding and oligoribonucleotide synthesis. Motif 3, also called the RNAP-basic motif, is a charged basic region with the consensus sequence G(R/K)X(V/I/L)X(F/Y)(G/S/A)(G/S/A)RX(V/I/L)X<sub>4</sub>P which is a motif both common and specific to primases, and prokaryote and eukaryote RNAP large subunits. However, the consensus sequence PKYLNSPET, which overlaps motif 3, is only found in bacterial primases, and may thus be a bacterial primase signature sequence (Versalovic and Lupski, 1993).

It has been found from sequence comparisons that the central region of DnaG, containing motifs 4-6, resembles a domain commonly found in topoisomerases and similar metal-binding phosphotransfer proteins. Consequently, this domain has been designated the toprim fold (see section 1.2.7). Hydrophobic residues that form a compact  $\beta/\alpha$  fold precede conserved negatively charged residues found in both motifs 4 and 5. Since these residues are strictly conserved in all prokaryote DnaG proteins studied, and cluster around the central domain, it has been suggested that this region forms the active site (Frick and Richardson, 2001). The physical interaction between DnaG and DnaB has been mapped to the C-terminal domain of DnaG. When the C-terminus of DnaG was removed by proteolysis, the truncated DnaG was able to synthesize oligoribonucleotides, but was inactive in helicase-dependent priming. The single amino acid substitution Q576A in the C-terminal domain of DnaG resulted in: (i) the synthesis of dramatically longer Okazaki fragments, (ii) a significantly lower affinity for DnaB, and (iii) a less efficient initiation of bidirectional replication than native DnaG (Keck *et al.*, 2000; Frick and Richardson, 2001). Mutational analysis in the catalytic centre of DnaG has shown that Lys-241 permits transcript initiation, but inhibits primer elongation, indicating that Lys-241 is an essential part of the catalytic centre (Keck *et al.*, 2000).

In summary, the step-wise assembly of eight proteins results in the formation of the primosome. This mobile multiprotein replication helicase/priming complex is necessary for the unwinding of dsDNA and the priming of the ssDNA for replication by the replisome. The primosome consists of DnaB, DnaG, SSB protein, DnaC and



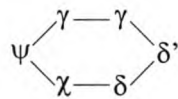
the products of the *dnaT* (DnaT, formerly protein i), *priA* (PriA, formerly protein n'), *priB* (PriB, formerly protein n), and *priC* (PriC, formerly protein n'') genes. Primosome assembly is initiated by the ATP-independent binding of PriA to DNA coated with SSB protein at specific sites termed PAS (primosome assembly sites) on the chromosome (Hiasa *et al.*, 1990, Zavitz and Marians, 1991). Once PriA is bound to the DNA template, PriB binds to the PriA-DNA complex. The DnaBC complex and DnaT transfers the DnaB hexamer to the PriA-PriB-DNA complex in an ATP-dependent, DNA-independent manner, thereby assembling the preprimosome. DnaG can then associate with this preprimosome complex forming the complete primosome which can now synthesize the pRNA. The interaction between DnaG and the primer is stabilized by SSB protein, resulting in DnaG remaining bound to the nascent RNA that serves as a primer for each Okazaki fragment (Zavitz and Marians, 1991; Benkovic *et al.*, 2001; Sutton and Walker, 2001).

#### 1.2.4) DNA Pol III holoenzyme

In *E. coli* the elongation phase of chromosomal replication involves the coordinated action of the DNA Pol III and DNA Pol I enzymes. In contrast to the single polypeptide chain of DNA Pol I, the sophisticated DNA Pol III HE consists of a ten-subunit, 18-polypeptide protein machine (Table 1.2).

The  $\gamma$  complex of DNA Pol III HE consists of five different proteins that form a seven-subunit ( $\gamma_3$ ,  $\delta$ ,  $\delta'$ ,  $\psi$ ,  $\chi$ ) processivity clamp loading complex that plays an important role in the initiation of the polymerase reaction by topologically tethering the DNA Pol III core complex to the DNA (Ellison and Stillman, 2001). The interaction of the  $\beta$  clamp with the  $\gamma$  complex initially drives the loading of the  $\beta$  clamp onto dsDNA. The  $\beta$  processivity clamp is a homodimer of the  $\beta$  subunit that is able to slide freely on dsDNA in both directions. The  $\beta$  clamp is ring-shaped having a diameter of  $\sim 35$  Å and a width of approximately one DNA helical turn (Benkovic *et al.*, 2001).

**Table 1.2:** Composition of DNA Pol III holoenzyme. Adapted from Snyder and Champness, (2003), and Benkovic *et al.* (2001).

Subunit	Gene	Mass (kDa)	Function	Contacts in subassemblies
$\alpha$	<i>dnaE</i> ( <i>polC</i> )	130	DNA polymerase catalytic subunit	} $\alpha$ - $\epsilon$ - $\theta$
$\epsilon$	<i>dnaQ</i> ( <i>mutD</i> )	27.5	3'-5' exonuclease	
$\theta$	<i>holE</i>	8.6	Stimulates exonuclease	
$\tau$	<i>dnaX</i>	71	Stimulates helicase, dimerizes core	} $\alpha$ - $\epsilon$ - $\theta$ $\theta\epsilon\alpha$ - $\tau$ - $\tau$ - $\alpha\epsilon\theta$ $\downarrow$ $\downarrow$ $\gamma$ - $\gamma$
$\gamma$	<i>dnaX</i> <sup>a</sup>	47.5	ATP-requiring clamp loader	} 
$\delta$	<i>holA</i>	38.7	Accessory protein, binds to $\beta$	
$\delta'$	<i>holB</i>	36.9	Accessory protein, stimulates ATPase of $\gamma$	
$\chi$	<i>holC</i>	16.6	Accessory protein, binds SSB	
$\psi$	<i>holD</i>	15.2	Accessory protein, with $\chi$ increases affinity for $\gamma$	
$\beta$	<i>dnaN</i>	40.6	Clamp protein	

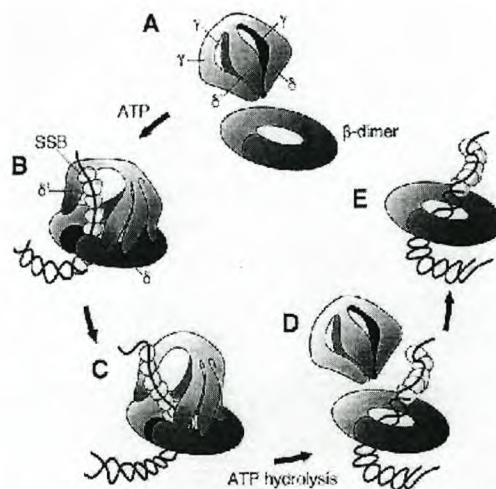
<sup>a</sup> The  $\gamma$  subunit is encoded in the N-terminal portion of the structural gene for  $\tau$  and in the same reading-frame.

Sequential assembly of the  $\beta$  clamp onto DNA by the  $\gamma$  complex occurs by the initial binding of ATP to the  $\gamma$  complex. The  $\gamma$ ,  $\delta$  and  $\delta'$  proteins of the  $\gamma$  complex are structurally similar to members of the AAA<sup>+</sup> ATPase protein superfamily. The AAA<sup>+</sup> ATPase protein superfamily are generally defined as proteins that must bind and hydrolyze ATP to perform their functions (Ellison and Stillman, 2001). In general, the AAA<sup>+</sup> superfamily proteins have three domains containing well conserved amino acid sequence motifs. Domain I (containing the P-loop, DExx box, and sensor 1 motif) is an anchor domain, which together with Domain II (containing the sensor 2 motif) constitutes the ATP binding and hydrolysis motifs. The Walker A and B motifs are also found in these domains. Domain III is a variable structure displaying considerable rotational ability dependent on ATP binding or hydrolysis (Ellison and Stillman, 2001).

The  $\gamma$  complex is an asymmetric heteropentamer that is thought to change its conformation upon binding ATP. The open conformation of the  $\gamma$  complex then binds the  $\beta$  processivity clamp homodimer (Fig 1.5). The precise function of ATP binding and hydrolysis is unclear. However, it is believed that the binding of ATP to the  $\gamma$  complex facilitates the presentation of the  $\delta$ -subunit to the  $\beta$  clamp. Following the



formation of the ATP- $\gamma$  clamp loader- $\beta$  clamp complex, a conformational change occurs in the  $\gamma$  complex resulting in the  $\delta$ -subunit acting like a molecular wrench to open the  $\beta$  clamp. The action of the  $\delta$ -subunit induces a conformational change in the  $\beta$  clamp, which allows DNA to insert through the open  $\beta$  clamp. Hydrolysis of ATP by the DNA- $\gamma$  complex- $\beta$  clamp assembly is thought to either position the dsDNA within the interior of the  $\beta$  clamp, or results in the release of the  $\beta$  clamp from the complex allowing it to close on the DNA (Benkovic *et al.*, 2001; Ellison and Stillman, 2001).

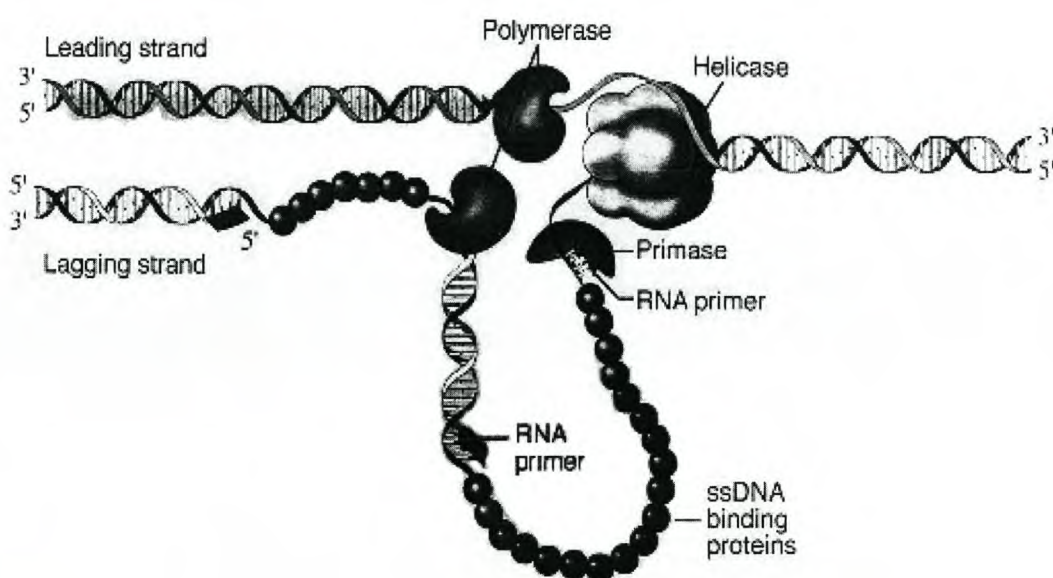


**Figure 1.5:** Clamp loading. Reproduced from Ellison and Stillman (2001).

It is this loading of the  $\beta$  subunit clamp that stimulates the intrinsic ATPase activity of the DnaA protein which inactivates DnaA for further replication initiation (Messer *et al.*, 2001).

Dissociation of the  $\gamma$  complex from DNA occurs by the hydrolysis of a second ATP molecule, leaving the  $\beta$  clamp behind. Subsequently, the  $\gamma$  complex is replaced by the core polymerase which is now tethered to the DNA by the  $\beta$  clamp. The binding of the DNA Pol III core complex to the  $\beta$  clamp prevents the  $\gamma$  complex from unloading the  $\beta$  clamp. This ensures that only once DNA synthesis has finished, can the  $\beta$  clamp be recycled. The  $\chi$  subunit of the  $\gamma$  complex interacts with SSB protein, and in doing so displaces DnaG from the pRNA of each Okazaki fragment allowing DNA Pol III to bind for subsequent elongation [section 1.2.3] (Benkovic *et al.*, 2001; Sutton and Walker, 2001).

The DNA Pol III core complex consists of the  $\alpha$  (catalytic),  $\epsilon$  (proofreading), and  $\theta$  (stimulates proofreading) subunits. A dimer of the  $\tau$  subunit serves to bridge two  $\alpha$  subunits to form a dimeric core  $[(\alpha\epsilon\theta)_2\tau_2]$ , resulting in the physical coupling of the two active core polymerases (Fig 1.6). This coupling is required for coordinated leading and lagging strand synthesis. The C-terminus of the  $\tau$  subunit appears to be necessary for binding to the DnaB helicase. The physical contact between the  $\tau$  subunit and DnaB serves to increase the unwinding rate of the helicase, so that the fork moves at the speed of DNA Pol III (Benkovic *et al.*, 2001).



**Figure 1.6:** Physical coupling of the core polymerase enzymes causing a looping of the lagging strand. Reproduced from Frick and Richardson (2001).

With DNA Pol III HE assembly complete, and coupled to the primosome, the newly formed replisome can initiate chain elongation. The core polymerase  $(\alpha\epsilon\theta)_2\tau_2$  synthesizes DNA at a rate of approximately  $20 \text{ nucleotides}\cdot\text{s}^{-1}$  with a processivity in the tens of nucleotides. However, in association with the  $\beta$  clamp, the DNA Pol III polymerization rate increases to  $\sim 750 \text{ nucleotides}\cdot\text{s}^{-1}$ , and has a processivity in excess of  $5 \times 10^4$  nucleotides (Benkovic *et al.*, 2001).

A molecular switch appears to coordinate the recycling of the  $\beta$  clamp during replication. In the presence of primed DNA the  $\beta$  clamp has an affinity for the core



polymerase, and it is this binding that blocks the ability of the  $\gamma$  complex to unload the  $\beta$  clamp. Only once DNA synthesis has finished can the  $\beta$  clamp be recycled by the  $\gamma$  complex. It has been found that in the absence of the core polymerase, the  $\beta$  clamp has a 30-fold greater affinity for the  $\gamma$  complex (Benkovic *et al.*, 2001). When DNA Pol III reaches a previously synthesized pRNA, the above process is repeated

### 1.2.5) DNA Pol I and DNA ligase

Little is known about the coordination of the DNA Pol I and DNA Pol III actions during Okazaki fragment fusion, but together with DNA ligase, and its 5'-3' activity, DNA Pol I brings about Okazaki fragment maturation (Sutton and Walker, 2001).

DNA Pol I is a 109 kDa protein composed of two functional domains, with the large C-terminal domain referred to as the Klenow fragment (Klenow and Henningsen, 1970; Patel *et al.*, 2001). DNA Pol I has been shown to possess three catalytic activities. The Klenow fragment contains the 5'→3' polymerase and 3'-exonuclease activities, while the N-terminal domain contains the 5'-exonuclease activity. It is the 5'-exonuclease activity of DNA Pol I that acts on the pRNA, cleaving ribonucleotides in advance of the polymerase activity. The 3'-exonuclease activity serves as a "proofreading" function to improve the fidelity of the polymerase activity. The metal-ion coordinated hydrolytic attack by the 3'-exonuclease domain of the Klenow fragment on misincorporated bases is thought to be induced by a destabilized DNA helix caused by mismatched base-pairs. This activation of the 3'-exonuclease leads to an increased ratio of exonuclease to polymerase activity resulting in rapid excision of the misincorporated base (Beese and Steitz, 1991). It is the combined action of these three activities that allows DNA Pol I to remove the pRNA by the mechanism termed nick translation. Nick translation replaces the ribonucleotides of the pRNA with deoxyribonucleotides (Patel *et al.*, 2001).

The *E. coli* DNA Pol I is a member of the 'A' family of polymerases, with the polymerase domain of the Klenow fragments having substantial blocks of highly conserved residues separated by regions of less amino acid sequence similarity (Polesky *et al.*, 1990; Patel *et al.*, 2001). There is, however, significant protein



tertiary structure similarity within this family. For instance, secondary structure elements,  $\alpha$ -helices and  $\beta$ -strands of *E. coli* DNA Pol I and *Taq* Pol I are all nearly superimposable (Patel *et al.*, 2001).

Since the DNA Pol III holoenzyme can only add deoxyribonucleotides to a 3'-OH terminus, when it reaches the 5'-end of the subsequent pRNA it dissociates. The nick translation activity of DNA Pol I hydrolytically degrades the pRNA allowing the polymerase activity of DNA Pol I to incorporate deoxyribonucleotides (Mathews and van Holde, 1990). The general mechanism for the DNA Pol I polymerase activity involves binding the template primer, binding the appropriate dNTP with the DNA Pol I:DNA complex, nucleophilic attack resulting in phosphodiester bond formation, and release of pyrophosphate (PPi). The rate limiting step is thought to be either phosphodiester bond formation, or a conformational change that precedes nucleotide incorporation. DNA Pol I is thought to undergo at least four significant conformational changes for nucleotide incorporation (Patel *et al.*, 2001). High-resolution complexes studied show that DNA Pol I interacts with the sugar-phosphate backbone along the minor groove, and these interactions are associated with bending of the DNA. This bending results in the DNA adopting an S-shaped conformation (Patel *et al.*, 2001).

A "base-flipping" mechanism has been proposed by Patel *et al.* (2001) for the polymerization reaction catalyzed by DNA Pol I. It is proposed that the first template base is rotated  $\geq 90^\circ$ , and the adjacent template bases are rotated  $180^\circ$ , such that these bases are 'flipped' outside the DNA helix axis and away from the dNTP binding site. The incoming dNTP then binds with the DNA Pol I polymerase domain, the template base rotates back into the helix axis, and the incoming dNTP interacts with its complementary template base by Watson-Crick base-pairing. The catalytic site of the polymerase potentially contains two metal ions, one of which is thought to facilitate deprotonation of the primer 3'-OH group. The close proximity of the nucleophile 3'-OH group and the  $\alpha$ -phosphate of the incoming base causes the phosphoryl transfer reaction (Patel *et al.*, 2001).

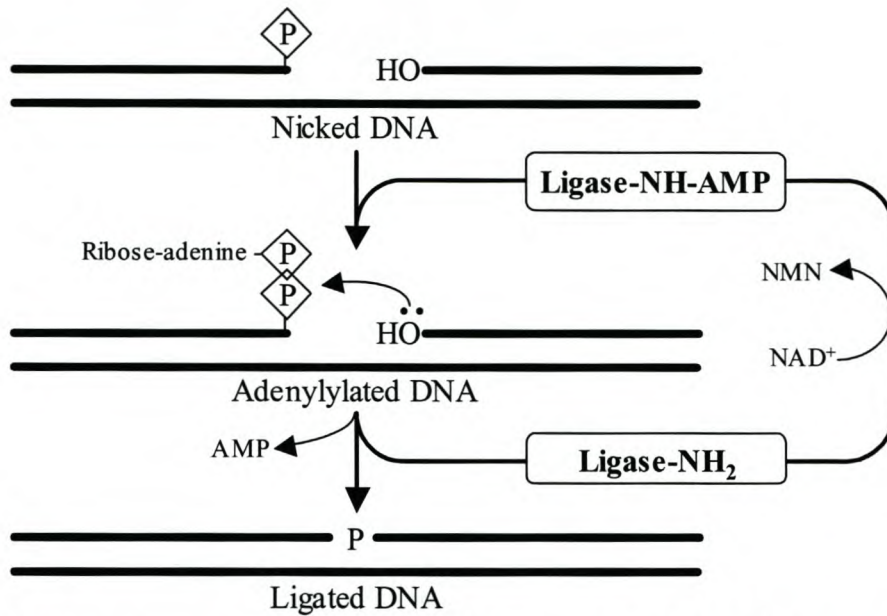


Patel *et al.* (2001) suggest that the “base-flipping” mechanism may have general relevance to the DNA polymerization mechanism.

The strain created by DNA helix distortion upon DNA Pol I binding to DNA, as well as the 90-180° base-flipping reaction, may help to guide the DNA Pol I 5'→3' direction of translocation with respect to the primer. As many of the DNA Pol I-DNA interactions are non-specific, this allows the DNA Pol I enzyme to translocate freely in a spiral motion (Patel *et al.*, 2001).

The DNA Pol I enzyme, however, cannot join the 5'-phosphoryl-terminus of the last incorporated deoxyribonucleotide to the 3'-OH terminus of the following Okazaki fragment. This is achieved through the action of DNA ligase coded for by the *E. coli lig* gene (Mathews and van Holde, 1990).

The active site of DNA ligase contains a lysine residue that is activated by adenylation. In the case of the *E. coli* ligase, NAD<sup>+</sup> is cleaved and the resultant AMP moiety is linked to the amino group of the lysine residue through a phosphoamide bond. Consequently, adenylylated DNA ligase and nicotinamide mononucleotide (NMN) are formed (Fig 1.7). The enzyme then transfers the adenylyl group to the 5'-terminal phosphate of the DNA substrate. This activates the 5'-terminal DNA substrate for nucleophilic attack by the 3'-OH group of the directly adjacent deoxyribonucleotide. The nucleophilic attack results in phosphodiester bond formation, release of AMP, and linkage of the Okazaki fragments to form a continuous daughter strand (Lehman, 1974; Mathews and van Holde, 1990).



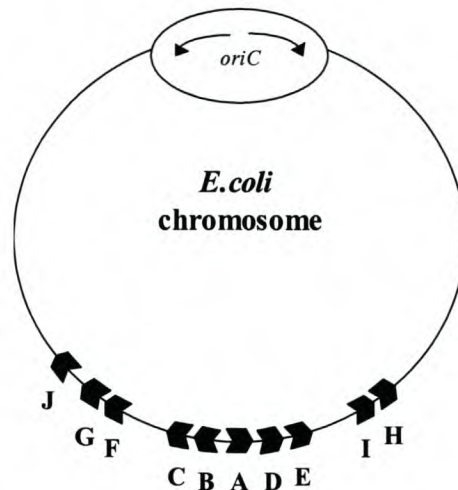
**Figure 1.7:** The reaction catalyzed by *E. coli* DNA ligase. Reproduced from Mathews and van Holde (1990).

### 1.2.6) Tus-mediated replication termination

Replication termination occurs at specific sequences called replication termini (*Ter*). In *E. coli* there are 10 *Ter* sites (Fig 1.8) having the consensus sequence 5'-(A/C/T)N<sub>2</sub>(T/A/C)(A/C/T)G(T/G/A)A(T/G)GTTGTAA(C/T)(T/G)AN<sub>4</sub>-3', and they are found diametrically opposite the *oriC*. These *Ter* sites have polarity as they are found in two clusters of five with each cluster having a polarity opposite to that of the other. This polarity is explained as the ability of a cluster to arrest a replication fork in one direction, but allow the replication fork to pass through unimpeded when approaching from the opposite direction (Mulugu *et al.*, 2001)..

Replication fork arrest occurs through a protein-protein interaction involving the DnaB helicase and the replication terminator protein Tus. The Tus protein specifically interacts with these *Ter* sites, thereby creating a replication trap that forces the two replication forks initiated at the *oriC* to meet within a well defined region of the chromosome (Mulugu *et al.*, 2001).





**Figure 1.8:** Diagrammatic representation of the *E. coli* *oriC* and the location of the 10 *Ter* sites (A to J). Arrows show the direction of the two replication forks originating from the *oriC*. Reproduced from Mulugu *et al.* (2001).

There are currently two models that attempt to explain the mechanism of fork arrest by Tus. The first of these is the “roadblock” model which suggests that the binding of Tus to a *Ter* site is strong enough to form a non-specific barrier to DnaB. This model fails to explain how replication forks are able to pass through DNA-protein complexes on the chromosome, some of which bind DNA with strong specificity, yet is unable to displace the Tus protein at the *Ter* site.

The second model is the helicase-contrahelicase model. This model regards the Tus protein as a polar contrahelicase as it interferes with the DNA unwinding activity of DnaB in an orientation specific manner. Consequently, the Tus protein is said to have a non-blocking and a fork-blocking face (Mulugu *et al.*, 2001). The helicase-contrahelicase model suggests that a helicase approaching from the non-blocking face of Tus can displace Tus, and pass through the site unimpeded. If the helicase approaches from the blocking face of Tus it would contact a critical region of the Tus protein called the L1 loop, and fork arrest would occur. This mechanism of replication fork arrest seems to explain the polarity of replication fork termination. It is not known, currently, whether the ATP-hydrolyzing activity of DnaB is abolished at this point due to Tus-mediated arrest. There is strong molecular evidence for the helicase-contrahelicase model as the crystal structure of the Tus-*Ter* complex has been solved, and yeast two-hybrid analyses have proven Tus-DnaB interaction (Mulugu *et al.*, 2001).

### 1.2.7) Topoisomerases

The two polynucleotide strands of dsDNA wind around each other in a right-handed helix, thereby generating one helical turn per 10.5 nucleotides. Supercoiling is the coiling of the helix axis itself. Supercoiled DNA is twisted opposite to the direction of the double helix, with about one negative supercoil per 15 double helical twists. Most dsDNA exists in physiological conditions in a negatively supercoiled form (Cozzarelli, 1980; Snyder and Champness, 2003).

As the circular prokaryote chromosome has no free ends to rotate, strand separation by the primosome increases torsional stress and supercoiling. To relieve this stress, topoisomerases undo the supercoiling ahead of the replication fork (Snyder and Champness, 2003). Topoisomerases are a group of enzymes able to interconvert different topological isomers of DNA.

Topoisomerases are divided into two groups. Type I topoisomerases, encoded by the *E. coli topA* gene, catalyze the removal of negative supercoils leading to the relaxation of supercoiled DNA (Kowalczykowski *et al.*, 1994; Noble and Maxwell, 2002). This is achieved by cleavage of only a single strand of dsDNA and then passing it through its complementary strand in an ATP-independent manner prior to religation. The type I topoisomerases consist of two distinct groups, type IA and type IB, which differ by their covalent linkage to the cleaved strand. The type IA topoisomerases form a covalent phosphotyrosine linkage to the 5'-OH group of the deoxyribonucleotide, whereas type IB form a covalent bond to the 3'-OH group (Noble and Maxwell, 2002).

The type II topoisomerases are all ATP-dependent and catalyze the relaxation of negatively supercoiled DNA by cleaving both phosphodiester backbones of dsDNA, transporting another segment of dsDNA through the cleaved segment (referred to as the 'gate' segment), and then religating the DNA (Noble and Maxwell, 2002). During the cleavage reaction the type II topoisomerases form a phosphotyrosine intermediate with the 5'-OH group of the deoxyribonucleotide (Noble and Maxwell, 2002).



A common feature identified in topoisomerases and primases is the toprim fold (Noble and Maxwell, 2002). The most highly conserved amino acid residues of the toprim fold are three acidic residues. However, four are conserved between the type IA and type IIA topoisomerases: the consensus being Glu separated by approximately 80 amino acids from the motif DXDXD (the third Asp is replaced by Glu in type IA topoisomerases). As these residues are close together in the toprim fold, and structures from DnaG-type primases have been shown to coordinate divalent cations, this fold probably binds  $Mg^{2+}$  (Noble and Maxwell, 2002).

Although capable of relaxation of negative supercoils, the preferred reaction of DNA gyrase is DNA supercoiling. This type II topoisomerase is an  $A_2B_2$  tetramer composed of GyrA and GyrB subunits (Hockings and Maxwell, 2002). The GyrA dimer is thought to assemble on DNA with two GyrB subunits. The GyrB subunit, which has ATPase activity, binds ATP, and in doing so, closes the GyrB clamp. The closure of the GyrB clamp traps the DNA segment that is to be transported through the gate segment. The active-site tyrosine residue responsible for nucleophilic attack on the phosphate group at the cleavage site is found on the GyrA subunit. A two metal-ion mechanism has recently been proposed for DNA cleavage by DNA gyrase. The mechanism involves two  $Mg^{2+}$  ions which polarise the active site tyrosine. The tyrosine residue then attacks the scissile phosphate group forming a negatively charged pentacovalent phosphate intermediate that is stabilized by both metal-ions (Noble and Maxwell, 2002).

The gate segment is cleaved at sites four bases apart, forming 5'-phosphotyrosine intermediates, and the two ends are separated. The segment to be transported is passed from GyrB through the gate segment to GyrA. The cleaved DNA is then aligned and religated, following which ATP is hydrolyzed causing the opening of the GyrB clamp (Hockings and Maxwell, 2002). This reaction introduces two negative supercoils into DNA, and may occur concomitantly with, or shortly after, the termination of replication (Mathews and van Holde, 1990; Snyder and Champness, 2003).

Another property of type II topoisomerases is decatenation of interlinked, or catenated, chromosomes formed at the end of a round of replication. Decatenation of



two interlinked chromosomes will allow for the separation of the newly replicated chromosome prior to cell division (Mathews and van Holde, 1990; Noble and Maxwell, 2002; Snyder and Champness, 2003).

### **1.3) Modes of plasmid replication**

A discussion of prokaryote chromosomal replication has been undertaken because the degree to which a plasmid relies on the host cell's replication machinery dictates its host range. The evolution of distinct plasmid replication mechanisms is probably a consequence of multiple horizontal transfers of genes from bacteria to their specific plasmids, suggesting that the ability of a plasmid to become established and maintained in a host is a function of plasmid-host coevolution. Moreira (2000) argues this point using bioinformatic analyses to show that bacterial chromosomally-encoded helicases do not form a monophyletic group in unrooted phylogenetic trees, but are intermixed with plasmid ortholog sequences.

The often used definition of a plasmid as an extrachromosomal element capable of autonomous replication, suggests that although some plasmids utilize elements of the host's replication machinery, plasmids must contain features very similar to chromosomes that allow for the initiation and control of their replication.

Osborn *et al.* (2000) state that the absolute minimum requirement for a plasmid is the ability to replicate, and thus the replicon is the defining feature of a plasmid. Generally, the replicon of a plasmid contains three features: (i) each plasmid has an origin(s) of replication (termed *ori*) distinct for that replicon, (ii) a plasmid encoded protein is generally required for the initiation of replication (often referred to as the Rep protein), and (iii) replication-controlling mechanisms are often plasmid-encoded (del Solar *et al.*, 1998).

The origin of replication (*ori*) found on plasmids can be defined as: (i) the minimal *cis*-acting region that can support autonomous replication of a plasmid, (ii) the region where DNA strands are melted to initiate the replication process, or (iii) the base(s) at which leading-strand synthesis starts (del Solar *et al.*, 1998).



Concomitant with the initiation of replication is the requirement for replication control. Replication control maintains a fixed level, or copy number, of the plasmid pool within a cell. As each plasmid copy replicates on average once per cell cycle, replication control mechanisms are very important for the correction of plasmid copy number fluctuations (del Solar *et al.*, 1998). The inability to sense copy number fluctuations in a population of different plasmid replicons can directly affect the segregational stability of a plasmid replicon within the plasmid pool. As individual plasmid replicons are selected at random for replication from a pool that includes replicated and non-replicated copies, replication control provides a means by which a plasmid can kinetically communicate its presence to the other copies of the same replicon. In doing so, it allows the plasmid to set its own replication frequency according to the total concentration of the particular replicon (Espinosa *et al.*, 2000; Paulsson, 2002).

A plasmid maintains its characteristic copy number in a particular host by encoding functions that autoregulate the initiation of replication. Replication of plasmids must be controlled, as unchecked plasmid copy number fluctuations can have an adverse effect on the metabolic fitness of the host cell. The plasmid's copy number, its gene expression levels, and the history of plasmid-host coevolution, all contribute to the metabolic burden imposed by the plasmid on the host (del Solar *et al.*, 1998; Paulsson, 2002).

Chattoraj (2000) affirms that the initiation rate per plasmid is a decreasing function of plasmid concentration. For instance, if partitioned randomly, plasmids should be present at random copy numbers in the daughter cells after cell division. A daughter cell receiving fewer than the average copy number of plasmids will have this fluctuation corrected by the plasmids undergoing compensatory replication until the characteristic copy number is reached (Gruss and Ehrlich, 1989). Compensatory replication is halted when an equilibrium between replication initiator and inhibitor is achieved. A higher plasmid concentration results in a higher inhibitor concentration, and this decreases the rate of replication initiator synthesis, thereby affecting the plasmid replication frequency (Paulsson, 2002). This generalized negative feedback mechanism (autoregulation) appears to be the primary regulatory strategy by which plasmids control replication initiation. Using this mechanism, plasmids are able to



sense and correct deviations from the average copy number in an individual cell (Novick, 1987; del Solar *et al.*, 1998).

Generally, three types of plasmid copy number control systems have been recognized. Each system is classified according to the type of negative control element utilized. These negative regulatory elements include: (i) directly repeated sequences (iterons) that complex with a replication initiating protein, (ii) antisense RNA's that inactivate a replication initiator, and are consequently termed countertranscribed RNA (ctRNA), (iii) or a combination of both ctRNA and a regulatory protein (del Solar and Espinosa, 2000). This third option, involving a combination of ctRNA and a regulatory protein, can be further subdivided into systems where the regulatory protein plays an auxiliary role, or systems where it acts directly as a transcriptional repressor (del Solar and Espinosa, 2000).

The negative feedback loop utilizing iterons includes three levels of replication control: (i) autoregulation of the plasmid replication initiating protein, (ii) sequestration of the replication initiating protein, (iii) or the inactivation of the origin by a pairing mechanism [handcuffing] (Espinosa *et al.*, 2000).

Circular plasmids replicate by either a rolling circle or a theta-type (including strand displacement) mechanism (del Solar *et al.*, 1998). Each of these replication mechanisms will be discussed with reference to well-studied plasmids of each replication type. The copy number control systems of plasmids will also be discussed, so as to demonstrate the various regulatory mechanisms employed by plasmids to counter copy number fluctuations.

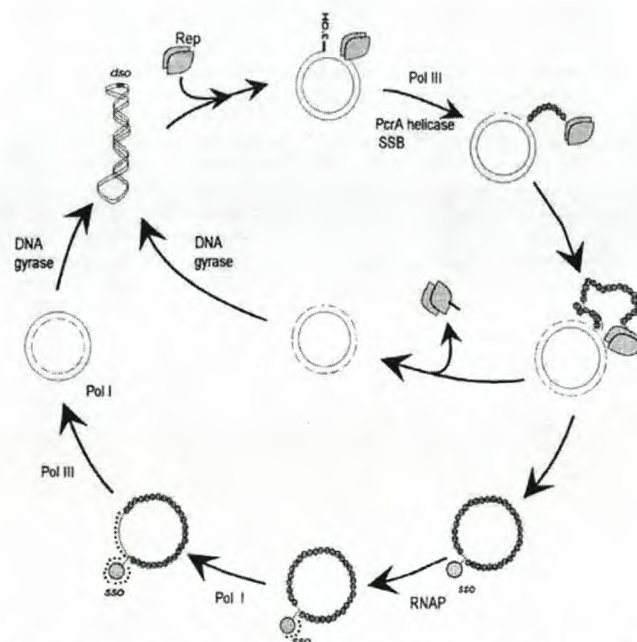
### **1.3.1) Rolling circle replication**

Initially thought to be exclusively used by ssDNA coliform bacteriophage, rolling circle replication has been identified in over 200 plasmids isolated from a variety of Gram-positive and Gram-negative bacteria, cyanobacteria and Archaea (del Solar *et al.*, 1998; del Solar *et al.*, 2002).



### 1.3.1.1) *General replication mechanism*

The general mechanism for rolling circle replication (Fig 1.9) involves the initiation of replication by the plasmid-encoded Rep protein. The Rep protein introduces a nick in the parental [+] -strand of the plasmid at a site termed the double strand origin (*dso*), leaving an exposed 3'-OH end from which leading strand synthesis can be initiated (Espinosa *et al.*, 2000). Consequently, no pRNA is required for leading strand synthesis, and the parental [+] -strand is displaced during synthesis of the new [+] -strand. Therefore, rolling circle replication is unidirectional, and is considered to be asymmetric as leading and lagging strand synthesis is uncoupled (Kramer *et al.*, 1997; del Solar *et al.*, 1998).



**Figure 1.9:** Diagrammatic representation of the general rolling circle replication mechanism. See text for details. Reproduced from Espinosa *et al.* (2000).

During leading strand synthesis the new [+] -strand is covalently bound to the parental [+] -strand. The parental [+] -strand is only displaced once the replisome reaches the reconstituted *dso* and strand-transfer reactions are completed to terminate replication. The result of leading strand synthesis is a reconstituted dsDNA molecule composed of the parental [-] -strand and the newly synthesized [+] -strand. Host replication proteins act on the single-strand origin (*sso*) of the displaced ssDNA parental [+] -strand converting this ssDNA replication intermediate into dsDNA. The *sso* is a non-coding DNA region physically separated from the *dso* from which lagging strand synthesis is



initiated with the aid of the host RNAP. Host replication proteins involved in rolling circle replication have not been fully identified, but it is assumed that a host encoded helicase and DNA Pol III HE extends the leading strand, and that SSB protein participates in the replication mechanism as ssDNA intermediates are formed (del Solar *et al.*, 1998; Kramer *et al.*, 1998; Espinosa *et al.*, 2000). There is no general mechanism that has been identified for replication termination as different interactions with the Rep proteins appear to take place at the end of a round of replication.

### 1.3.1.2) *Origin comparisons*

Plasmids utilizing rolling circle replication have an essential region composed of the *dso*, the *rep* gene, and the elements involved in plasmid replication control. Together these regions are referred to as the leading-strand initiation and control (LIC) module (del Solar *et al.*, 1998).

Within the *dso* is found the *bind* and *nic* regions where the Rep protein binds to the plasmid template and introduces the initial nick. The *nic* site is highly conserved amongst replicons of the same family, whereas the *bind* region appears to define replicon specificity as these regions are not conserved (Gruss and Ehrlich, 1989; del Solar *et al.*, 1998).

Based on the homologies in the LIC module, four plasmid families utilizing rolling circle replication have been defined. These families are represented by pT181, pC194, pMV158, and pSN2 (Espinosa *et al.*, 2000). To date, little is known about the pSN2 family of plasmids. Comparisons of the organizational structure of the LIC module of these four families shows that the pT181 and pC194 families have a *bind* region composed of inverted repeats contiguous to the *nic* site, while the pMV158 family has a *bind* region composed of a set of two or three direct repeats separated from the *nic* site by a spacer region of 14 to 95 nucleotides (del Solar *et al.*, 1998; Espinosa *et al.*, 2000). Hairpin structures in the *nic* region that contains the DNA sequence where the nick is introduced have been identified for members of the pT181 (IR-II region) and pMV158 (hairpin I region) families. A stem-loop structure has been identified by genetic analysis in the *nic* region of pC194 (del Solar *et al.*, 1998; Espinosa *et al.*, 2000). An absolute requirement for rolling circle replication is that



DNA is supercoiled. The supercoiled DNA topology is most likely required for cruciform extrusion allowing for the Rep proteins to act on these stem-loop structures (Espinosa *et al.*, 2000).

The *sso* of the ssDNA intermediate molecule is generally only fully functional in its original or closely related hosts (Kramer *et al.*, 1997). The *sso* contains imperfect palindromic structures capable of forming imperfect stem-loop structures which function in an orientation-dependent manner (Gruss and Ehrlich, 1989; del Solar *et al.*, 1998). Various types of *sso* have been studied, but in light of the above discussion the *ssoA* of pT181, pC194, and pMV158 will be considered. The pMV158 plasmids are unique in that they have two distinct *sso* regions designated *ssoA* and *ssoU* (Kramer *et al.*, 1998). Analysis of various *ssoA* sequences and structures revealed a highly conserved site. This site, termed  $RS_B$ , is a recombination site involved in inter-plasmid recombination, and is the site to which RNAP binds for the RNAP-directed synthesis of the pRNA for lagging strand synthesis (Kramer *et al.*, 1997; Kramer *et al.*, 1998). Comparisons between the pT181 and pMV158 *ssoA* revealed a conserved sequence 5'-TAGCG(T/A)-3', designated CS-6, which has been shown to function as a transcriptional terminator of pRNA synthesis (Kramer *et al.*, 1997; del Solar *et al.*, 1998).

### **1.3.1.3) Replication initiator protein comparisons**

As the Rep proteins appear to be involved with the initiation and termination of replication, and are thus responsible for a DNA cleavage and rejoining reaction, they are said to have a type I topoisomerase-like activity (del Solar *et al.*, 1998).

The Rep proteins of these families have differing quaternary structures. The active RepC of pT181 is a dimer with one monomer participating in the DNA replication initiation reaction, while the other monomer participates only in the replication termination reaction. The RepC monomer responsible for the replication initiation reaction is involved in both sequence-specific recognition of the *dso*, and cleavage of the *nic* site (Chang *et al.*, 2000). The RepB of pMV158, however, appears to be a hexamer. By analogy with another Rep protein of the same family, the RepA of pC194 is likely to be a monomer (del Solar *et al.*, 1998).



There are several conserved motifs found on the Rep proteins with similarity to motifs found in Tra and Mob proteins responsible for plasmid transfer and conjugative mobilization. There are, however, a further two motifs which correspond to a catalytic domain and a metal-binding domain. The metal-binding domain, termed the HUH domain, is not found in the Rep proteins of the pT181 family. None of the Rep proteins appear to have the classic helix-turn-helix (HTH) motif of DNA-binding proteins, although a LZ motif has been tentatively identified in the RepB protein of pMV158 (Gruss and Ehrlich, 1989; del Solar *et al.*, 1998).

Residues 267-270 in the C-terminal of the 314 amino acid RepC monomer have been identified as being critical for the sequence-specific recognition of the *dso*. In the catalytic domain a conserved Tyr residue is involved in the nucleophilic attack on the plasmid DNA. It is Tyr-191 of the pT181 RepC protein that is involved in the cleavage of the phosphodiester bond. The residue responsible for this activity in RepA proteins of pC194 is Tyr-214, although Glu142 and Glu-210 are also required. This has led to the suggestion that for the pC194 RepA protein the Tyr-214 is involved in the covalent binding to the *bind* site for the initiation of replication, while Glu-210 participates in the termination step by directing the hydrolysis of the phosphodiester bond. Stable covalent binding to the plasmid template does not appear to be a general feature as RepB of pMV158 does not form a stable covalent phosphotyrosine linkage with its target DNA (del Solar *et al.*, 1998; Chang *et al.*, 2000).

#### **1.3.1.4) Replication and control mechanism comparisons**

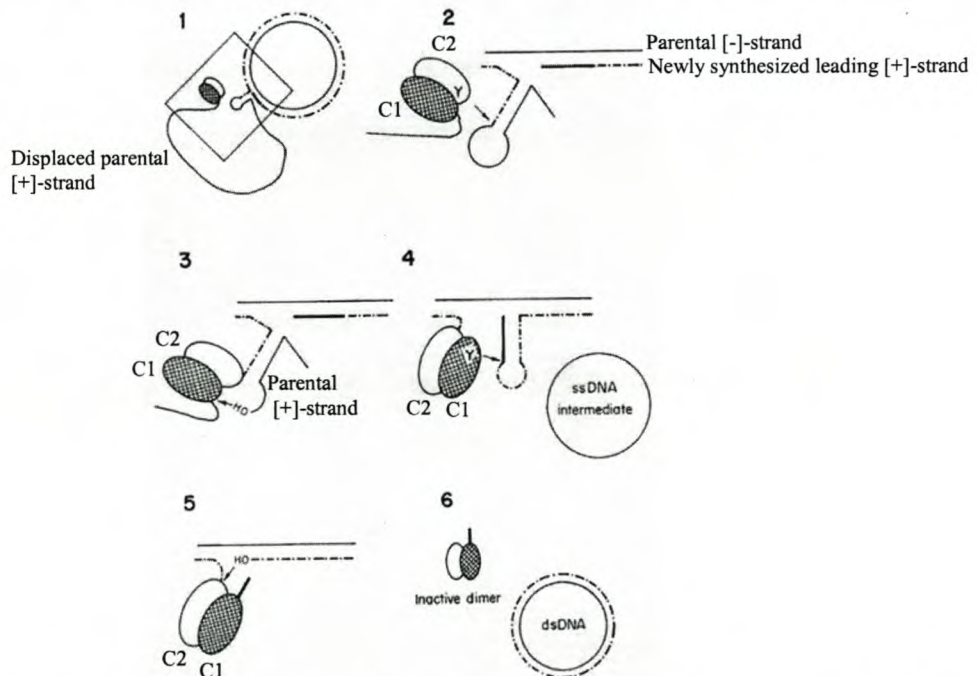
The rolling circle replication mechanism has only been studied in depth for the pT181 family of plasmids. Even so, information about the initiation complex is limited. Replication is initiated by one monomer ('C1'; Fig 1.10) of the pT181 RepC homodimer recognizing the *dso* IR-III region of pT181 in a sequence-specific manner. The Tyr-191 residue of the same monomer (C1) is responsible for cleaving the *nic* site found in the *dso* IR-II region by nucleophilic attack. The replication initiating monomer of the RepC homodimer (C1) becomes covalently attached to the 5'-end of the nicked strand (Fig 1.10). This covalent attachment is achieved through a



phosphotyrosine bond with the thymidine base exposed by the cleavage reaction (Wang *et al.*, 1993; Chang *et al.*, 2000).

In pT181's natural host, *Staphylococcus aureus*, the helicase which acts to unwind the dsDNA is chromosomally encoded. The chromosomal PcrA protein, which has similarity to UvrD helicase and Rep helicase of *E. coli*, is recruited to the *nic* site by physical interaction with RepC. It is possible that the RepC now also becomes attached to the PcrA monomer, and may translocate with the PcrA helicase as it moves just ahead of the replication fork. As RepC is covalently attached to the displaced leading strand of the DNA, PcrA helicase may either translocate on the template strand (in a 3'→5' direction), or on the displaced strand (in a 5'→3') direction. Both helicase activities have been demonstrated, with a much stronger 5'→3' activity suggesting that translocation occurs on the displaced strand (Chang *et al.*, 2002).

The 3'-OH end generated by the nicking reaction serves as a primer for displacement synthesis of the new [+] strand by DNA Pol III (Fig 1.10; [1]) (Chang *et al.*, 2002).



**Figure 1.10:** Rolling circle replication model for the pT181 family of plasmids. See text for replication mechanism. C1 and C2 distinguishes between the two monomers of the RepC homodimer, and Y refers to the catalytic Tyr-191 residue of each RepC monomer. Reproduced from del Solar *et al.* (1998).



Leading strand synthesis proceeds further than one round of replication, reconstituting the *nic* site of the IR-II hairpin structure in both the displaced strand and the newly synthesized leading strand (Fig 1.10; [2]). The Tyr-191 ('Y') of the RepC monomer not involved in replication initiation (C2) now cleaves the regenerated, but displaced *dso*, by nucleophilic attack, and becomes covalently bound to the 5'-end of the newly synthesized leading strand (Fig 1.10 [2,3]). The 3'-OH end of the parental [+]-strand released in this reaction now attacks its own tyrosyl-phosphodiester bond formed between its 5'-end and the RepC monomer (C1) that was active in the initiation step (Fig 1.10; [3]). In doing so it releases the ssDNA parental [+]-strand intermediate. The now free Tyr-191 OH-group of the RepC monomer (C1) involved in replication initiation attacks the *nic* site generated by further extension of the leading strand, and cleaves off a 10 to 12-bp oligonucleotide which becomes attached to this monomer (Fig 1.10; [4,5]). The 3'-OH of the newly synthesized leading strand now attacks the tyrosyl-phosphodiester bond formed between its 5'-end and the second RepC monomer (C2) (Fig 1.10; [5]). The result of these two reactions is the synthesis of an intact dsDNA molecule, and the release of an inactive heterodimer (Fig 1.10; [6]). The heterodimer is formed by the attachment of the oligonucleotide to the initiating RepC monomer (C1). This heterodimer is thus unable to reinitiate replication (Rasooly and Novick, 1993; del Solar *et al.*, 1998).

The ssDNA parental [+]-strand intermediate must now be converted to a dsDNA molecule, and this occurs via lagging strand synthesis (Fig 1.10; [6]). No known plasmid functions seem to be encoded for this conversion, and it appears to be mediated by RNAP and DNA Pol I. Analysis of the conversion of the ssDNA intermediate to dsDNA initiated from a pMV158-*ssoA* suggests that host encoded RNAP binds to the RS<sub>B</sub> and produces a pRNA of 20 nucleotides (Kramer *et al.*, 1997). The host DNA Pol I then probably extends the polynucleotide chain for a further 80-100 bases before the host DNA Pol III completes the synthesis of the growing chain. It is interesting to note that for efficient DNA Pol I activity, the *Streptococcus pneumoniae* DNA Pol I used by Kramer *et al.* (1997) could not have the 5'-exonuclease domain removed from the enzyme. Since this domain is required for chromosomal pRNA processing, this suggests that DNA Pol I processes the pRNA on the ssDNA intermediate by passing the primer through a helical arch in the exonucleolytic domain which degrades the primer (Kramer *et al.*, 1997). Sequences

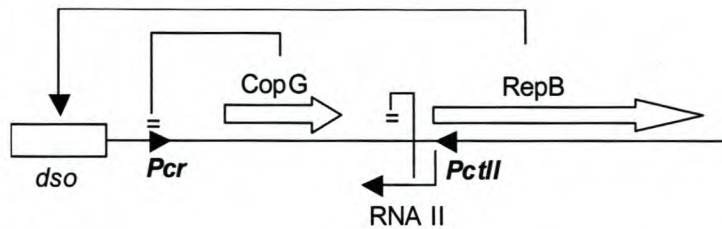


in the CS-6 region, which seem to act as a transcriptional terminator, end RNAP-directed pRNA synthesis (Kramer *et al.*, 1997).

Analysis of pMV158 without an *sso* region that was able to accomplish the ssDNA intermediate to dsDNA conversion suggested an alternate pathway to achieve ssDNA conversion. This pathway may involve plasmid-encoded antisense RNA's which anneal to complementary regions and act as primers in a process mediated by the host RecA protein (del Solar *et al.*, 1998).

Transcriptional attenuation is the mechanism used by pT181 for controlling its copy number. This inhibitor-target regulatory mechanism involves two ctRNA species (having the same 5'-end but different 3'-ends) which are complementary to the untranslated 5'-end of the *repC* mRNA (Kumar and Novick, 1985; Novick, 1987; del Solar *et al.*, 1998). The initial kissing-complex between the ctRNAs and the mRNA leads to a conformational change in the *repC* mRNA transcript distal to the region of complementarity. This conformational change in the mRNA causes a new hairpin structure to form. As the new hairpin structure ends in an A(U)<sub>6</sub> sequence that resembles a Rho-independent transcription terminator, it truncates the *repC* mRNA molecule rendering it non-functional (Novick *et al.*, 1989; del Solar *et al.*, 1998).

The pMV158 plasmid regulates its copy number by a dual system in which it uses both a transcriptional inhibitor and a ctRNA molecule, designated RNA II (del Solar, 1998, Espinosa *et al.*, 2000). This is an example of a copy number control mechanism in which the transcriptional repressor plays a direct role in the regulatory mechanism. The *copG* and *repB* genes are co-transcribed on a bicistronic mRNA molecule regulated by a single promoter [ $P_{cr}$ ] (Fig 1.11). The very stable homodimeric CopG protein represses transcription of this bicistronic mRNA by binding to the  $P_{cr}$  promoter, hindering the binding of RNAP and thus regulating its own synthesis and that of *repB* (del Solar and Espinosa, 2000).



**Fig 1.11:** pMV158 copy number control mechanism. CopG repressor protein blocks transcription from *Pcr* promoter which directs synthesis of *copG-repB* mRNA. The RNA II ctRNA molecule transcribed from the *Pctl* promoter blocks translation of RepB. Reproduced from del Solar and Espinosa (2000).

The current model of CopG repression involves the cooperative binding of four CopG dimers to an inverted repeat that encompasses the  $-35$  box of the *P<sub>cr</sub>* promoter. A ribbon-helix-helix motif is involved in CopG dimerization, and in specific interactions with the DNA bases in the operator. RNA II binds to a complementary sequence between the *copG* and the *repB* coding sequences on the mRNA inhibiting the translation of the *repB* gene by direct interaction with the ribosomal binding sites (RBS, Shine-Dalgarno) present in the *repB* mRNA coding sequence (Espinosa *et al.*, 2000; del Solar *et al.*, 2002).

The copy number control mechanism of pC194 appears to be regulated only by a ctRNA species (Espinosa *et al.*, 2000).

### 1.3.2) Theta replication

The theta replication mechanism derives its name from the appearance of the replicating plasmid molecule. Electron microscopy studies of these plasmids utilizing this replication mechanism identified a characteristic theta ( $\theta$ )-shape during replication (del Solar *et al.*, 1998).

#### 1.3.2.1) General replication mechanism

Defining features of this plasmid replication mechanism include the use of a plasmid encoded replication initiator protein (Rep) that acts on a series of direct repeats (iterons) found in one or several plasmid origins (*ori*). Through a series of DNA-protein and protein-protein interactions the Rep proteins recruit other initiation factors to the origin which assemble into the initiation nucleoprotein complex (Giraldo *et al.*,



1998). The consequent initiation of replication causes helix distortion which precedes localized melting of the origin parental strands allowing the nucleoprotein complex to synthesize a pRNA. Subsequently, covalent extension of the pRNA by the replisome results in daughter strand synthesis. DNA synthesis is continuous on the leading strand, discontinuous on the lagging strand, and can proceed either uni- or bidirectionally. Alternatively, some replicons require DNA Pol I to assist in the early stages of leading-strand synthesis (del Solar *et al.*, 1998, Espinosa *et al.*, 2000).

A general mechanism is difficult to define as the ColE1-type replicons utilize another form of theta replication. For these replicons, replication commences with transcription across the origin, followed by the formation of a RNA:DNA hybrid. The RNA transcript is processed to generate the pRNA from which DNA Pol I polymerization initially extends the leading strand until a point is reached where DNA Pol III takes over and completes strand synthesis (Espinosa *et al.*, 2000). This mechanism is referred to as a DNA Pol I-dependent mechanism.

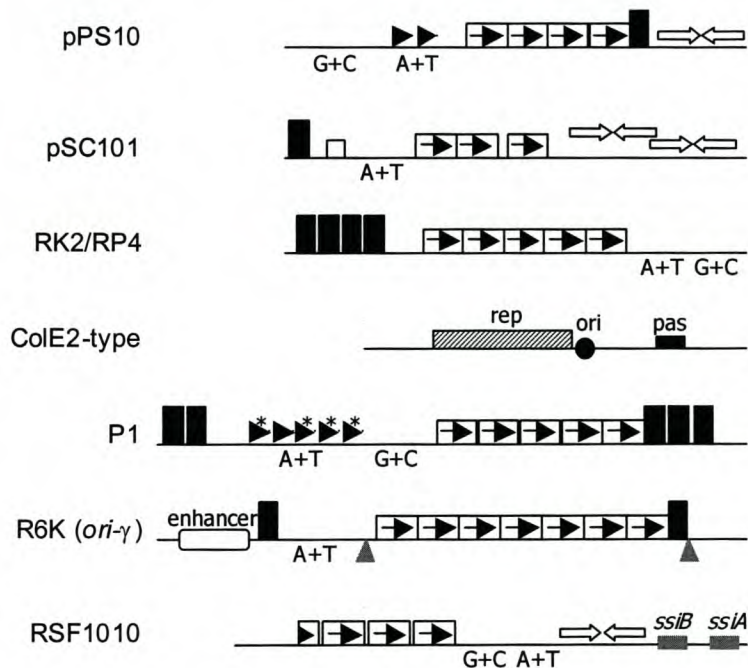
Host replication factors are important components of theta-type replication. Discussed above is the requirement for DNA Pol I by some plasmid replication mechanisms. In some mechanisms the Rep proteins interact directly with DnaA, and DnaB (Konieczny and Helinski, 1997[a]; Espinosa *et al.*, 2000). DnaA appears to be an important host-encoded factor for most theta-type replication mechanisms. DnaA assists plasmid origin open complex formation, probably as a consequence of the inability of Rep proteins to bind ATP (Giraldo and Díaz-Orejas, 2001). DNA gyrase most certainly plays a role to overcome topological constraints, while Topoisomerase IV is responsible for the decatenation of the two DNA daughter molecules generated at the end of a replication cycle (Espinosa *et al.*, 2000).

Some plasmids, such as the DNA Pol I-dependent ColE1 plasmids, depend entirely on host-encoded replication proteins, while others, like RSF1010, have minimal host factor requirements.

### 1.3.2.2) Origin comparisons

The origin of replication for theta replicating plasmids can be divided into iteron-containing origins (e.g. plasmids: pPS10, pSC101, RK2/RP4, ColE2, ColE3, P1, R6K, and RSF1010) and origins lacking iterons (e.g. plasmids: R1, ColE1, and pLS20). Analysis of iterons found at the origin has revealed that, in general, these repeats tend to be multiples of eleven. This is close to the helical periodicity of the DNA double helix, and it is believed that this may align the Rep proteins, bound to the iterons, on the same face of the DNA helix (del Solar *et al.*, 1998). Auxiliary iterons found outside the origin of some plasmids are usually not required for replication, but assist in the control of replication. Iterons of a particular origin need not be identical, although a consensus sequence can usually be defined for the particular set of iterons.

Iteron-containing origins have a range of organizational structures (Fig 1.12).



**Figure 1.12:** Structure of theta replicating plasmids with iteron-containing origins of replication. Boxed arrows correspond to direct repeats to which Rep proteins bind, with G+C and A+T-rich regions indicated. Features shown are: DnaA-boxes (solid rectangles), inverted repeats (open arrows pointing facing each other), repeats in A+T-rich regions (solid arrow heads), IHF-binding sites (open rectangles), *dam* methylation sites (stars), and FIS-binding sites (grey triangles). The primosome assembly site (PAS) for ColE2-type origins, and the single-stranded initiation sites (*ssiA* and *ssiB*) for RSF1010 are indicated. Note, for RSF1010 it has three and a half iterons with a 2-bp spacer between each iteron. Adapted from del Solar *et al.* (1998) and Rawlings and Tietze (2001).



The narrow host range plasmid pPS10, isolated from *Pseudomonas syringae* pv. *savastanoi*, has an origin containing four consecutive perfectly conserved 22-bp iterons to which its Rep protein binds (Nieto *et al.*, 1992). These iterons are flanked by an A+T-rich sequence and a DnaA-box. Next to the DnaA-box lies two inversely repeated sequences of 8-bp, designated S1 and S2, that flank the -35 region of the *rep* gene promoter. This inverted repeat forms an operator to which the Rep protein binds to autoregulate its own transcription. The inverted repeats share homology with the internal sequences of the origin iterons (García de Viedma *et al.*, 1995).

The origin of replication of pSC101 contains three iterons of approximately 20-bp, an A+T-rich region preceding an IHF binding site, and a single DnaA-box (Yamaguchi and Yamaguchi, 1984; Stenzel *et al.*, 1991). An A+T-rich region adjacent to the iterons contains putative DnaBC helicase complex entry sites (Espinosa *et al.*, 2000).

Plasmid RK2/RP4 is a 60-kb broad host range conjugative plasmid of the incompatibility group IncP $\alpha$ . The five iterons of 17-bp (with a 4-6-bp iteron spacer) found in the minimal origin of RK2/RP4 are flanked by four DnaA-boxes, an A+T-rich region, and a G+C-rich region (Perri and Helinski, 1993). The four DnaA boxes are arranged as two pairs in an indirect orientation that can potentially form a cruciform structure (Doran *et al.*, 1999[a]). Three of the DnaA-boxes contain one mismatch from the *E. coli* DnaA box consensus sequence, and the fourth box contains two mismatches. The DnaA-box proximal to the iterons appears to be the most critical. Deletion of three of the DnaA-boxes, but not the iteron-proximal DnaA-box, results in severe impairment of replication in *E. coli*. However, deletion of all four DnaA-boxes results in complete loss of replication in *E. coli*, and reduces the level of stable maintenance of the plasmid in *Pseudomonas aeruginosa* (Doran *et al.*, 1999[b]). When only the iteron-proximal DnaA-box was mutated, replication activity was abolished in *E. coli* (Doran *et al.*, 1999[a]).

The smallest known origins described so far are found in the ColE2 and ColE3 families. These origins contain two directly repeated sequences which are 47-bp and 33-bp long, respectively (del Solar *et al.*, 1998).



P1 is a bacteriophage that can also be propagated as a low copy number plasmid in *E. coli* (Abeles *et al.*, 1984). At the *IncC* locus, the P1 origin contains five iterons which are flanked by two tandem DnaA-boxes on one side and three tandem DnaA-boxes on the other. Unlike R6K (see below), all five iterons are required for replication *in vivo*. Three highly conserved sequence patches are found within the iterons of this plasmid. A G+C region, and an A+T-rich region containing five *dam* methylation sites that appear to modulate origin activity, precede the iterons (Abeles *et al.*, 1984; del Solar *et al.*, 1998; Espinosa *et al.*, 2000).

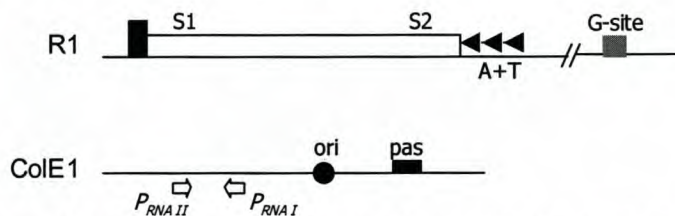
All the above plasmids contain a single origin, but plasmids containing more than one origin have been known for some time. Plasmid R6K has three different origins of replication. The *ori-γ* is the regulatory origin from which replication is initiated, and it contains seven iterons together with a *cis*-acting enhancer and an A+T-rich region. Binding sites for DnaA, IHF, FIS, HU and RNAP are found in this A+T-rich region, while the enhancer region contains an inverted repeat, a DnaA-box and a stability locus, *stb*, which confers stable inheritance on R6K. Conversely, *ori-α* contains only one complete iteron, and *ori-β* has half an iteron (Filutowicz and Rakowski, 1998). Deletion studies have shown that removal of one iteron from *ori-γ* has no effect on plasmid replication, but removal of two reduces the efficiency of replication, while deletion of three or more completely abolishes replication (del Solar *et al.*, 1998; Espinosa *et al.*, 2000). This indicates that *ori-γ* is essential for replication and that the other origins cannot compensate for the loss of this origin.

Plasmids with multiple origins are also found in the IncQ family of plasmids. Plasmid pIE1107 contains two origins designated *oriVa* and *oriVb* (Tietze, 1998; Rawlings and Tietze, 2001). Both of the origins of pIE1107 contain a G+C-rich and A+T-rich region, although *oriVa* contains three and a half iterons, while *oriVb* has four iterons. However, only one origin is active for replication. Deletion of *oriVb* rendered the plasmid unable to replicate, but as pIE1107 was still able to replicate with *oriVa* deleted, this indicated that *oriVa* was non-functional for replication (Tietze, 1998; Rawlings and Tietze, 2001). The prototype plasmid of the IncQ family of plasmids is RSF1010. This plasmid, however, contains only one origin having



three and a half iterons of 20-bp with a 2-bp spacer, and a G+C-rich and A+T-rich region. The origin also has template specific single-stranded initiation sites, designated *ssiA* and *ssiB*, upon which the plasmid-encoded primase acts (Honda *et al.*, 1989; Rawlings and Tietze, 2001).

The plasmid resistance factor R1 (Fig 1.13) and prototype of the IncFII family of plasmids has an origin, *oriR*, that is an example of an origin that does not contain iterons. Instead, *oriR* has a DnaA-box, a contiguous 100-bp region where the Rep protein interacts, and an A+T-rich region containing a putative DnaB-binding site. The contiguous 100-bp region at which the Rep protein interacts contains two imperfect palindromes, designated S1 and S2, but the Rep protein preferentially binds to the S1 site. It has been shown that the DnaA-box is dispensable, but optimizes the efficiency of replication initiation (del Solar *et al.*, 1998; Espinosa *et al.*, 2000).



**Figure 1.13:** Structure of theta replicating plasmids with origins of replication that do not contain iterons. Features shown for R1 are: DnaA-boxes (solid rectangles), repeats in A+T-rich regions (solid arrow heads), the DnaG-type priming signal (G-site), and the imperfect palindromes (S1 and S2). The primosome assembly site (PAS) for ColE1-type origins is shown, including the promoters (open arrows) for RNA I and RNA II. Reproduced from del Solar *et al.* (1998).

A plasmid which does not have a replication initiator protein nor iterons is the ColE1 plasmid. This plasmid does, however, have a replication requirement for DNA Pol I. The ColE1 origin spans a region of about 1-kb and encodes two RNA molecules. RNA I is a ctRNA molecule active in copy number control, while RNA II is an RNA primer used for leading strand synthesis. The origin has sequences complementary to RNA II, promoting the formation of a stable RNA II/DNA hybrid, and it also includes sequences that favour specific processing of this coupled complex by RNase H (see section 1.3.2.4). It has a primosome assembly site, which together with an adjacent DnaA-box, allows for the DnaA-dependent DnaBC helicase complex assembly for discontinuous lagging strand synthesis. Finally, the origin also contains a termination site for lagging strand synthesis, *terH*, which determines its unidirectional replication mode (del Solar *et al.*, 1998).



There are even plasmids such as pLS20 in which the origin is only composed of several palindromes flanking a DnaA-box. This plasmid's replication mechanism is independent of both a replication initiator protein, and DNA Pol I (del Solar *et al.*, 1998).

### 1.3.2.3) *Replication initiator protein comparisons*

The three functional domains of the 26.7 kDa RepA protein of pPS10 that have been identified include a LZ motif, and two globular domains capable of binding specific DNA sequences within the iterons and the operator (García de Viedma *et al.*, 1995; García de Viedma *et al.*, 1996). Mutational analysis of the LZ motif determined that this motif was not directly involved in binding of RepA to DNA, but is a protein-protein interaction interface by which it regulates RepA dimerization and probably promotes interaction between DnaA and host replication factors (García de Viedma *et al.*, 1996; del Solar *et al.*, 1998). Heterodimerization assays confirmed that RepA contacts the iterons in the origin and initiates replication as a monomer, whereas the dimers bind to the inverted repeats at the operator and repress transcription (García de Viedma *et al.*, 1996). Apart from the LZ motif, the RepA monomer has two globular domains which have secondary structure elements that resemble a winged-helix domain. A winged-helix domain recognizes DNA using a HTH motif in a three helix bundle, and a strand-loop-strand 'wing' (Giraldo *et al.*, 1998; Giraldo and Díaz-Orejas, 2001). The HTH motif is characteristic of proteins which bind DNA (Brennan and Matthews, 1989). The N-terminal LZ motif stabilizes RepA dimers, but DnaK-mediated dissociation of dimers results in the monomer undergoing a conformational change to an extended form which allows both winged-helix domains of the RepA monomer to interact with DNA. Conversely, only one winged-helix domain of each oligomer of the dimer contacts the inverted repeats at the operator by the HTH motif (Giraldo *et al.*, 1998; Giraldo and Díaz-Orejas, 2001).

Like pPS10, the RepA of pSC101 and P1 both exist in a monomer-dimer equilibrium (Ingmer *et al.*, 1995; Espinosa *et al.*, 2000). As a monomer the RepA of pSC101 binds to the iterons in the origin and initiates replication, while the dimer autoregulates RepA transcription by binding to the inverted repeat in the *repA* promoter repressing transcription (Manen *et al.*, 1992). Sequence analysis has



revealed that the pSC101 RepA shares homology at the N- and C-terminal ends with the pPS10 RepA where the LZ motif and HTH motif have been identified (Espinosa *et al.*, 2000).

The R6K  $\pi$  protein, encoded by the *pir* gene, also contains a LZ-like motif. Mutation of the LZ-like motif resulted in  $\pi$  protein that was unable to activate *ori- $\alpha$*  or *ori- $\beta$*  (del Solar *et al.*, 1998; Espinosa *et al.*, 2000). The *pir* gene encodes two  $\pi$  polypeptides which differ in size: a 35.0 kDa species ( $\pi^{35.0}$ ), and a smaller 30.5 kDa ( $\pi^{30.5}$ ) species expressed at a lower level. The  $\pi^{35.0}$  species is active for replication initiation, while the  $\pi^{30.5}$  appears to act as an inhibitor (Filutowicz and Rakowski, 1998). The functional difference between these species appears to be controlled by a LZ motif present in the larger  $\pi^{35.0}$  species, but absent in the  $\pi^{30.5}$  protein. Although both species form dimers, the stability of these dimers differ. The LZ motif present in  $\pi^{35.0}$  allows for the formation of stable dimers, whereas dimers formed by  $\pi^{30.5}$  are unstable (Filutowicz and Rakowski, 1998). Both monomers and dimers can bind to an iteron, but only dimers can bind to the non-iteron site that resides in the A+T-rich region. It is the monomeric form that activates replication at the iterons (Krüger and Filutowicz, 2000). The C-terminal of the  $\pi$  protein monomer has been identified as the DNA binding domain, while the N-terminal domain contains at least two dimerization surfaces (Levchenko *et al.*, 1994; Urh *et al.*, 1998).

The replication initiation protein TrfA of RK2/RP4 exists in two oligomeric forms (TrfA-44 [TrfA1] and TrfA-33 [TrfA2]) that also differ in size, but both forms bind the iterons as a monomer (Toukdarian *et al.*, 1996). The two oligomeric forms of this protein are due to the two translational start sites within the *trfA* gene (Pansegrau *et al.*, 1994). Both of these translation products are able to initiate replication at the RK2/RP4 origin in *E. coli*, except that the smaller TrfA-33 protein is unable to initiate efficient replication in *P. aeruginosa* (Thomas *et al.*, 1981; Espinosa *et al.*, 2000).

The RepC protein of RSF1010 is also thought to be in a monomer-dimer equilibrium. The DnaK-stimulated binding of RepC to the three and a half iterons suggests that RepC is active as a monomer (Sakai and Komano, 1996).



The 35 kDa ColE2 Rep protein, in comparison to the other Rep proteins that interact with the origin, has the added ability to synthesize a pRNA at a specific *ssi*-site. Once the pRNA is synthesized, DNA Pol I acts to initiate leading strand synthesis. Thus this Rep protein possesses primase activity by which it synthesizes a three ribonucleotide primer of the sequence 5'-ppApGpA-3' (Takechi *et al.*, 1995). The C-terminal domain of this Rep protein also has a HTH motif probably used for DNA-binding (del Solar *et al.*, 1998; Espinosa *et al.*, 2000).

#### 1.3.2.4) *Replication mechanism comparisons*

The replication initiating proteins of the plasmids discussed in section 1.3.2.3 exist in a monomer-dimer equilibrium, with the monomer form being active for replication initiation. Chatteraj (2000) speculates that dimerization may be a mechanism by which initiator protein availability is controlled, thereby delaying replication initiation. It appears that the molecular chaperone proteins DnaK (ATP-dependent binding of polypeptide substrates), DnaJ (tags proteins for recognition by DnaK), GrpE (nucleotide exchange factor) or ClpA mediate this equilibrium, thereby participating in the Rep-type replication initiation pathways (McCarty *et al.*, 1995). Many chaperone and Clp proteins are heat-shock proteins that are active in the refolding and remodeling of protein structures, the degradation of irreversibly denatured proteins, as well as dissolution of protein aggregates and reactivation of the active-species (Pak and Wickner, 1997; Kim *et al.*, 2002). When ClpA and ClpX are associated with the catalytic protease component ClpP, they bring about protein degradation. However, in the absence of ClpP they display chaperone activities.

The replication initiation of plasmid P1 requires the monomerization and activation of its RepA protein by DnaK, DnaJ and GrpE chaperones (Wickner *et al.*, 1992). Further *in vitro* studies on chaperone activation of P1 RepA showed that the ATP-dependent ClpA can substitute for DnaK and DnaJ to activate RepA, but that ClpX had no activation ability on RepA (Pak and Wickner, 1997). Although ClpX is able to convert inactive RK2/RP4 TrfA dimers to active monomers, monomerization and activation of the TrfA dimers is most efficiently brought about by the synergistic action of ClpB, DnaK, DnaJ and GrpE chaperones (Konieczny and Helinski, 1997[b]; Konieczny and Liberek, 2002). The mechanism of activation of monomers is unclear,



but recent studies have demonstrated that DnaJ and DnaK have distinct binding sites on P1 RepA (Kim *et al.*, 2002).

Whereas bacterial chromosomal replication requires the single evolutionarily conserved DnaA protein to initiate replication, many plasmids require both host-encoded DnaA and the plasmid-encoded Rep-type initiator proteins to act as dual plasmid replication initiators (Sharma *et al.*, 2001). The concerted participation of DnaA and plasmid-encoded Rep proteins to initiate replication has been described for many plasmids, but the exact role of DnaA in open complex formation depends on each replicon.

Investigations of plasmid replicon replication have revealed a range of dependency on host encoded replication factors, and the exact participation of these factors is still being determined. For instance, plasmid pPS10 has shown an *in vitro* and *in vivo* requirement for DnaK, DnaA and the DnaBC helicase complex for the initiation of replication. Recent studies suggest that DnaA entry into the pPS10 initiation complex is mediated by RepA-DnaA interactions (Fernández-Tresguerres *et al.*, 1995; Maestro *et al.*, 2002).

Plasmid P1 requires the RepA protein together with DnaA, HU and IHF to bring about replication initiation. The presence of DnaA boxes on both ends of the origin helps localize the unwinding of the A+T-rich region (Park and Chattoraj, 2001). Through direct molecular interaction the DnaA loads the DnaB helicase into the open complex (Seitz *et al.*, 2000). Once the nucleoprotein complex has assembled, bidirectional replication completes a new round of plasmid synthesis (Park and Chattoraj, 2001). It has been demonstrated in the P1 mechanism that DnaA protein acting in the absence of RepA can induce open complex formation (Mukhopadhyay *et al.*, 1993; Park *et al.*, 1998).

Contrary to this, binding of DnaA to its four DnaA-boxes in the RK2/RP4 origin cannot create an open complex on its own, but together with TrfA enhances strand opening. The TrfA protein recognizes the iterons of the origin and together with HU and DnaA produces an opening specifically at a set of four 13-mers located in the A+T-rich region (Konieczny *et al.*, 1997). The two DnaA-boxes proximal to the set



of iterons are critical for the initiation of RK2/RP4 replication in *E. coli*. This clearly demonstrates the reliance of RK2/RP4 on the *E. coli* DnaA protein for replication initiation in this bacterium. In *E. coli*, DnaA is critical in delivering the DnaBC helicase complex to the DnaA-box region of the RK2/RP4 origin. Despite the interaction of DnaA with the DnaBC helicase complex, TrfA is required to bind to the DnaABC complex for both loading and activation of DnaB at the open complex (Konieczny and Helinski, 1997[a]; Pacek *et al.*, 2001). The binding of DnaA to these two DnaA boxes occurs in a cooperative manner in *E. coli*, with the DnaA-box closest to the set of iterons coordinating the binding of DnaA to the other three DnaA-boxes (Doran *et al.*, 1999[a]). However, non-cooperative binding of the *Pseudomonas* DnaA ortholog to the RK2/RP4 origin DnaA-boxes has been demonstrated in *Pseudomonas putida* and *P. aeruginosa*. Replication of RK2/RP4 in *P. aeruginosa* is independent of the DnaA-boxes suggesting an alternate mechanism of replication initiation in this species (Caspi *et al.*, 2000). Caspi *et al.* (2000) suggest that as *P. aeruginosa* has a preference for TrfA-44, the RK2/RP4 plasmid produces two Rep proteins as a means to overcome a species-dependent preference for a particular oligomer at later stages in replication; such as for the loading and activation of the helicase.

The DnaA protein is strictly required for the replication of plasmid pSC101 (Sharma *et al.*, 2001). The loading of the DnaBC helicase complex onto the pSC101 origin is a RepA-dependent system that together with DnaA and IHF induces open complex formation (Datta *et al.*, 1999; Seitz *et al.*, 2000). The model proposed by Sharma *et al.* (2001) for pSC101 initiation of replication involves the initial binding of IHF to its site on the origin, which then promotes and stabilizes the binding of DnaA to the single DnaA-box. IHF also promotes the physical interaction between DnaA and RepA bound to the iterons. These factors, acting in concert, induce directed melting of the origin. It is unclear as to whether RepA alone, or a RepA-DnaA complex, recruits the DnaBC helicase complex to the origin, but direct interaction between RepA and DnaB has been demonstrated using a yeast forward and reverse two hybrid system (Datta *et al.*, 1999; Sharma *et al.*, 2001). Since ADP-DnaA, inactive for *oriC* replication initiation, is able to support R1, P1, RK2/RP4, pSC101 and R6K replication, it is believed that only the oligomerization (domain I) and the DNA-



binding domain (domain IV) of DnaA are required to initiate replication of these plasmids (Doran *et al.*, 1999[a]; Sharma *et al.*, 2001).

The R6K  $\pi$  protein acts on all three origins, but the requirements for activation differs between the origins. For instance, DnaA is required for activation of *ori- $\gamma$*  and *ori- $\alpha$* , but not *ori- $\beta$*  (Kelley *et al.*, 1992). The  $\pi$  protein also preferentially binds *ori- $\gamma$* , but the activation of *ori- $\alpha$*  and *ori- $\beta$*  requires the enhancer which appears to transfer the  $\pi$  protein to these two origins (del Solar *et al.*, 1998). Replication initiation at *ori- $\gamma$*  involves not only the two regulatory forms of  $\pi$  protein, but binding of IHF, DnaA, FIS and RNAP to their respective sites. It is, however, not known whether strand opening is achieved by  $\pi$  protein binding to the iterons in *ori- $\gamma$* , or to the A+T-rich region, or both (Filutowicz and Rakowski, 1998). The  $\pi$  protein regulates R6K replication on a number of levels. Firstly, two oligomeric forms of the  $\pi$  protein regulate replication. The  $\pi^{30.5}$  oligomer is a regulatory inhibitor, while the  $\pi^{35.0}$  oligomer stimulates replication (Wu *et al.*, 1997). Secondly, the  $\pi$  protein autoregulates itself by binding as a dimer to the inverted repeats that make up the *pir* gene operator. Finally, replication can also be regulated by the  $\pi$  protein binding to the iterons as either a monomer or dimer (Urh *et al.*, 1998). The oligomerization model of  $\pi$  protein activity proposed by Filutowicz and Rakowski (1998) suggested that the monomer binds to the iterons in a head-to-tail association which is active for replication. However, dimeric binding of  $\pi$  to the iterons in a sideways association inhibits replication through a ‘handcuffing’ mechanism (see section 1.3.2.5). A specific head-to-head binding of dimeric  $\pi$  protein to the inverted repeat in the operator is believed to repress transcription of  $\pi$  protein, thereby regulating replication.

The replication initiation model proposed by Filutowicz and Rakowski (1998) suggests that  $\pi^{30.5}$  protein, whether as a  $\pi^{30.5}$ - $\pi^{35.0}$  heterodimer or by competition with  $\pi^{35.0}$  for the direct repeats, inhibits replication. The binding of IHF to its site, *ihf1*, is thought to result in folding of the *ori- $\gamma$* , and this folding is believed to reverse replication inhibition by blocking access of the inhibitory- $\pi$  protein to the direct repeats. This folding of *ori- $\gamma$*  is also believed to permit interactions between proteins



bound at the enhancer and the *ori-γ* core. Such DnaA-DnaA and DnaA- $\pi$  protein interactions would then relieve initiation constraints (Wu *et al.*, 1994; Dellis *et al.*, 1996). The DnaA- $\pi$  protein complex may be responsible for recruiting the DnaBC helicase complex as  $\pi$  protein and DnaB have been shown to directly interact (Ratnakar *et al.*, 1996). The inability of  $\pi^{30.5}$  to stimulate replication could be explained by this oligomer lacking the LZ motif in the truncated N-terminal domain. This could render  $\pi^{30.5}$  unable to interact with DnaB, and thus unable to initiate primosome assembly (Wu *et al.*, 1997). An alternative argument is that  $\pi^{30.5}$  dimerizes with  $\pi^{35.0}$  to form the  $\pi^{30.5}$ - $\pi^{35.0}$  heterodimer which in turn inhibits priming for leading strand synthesis by binding to a non-iteron site that can only be bound by dimers, and also happens to be located in the A+T-rich region where DnaG priming occurs in the *ori-γ* (Krüger and Filutowicz, 2000).

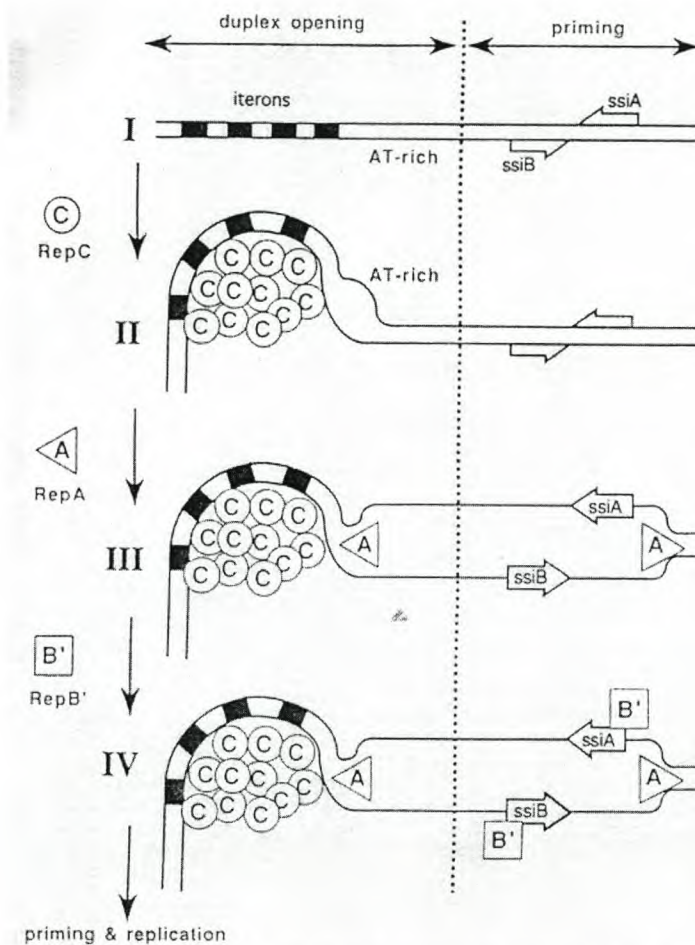
Initiation of R1 replication begins with RepA dimers recognizing the cores of the *oriR* S1 and S2 partially palindromic sequences. Subsequently, DnaK-mediated formation of monomers results in RepA monomers which bind the intermediate region of *oriR*. Following this, DnaA binds to its DnaA-box (adjacent to the RepA-binding region), but this binding requires interaction with the RepA protein. If, however, the DnaA-box is removed from the origin, it appears that functional interaction between DnaA and RepA is sufficient to promote replication (del Solar *et al.*, 1998). Initiation of leading strand synthesis occurs at the G-site located 400-bp downstream of *oriR*, and is believed to be activated when lagging strand synthesis, which initiates in the A+T-rich region, reaches the G-site. The G-site resembles the bacteriophage G4 origin for complementary strand synthesis. In the case of R1, this G4-like site is primed by DnaG primase alone. It appears R1 G-site priming requires only DnaG, SSB protein and DNA Pol III, while DnaB, DnaC and DnaT are not required (Masai and Arai, 1989).

The origin of RSF1010 contains two small palindromic *ssi*-sites located on opposite strands, designated *ssiA* and *ssiB*, which are two independent *cis*-acting elements that are primed by the RSF1010-encoded primase. The primase initiates continuous leading strand synthesis in opposite orientations (Honda *et al.*, 1989; Sakai and Komano, 1996). The replication initiating protein of RSF1010, RepC, is capable of



initiating replication independently of host replication factors such as DnaA, DnaB, DnaC, and DnaT (Scherzinger *et al.*, 1991). In other words, the RSF1010 replicon encodes a plasmid-template specific helicase (RepA) and primase (RepB) which, together with RepC, renders this plasmid independent of most host replication factors which might pose a constraint on this plasmid's replication ability in a range of species. The replication mechanism of this distinctive replisome complex displaces the non-replicated parental strand as a D-loop. Leading strand synthesis in opposite directions results in a double-stranded theta-shaped structure. As this replication mechanism occurs without concomitant lagging strand synthesis, the replication mechanism is seen as unique amongst the theta-replicating plasmids, and is termed the strand displacement mechanism (Sakai and Komano, 1996; del Solar *et al.*, 1998).

The strand displacement mechanism of RSF1010 can proceed either uni- or bidirectionally after initiation (Sakai and Komano, 1996). With reference to Fig 1.14, initiation of replication occurs through the binding of the RepC to the iterons causing a localized melting in the A+T-rich region allowing for the entry of the plasmid-encoded RepA helicase (Haring *et al.*, 1985; Kim and Meyer, 1991). The 5'→3' translocation of the hexameric RepA helicase along the *l*-strand (the strand which is in correct orientation for mRNA transcription of 10 of the 11 known RSF1010 proteins) exposes and activates the two *ssi*-sites on each opposing strand (Niedenzu *et al.*, 2001; del Solar *et al.*, 1998).



**Figure 1.14:** Strand displacement replication mechanism of RSF1010. The replication initiating protein RepC binds to the iterons in the origin resulting in localized melting of the A+T-rich region and strand opening. The RSF1010 RepA helicase is then recruited to the open complex and unwinds the double helix. Translocation of the RepA helicase exposes the *ssi* sites that are then primed by the RSF1010 RepB' primase. Replication can be initiated independently from either of these primed *ssi* sites, or from both of them simultaneously. Replication progresses continuously and in opposite directions. The non-replicated parental strand is displaced during the replication process. Reproduced from Sakai and Komano (1996).

The *ssi*-sites are highly conserved 40-bp sequences which form a stem-loop structure. The structure and dimensions of the stem-loop are critical for *ssi* activity as base composition and base-pairing rather than nucleotide sequence are the important characteristics of the stem-loop. The trinucleotide 5'-GTG-3' in the 3'-flanking region of the stem is a possible start site for DNA synthesis (Honda *et al.*, 1993; Miao *et al.*, 1993; Miao *et al.*, 1995[a]). The plasmid-encoded RepBAC proteins effectively negate the requirements for host replication proteins DnaA, DnaB, DnaC, DnaG and RNAP, although the RSF1010 replication mechanism is still dependent on DNA Pol III HE and DNA gyrase (Sakai and Komano, 1996).



### 1.3.2.5) *Copy number control*

Plasmid copy number is regulated by the opposing action of positive effectors and negative effectors of replication. The Rep protein and DnaA are positive effectors of replication for many plasmids utilizing theta-replication. However, these are not the only initiators of replication, as the ColE1 family of plasmids use an RNA species as a positive effector, while a ctRNA and a regulatory protein act as the negative regulators of replication (see below). The mechanism of negative regulation of replication is varied amongst plasmids. Plasmids do not necessarily utilize a single negative regulator, but may use combinations of ctRNA, regulatory proteins, autoregulation of their positive initiator, and iterons to control replication.

Therefore, there are many levels of replication regulation available to a plasmid. Autoregulation of the positive initiator is an important mechanism by which plasmids maintain an optimal concentration of Rep protein independent of copy number (Espinosa *et al.*, 2000). An independently regulated Rep protein concentration allows for increases in Rep concentration to be immediately counteracted by a negative regulatory loop, thereby preventing uncontrolled replication initiation, and possible run-away replication.

Two forms of autoregulation are recognized. In the first form of autoregulation, utilized by plasmids such as pPS10, pSC101 and R6K, the operator(s)/promoter(s) and origin of replication are independently located. However, in the second form, exemplified by P1, the operator(s)/promoter(s) are buried in the iterons meaning that the Rep monomer performs both replication initiation and autoregulation (Park *et al.*, 1998). With the P1 *repA* promoter buried in the iterons, the RepA bound to the iterons excludes RNAP from the promoter. Therefore, as the replication fork crosses the origin iterons it 'cleans' the promoter of bound RepA allowing RNAP access to the promoter. Consequently, there is a transient burst of *repA* transcription until transcription is repressed again by binding of RepA to the iteron region (Mukhopadhyay and Chattoraj, 2000).

Another level of regulation is the active monomer/inactive dimer equilibrium mediated by chaperones. It is possible that dimerization may limit Rep availability



and thereby delay replication (Chattoraj, 2000), although this does not seem to be the case for R6K and R1. Some dimers bind auxiliary iterons found outside the origin. Examples of plasmids which have auxiliary iterons include F, R6K and P1 (*incA* locus). Although not active for replication initiation, these auxiliary iterons may serve to titrate the available Rep protein, thereby providing a more refined control mechanism by reducing copy number deviations from the copy average (Chattoraj, 2000).

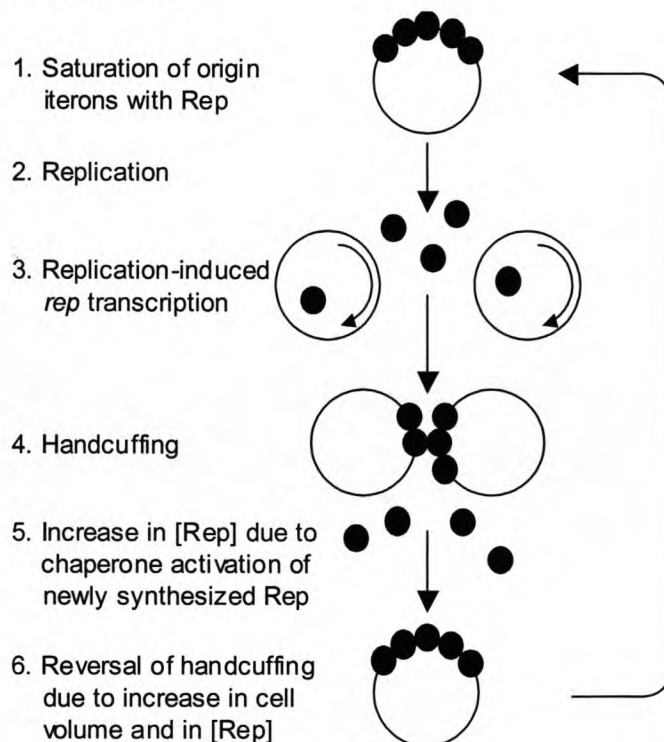
For iteron-containing plasmids, the model of replication initiation dictates that the Rep protein binds to the iterons until a saturating level is reached that causes strand melting and open complex formation. Exactly how iterons act as negative effectors has been a topic of debate for many years as the mechanism by which they negatively regulate replication has not been conclusively demonstrated. For instance, confusion over how the copy number of RK2/RP4 and R6K is maintained has resulted from studies revealing that the apparent concentration of Rep protein was too high to be iteron saturation limiting (del Solar *et al.*, 1998). This has led to the proposal of two models of iteron-mediated plasmid copy number control.

The ‘titration model’ was the first model proposed for iteron-mediated replication control, but is currently not favoured. This model assumed that Rep protein was rate limiting, and that the Rep protein was titrated by the iterons until a certain level of Rep-iteron interaction had been reached causing another round of replication to be initiated (Chattoraj, 2000). Therefore, this sequestration model suggests that replication frequency is a function of Rep protein concentration, and it assumes that Rep protein expression is constitutive. The model appears to hold true as copy numbers of iteron-containing plasmids decreased when the iteron concentration was increased *in cis* or *in trans* (Chattoraj, 2000). However, it fails to explain why a 200-fold excess of TrfA protein resulted in only a 30 % increase in RK2/RP4 copy number, or why a two fold decrease in  $\pi$  protein concentration had no significant effect on R6K copy number (Durland and Helinski, 1990; Filutowicz and Rakowski, 1998). Similarly, over-expressing the RepA protein beyond physiological levels was found to have very little effect on pPS10 and P1 (when *incA* was present) copy number (Pal and Chattoraj, 1988; García de Viedma *et al.*, 1996). Added to this, the



Rep proteins of iteron-containing plasmids are autoregulated, or under transcriptional control, yet titration would negate the need for such mechanisms (Nordström, 1990). These systems are difficult to merge with the titration model (del Solar *et al.*, 1998; Chattoraj, 2000).

The second model for copy number control for iteron-containing plasmids is the ‘handcuffing’ model (Fig 1.15).



**Figure 1.15:** Handcuffing model for iteron-mediated control of copy number. Reproduced from Chattoraj (2000).

This model incorporates the three levels of replication control. The model proposes that Rep protein sequestration to the iterons in an origin results in a saturation level being achieved that promotes strand opening, and initiation replication. Replication induces *rep* transcription which increases the cellular concentration of Rep. The newly synthesized Rep proteins readily dimerize, and the pre-existing active Rep protein monomers are titrated by the newly formed plasmid’s origin. The iteron-bound Rep protein monomers on both daughter plasmids enter an associate-dissociation equilibrium, probably through a specific protein handcuffing domain in the monomer, and this protein-protein interaction results in plasmid “handcuffing” which blocks a new round of replication. Origin inactivation is explained as a consequence of origin pairing (handcuffing) which results in steric hindrance to open

complex formation. Thus, this handcuffing mechanism would prevent a new round of replication, and represses *rep* operator(s)/promoter(s) (Chattoraj, 2000).

There are a number of means, probably acting in concert, by which handcuffing can be reversed. An increase in cell volume prior to chromosome segregation may upset the protein-protein equilibrium of iteron-bound monomers resulting in prolonged periods of intramolecular dissociation. Furthermore, the frequency of chaperone-mediated dimer dissociation is a rate determinant of cellular Rep monomer concentration that, together with a possible conformational change to initiation-active monomer, regulates replication initiation by increasing the intracellular concentration of Rep monomer for iteron saturation (Chattoraj, 2000).

Chattoraj (2000) thus suggests that dimerization contributes to copy number control in two ways: in regulating initiator supply, and in effecting handcuffing.

What is evident is that a large number of regulatory mechanisms are available to plasmids to control replication initiation. The utilization of combinations of these mechanisms provides a tighter degree of copy control, as the greater the sensitivity to random copy number fluctuations, the greater the capacity to restrict the copy number deviation from the mean (Park *et al.*, 2001; Paulsson, 2002).

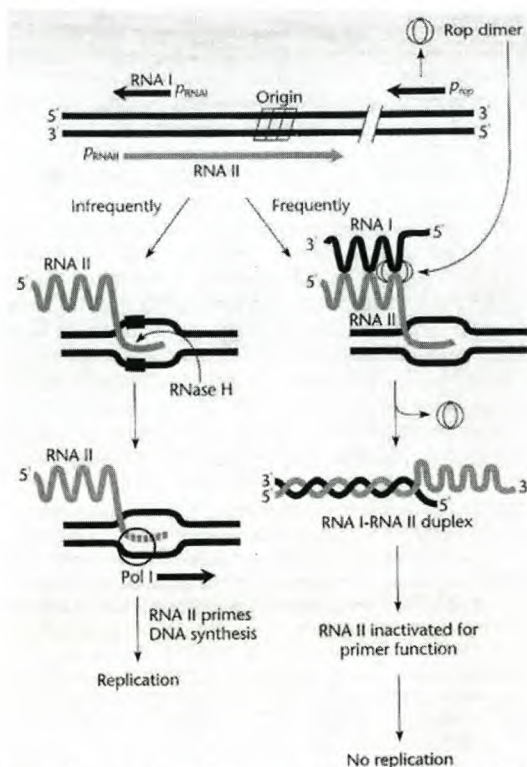
Plasmids of the ColE1 and R1 families use a combination of a ctRNA and an auxiliary regulatory protein to regulate the frequency of replication.

Initiation of ColE1 plasmid replication occurs by the synthesis of an RNA transcript, RNA II, by RNAP (Fig 1.16). The RNA II preprimer anneals to the ColE1 origin, and the 5'-end of the transcript adopts specific conformations that lead to the establishment of a persistent RNA:DNA hybrid. RNase H recognizes this hybrid structure and then cleaves the 3'-end at one of three consecutive adenosine residues, thereby generating a 3'-OH terminus which can now be extended by DNA Pol I. DNA Pol I uses the primer to initiate leading strand synthesis by extending the primer for about 400 nucleotides. The extension of the leading strand exposes on the displaced strand a primosome assembly site (PAS) site. Once the primosome has assembled at this site it translocates in a 5'→3' direction, unwinding the helix and



initiating priming of the lagging strand. DNA Pol I is now replaced by DNA Pol III HE (del Solar *et al.*, 1998; Espinosa *et al.*, 2000).

The dependence on DNA Pol I by ColE1 is probably due to a number of factors. Firstly, steric hindrance by the folded RNA II structure possibly prevents entry of the larger DNA Pol III, and/or the regions to which the primer anneals is not sufficiently melted to allow entry of the DnaB helicase. Also, the unexposed 3'-OH end of the preprimer is likely to be a suitable substrate only for DNA Pol I (del Solar *et al.*, 1998; Espinosa *et al.*, 2000).



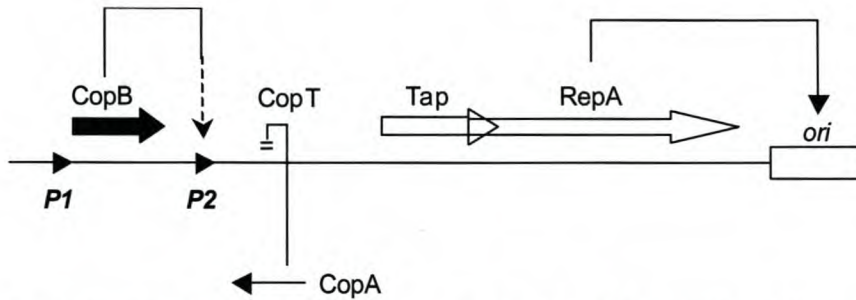
**Figure 1.16:** Replication and copy number control mechanism of the ColE1 plasmid family. Reproduced from Snyder and Champness (2003). Note, the Rop dimer is referred to in the text as the Rom protein, being a product of the *rom* gene (del Solar and Espinosa, 2000).

The initiation of replication, and therefore copy number, of ColE1 plasmids is controlled by a ctRNA molecule, and a regulatory protein which plays an auxiliary role (Fig 1.16). This antisense RNA molecule (RNA I) binds to RNA II by complementary base-pairing which alters the secondary structure of RNA II preventing the establishment of the stable RNA:DNA hybrid. Since the RNA:DNA hybrid is not formed the RNA II molecule cannot be processed by RNase H. Therefore, the availability of the 3'-OH group is rate limiting for initiation (del Solar and Espinosa, 2000). The inhibitory effect of RNA I is enhanced by the Rom protein



(also called Rop) which facilitates the formation of the stable RNA:RNA hybrid complex between RNA II and RNA I (del Solar *et al.*, 1998; del Solar and Espinosa, 2000). As RNA I blocks synthesis of the replication primer, this form of regulation is an example of a direct inhibitor-target regulatory mechanism (Novick, 1987).

The copy number regulation of R1 plasmid is another example of a system involving ctRNA and regulatory protein that plays an auxiliary role (Fig 1.17).



**Figure 1.17:** R1 copy number control mechanism. The CopA ctRNA transcript acts on the *tap-repA* mRNA at the CopT region. CopB acts on promoter *P2* resulting in expression of *repA* being driven almost entirely off the *P1* promoter. Reproduced from del Solar and Espinosa (2000).

The synthesis of R1 RepA is translationally coupled to the leader peptide Tap (translational activator peptide). The ctRNA, CopA, is the main regulatory element which inhibits the translation of *tap*, and therefore, indirectly inhibits translation of *repA*. CopA inhibition of translation occurs through the formation of a RNA:RNA hybrid brought about by the complementary binding (kissing-complex) of CopA to a leader region, termed CopT, of the *tap-repA* mRNA. This kissing-complex sequesters the *repA* RBS preventing translation of the *repA* reading frame. The CopA-CopT hybrid is specifically cleaved by RNase III, but bulged nucleotides in the stem region of CopA result in a structural destabilization which seems to protect the kissing-complex from RNase III. The CopB protein is the second regulatory element which binds to the *P2* promoter of this replicon. The expression of the *tap-repA* mRNA transcript can be driven from either the *P1* or *P2* promoter. At steady state, CopB is present at saturating concentrations thereby blocking transcription from *P2* such that expression of *repA* is driven almost entirely from the *P1* promoter. The consequence of this is that RepA is translated off a polycistronic *copB-tap-repA* mRNA. However, when the copy number decreases, or is low due to early establishment in a new host, the *P2* promoter is derepressed allowing for transcription of the *tap-repA* mRNA. This leads to a transient increase in RepA synthesis, which increases the frequency of



replication initiation to increase the copy number. Deletion of the *copB* gene results in an eight-fold increase in R1 copy number (Blomberg *et al.*, 1992; del Solar and Espinosa, 2000).

Between the *repB* primase and *repA* helicase of RSF1010 is found the *cac* gene which regulates the expression of the *repAC* operon at a transcriptional level (Maeser *et al.*, 1990). The product of the *cac* gene acts as a repressor of the *repAC* P<sub>4</sub> promoter of this replicon, and in doing so, the RepC is rate-limiting for replication initiation (Sakai and Komano, 1996). The discovery of a ctRNA molecule in the highly related (almost identical) plasmid R1162, suggests that a translational control mechanism may also be regulating copy number in the RSF1010 family. This small 75-bp ctRNA defines a complementary coding sequence that overlaps the RBS and the first two codons of the *repA* gene. Mutational analysis revealed that this ctRNA is a negative regulator of replication initiation, as mutation of this ctRNA resulted in a detectable increase in copy number (Kim and Meyer, 1986). The interaction between the transcriptional regulator (*cac* gene product) and the ctRNA has not been investigated further. Consequently, no model for copy number control involving both elements has been proposed.

The involvement of ctRNA in the control of Cole2 replication has recently been proposed. A stable kissing-complex is thought to form between the ctRNA and a complementary region on the *rep* mRNA. Even though the ctRNA binding site does not overlap the RBS, stable complex formation appears to be enough to inhibit translation of the Rep protein (del Solar and Espinosa, 2000).

In a discussion of plasmid replication the mechanism of replication termination is relevant. The first replication arrest sequence was identified in R6K, designated *ter*, and determined to be a 20-bp inverted repeat. The *ter* sequence is a barrier to unidirectional replication initiated at *ori-α* or *ori-β* of R6K (Khatri *et al.*, 1989; del Solar *et al.*, 1998). The organization of the *ter* sequence was found to be two separable and polar terminus sites that are bound by the *E. coli* Tus protein (Horiuchi and Hidaka, 1988). It is the interaction of the Tus protein with the replisome which brings about plasmid replication arrest. The identification of the essential features of



the *ter* site has allowed for the identification of *ter*-like sequences in R1, and assisted in the identification of the *ter* sequences found on the *E. coli* chromosome that bring about chromosomal replication arrest (section 1.2.6). A sequence termed *terH* has been identified close to the PAS site of ColE1 and is active in ColE1 lagging strand synthesis termination (del Solar *et al.*, 1998). The Tus-*ter* interaction was also utilized to investigate and help define the *ssi*-directed strand displacement mechanism of RSF1010/R1162 (Zhou *et al.*, 1991).

Once plasmid replication has been completed the two catenated plasmid molecules need to be resolved. A specific type II topoisomerase, Topo IV, is involved in plasmid segregation. The decatenation reaction brought about by Topo IV is followed by the action of DNA gyrase which introduces supercoiling into the daughter plasmids (del Solar *et al.*, 1998).

Finally, hemimethylation is also thought to regulate plasmid replication. One mechanism by which this might occur is through the sequestration of hemimethylated origins of recently replicated plasmids to the cell membrane by SeqA. As methylation increases the replication frequency, the methylation of hemimethylated plasmid DNA could stimulate replication by enhanced bending or unwinding of the DNA (del Solar *et al.*, 1998).

In conclusion, the mechanism of strand synthesis by the replisome is likely to be the same between all plasmid families, irrespective of the range of replication initiation and copy number control mechanisms utilized. However, differences lie in the dependence of a plasmid's replication mechanism on DNA Pol I. The replisome of Pol I-independent plasmids initiate and carry out DNA synthesis, whereas for DNA Pol I-dependent plasmids, the replisome is only formed after the initial synthesis of the leading strand by DNA Pol I (Espinosa *et al.*, 2000). Therefore, the degree of dependency on host-encoded replication factors, together with a diversity of replication initiation and copy number regulation mechanisms, contributes to a plasmid's host range.



#### **1.4) Plasmid host range constraints**

The host range of a plasmid is influenced by: (i) dependence on host replication proteins, (ii) origin structure and topology, (iii) Rep protein interactions, and (iv) efficient plasmid-host communication to coordinate transcription and translation of replication genes.

##### **1.4.1) Dependence on host replication proteins**

DNA Pol I and DnaA protein are emerging as the two critical factors which constrain plasmid host range. As discussed above, the ColE1 family of plasmids is a group of DNA Pol I-dependent, narrow host range plasmids which originate from purple bacteria of the  $\gamma$  subdivision (Espinosa *et al.*, 2000). The narrow host range phenotype of these plasmids is probably related to the inability of these plasmids to functionally interact with the DNA Pol I protein of other bacterial species (del Solar *et al.*, 1996).

The discussion on theta replication mechanisms (section 1.3.2.4) highlights the varying degree of dependence on DnaA to assist in the initiation of plasmid replication. This dependence on DnaA can also contribute to the host range of a plasmid. A study with the broad host range plasmid RK2/RP4 revealed that the inability of a DnaA ortholog of a host bacterium to form a stable and functional complex with the DnaA-boxes of the RK2/RP4 origin can constrain this plasmid's host range (Caspi *et al.*, 2000). This study showed that the DnaA protein from *E. coli*, *P. putida* and *P. aeruginosa* bind to the DnaA-boxes of the RK2/RP4 origin and induce open complex formation. However, the DnaA protein of *Bacillus subtilis* and *Streptomyces lividans* are capable of interacting with the DnaA-boxes of RK2/RP4 origin, but do not bind stably, and are unable to induce open complex formation (Caspi *et al.*, 2000).

Maestro *et al.* (2002) report the first broadening of a plasmid's host range by mutagenesis of a host DnaA protein, thereby providing support for the hypothesis that DnaA-Rep protein interactions modulate plasmid host range. Plasmid pPS10 is a



narrow host range plasmid that is able to efficiently replicate in *P. putida* and *P. aeruginosa*, but is unable to successfully establish itself in *E. coli*, particularly at temperatures above 30°C. Neither the expression of *Pseudomonas* DnaA protein *in trans*, nor the adjustment of the pPS10 DnaA-box to the *E. coli* DnaA-box consensus sequence was able to significantly improve the establishment of pPS10 in *E. coli* (Nieto *et al.*, 1992). By ethylmethanesulphonate mutagenesis of *E. coli* LE392, Maestro *et al.* (2002) introduced three mutations in the *E. coli dnaA* gene. Due to the LE392 *supE* and *supF* genotype, this triple mutant *dnaA* gene expresses a heterologous population of three DnaA protein variants. The first variant is a truncated protein owing to the introduced Q14amber mutation. The second mutant, brought about by the *supE* suppression of the amber codon to Gln (same amino acid as wild-type), is a double mutant owing to P297S/A412V. The third variant, arising from the translation of the amber codon to Tyr by *supF*, is a Q14Y/P297S/A412V mutant (Maestro *et al.*, 2002). This triple mutant DnaA protein (DnaA403) allows for efficient propagation of pPS10 in *E. coli* LE392. Although physical interaction between DnaA403 and pPS10 RepA was demonstrated, the effect that these three mutations have on RepA-DnaA interaction and how they extend the pPS10 host range to include *E. coli* is still speculative (Maestro *et al.*, 2002).

#### 1.4.2) Origin structure and topology

The four DnaA-boxes found in the RK2/RP4 origin provide insights into the influence of origin structure on host range. Two out of the four DnaA-boxes are required for replication in *E. coli* or *P. putida*, yet three DnaA-boxes are dispensable for replication in *Azotobacter vinelandii* (Doran *et al.*, 1999[b]). This gives an indication as to the importance of the DnaA-box proximal to the iterons for replication initiation in these species. Contrary to this structural requirement for replication in the above species, RK2/RP4 is able to dispense with all four DnaA-boxes to replicate in *P. aeruginosa* (Doran *et al.*, 1999[b]).

Origin topology can affect host range as alteration of helical phasing and intrinsic DNA curvature can prevent strand opening at the origin and prevent plasmid replication. This was demonstrated by interruption of the RK2/RP4 origin structure through the insertion of 6-bp at designated intervals in this plasmid's origin. As 6-bp



is approximately a half of a helical turn, this insertion alters the helical phasing. The results of this study showed that the insertion of 6-bp between the iterons and the four DnaA boxes had a significant effect on RK2/RP4 replication in *E. coli*, but had no effect on replication in *P. aeruginosa*. However, interruption of the A+T-rich region with a 6-bp fragment, or insertion of this 6-bp fragment between the iterons and the A+T-rich region, abolished replication in *P. aeruginosa* and *P. putida* (Doran *et al.*, 1998).

Origin topology can also affect host range as structures critical for replication initiation and control can be incorrectly displayed in a particular host. Therefore, the *dso* stem-loop of plasmids using the rolling-circle replication mechanism, or a transcription terminator stem-loop, may not be functional in a specific species (del Solar *et al.*, 1996).

For plasmids utilizing rolling circle replication the *sso* element appears to be the limiting factor for host range. The cleavage of the *dso* of these plasmids creates a 3'-OH end which is used to prime leading strand synthesis, negating the need to rely on the host primase. However, the *sso* of the same plasmid is reliant on the host RNAP and DNA Pol I for the conversion of the ssDNA intermediate to dsDNA, although alternate priming pathways can accomplish this conversion (Espinosa *et al.*, 1995). Another strategy used by rolling circle plasmids to overcome this problem is to carry two active *sso* elements such as the *ssoA* and *ssoU* found on pMV158 which extends the host range (Espinosa *et al.*, 2000).

However, containing more than one origin element is not necessarily indicative of a broad host range. For example, plasmid R6K has three origins, yet still has a narrow host range. The close association of R6K  $\pi$  protein with host specific factors DnaA, DnaB and DnaG restricts this plasmid to closely related species (del Solar *et al.*, 1996).



### 1.4.3) Rep protein interactions

Rep protein interactions can influence the ability of a plasmid to initiate replication in different host species. A conservative Ala→Val substitution in the LZ motif of pPS10 RepA resulted in the efficient establishment of this plasmid in *E. coli* (otherwise restricted to *P. aeruginosa* and *P. putida*), and stable maintenance at 37°C (Fernández-Tresguerres *et al.*, 1995). As this mutation had neither a detectable effect on the binding efficiency of RepA to the pPS10 origin, nor influenced host range through increased expression of RepA by affecting autoregulation, it suggests that this alteration in the LZ motif improved interactions between the pPS10 and host replication factors (Fernández-Tresguerres *et al.*, 1995). In a recent study, hydroxylamine mutagenesis of the pPS10 *repA* gene resulted in the isolation of five different mutant pPS10 plasmids which displayed host range alterations. Four of the mutant pPS10 plasmids broadened their host range by demonstrating an ability to replicate in *E. coli* at 37°C. One of the mutant plasmids could no longer replicate in *P. aeruginosa*, but could establish itself in *E. coli*, thus changing its host range (Maestro *et al.*, 2003). Isolation and characterization of the five mutant pPS10 RepA proteins showed different single amino acid substitutions, and all of these amino acid changes are buried deep in the protein structure. These mutations were also found to positively affect the binding of RepA to both the iterons and the *repA* operator region of pPS10 (Maestro *et al.*, 2003). More significantly, these mutations all displayed an improved pPS10 RepA-DnaA interaction which may account for the broadening in pPS10 host range (Maestro *et al.*, 2003).

Similarly, mutational analysis of the C-terminal end of the RK2/RP4 TrfA protein found that this domain affected host range, suggesting that the C-terminus of TrfA interacts with host replication factors. For example, an R382K substitution resulted in the loss of replication ability in *E. coli* and *A. vinelandii*, but had no effect on replication in *P. putida* or *Agrobacterium tumefaciens* (Cereghino and Helinski, 1993; del Solar *et al.*, 1996; del Solar *et al.*, 1998). The truncated TrfA-33 protein may also be unable to initiate replication in certain species due to inefficient interaction with host replication factors.



#### **1.4.4) Efficient plasmid-host communication**

Another consideration affecting host range is efficient communication between plasmid and host specific factors. Balanced transcription and translation of plasmid genes is essential to prevent loss of a plasmid from a population due to host inhibition of an essential plasmid-encoded regulator (del Solar *et al.*, 1996). The inhibition of a critical copy number regulator could prevent the establishment of the plasmid in the host by complete inhibition of plasmid replication, or result in uncontrollable copy number fluctuations resulting in plasmid loss.

In summary, all the factors discussed above contribute significantly to plasmid host range. The two broad host range plasmids RK2/RP4 and RSF1010 are able to establish themselves in a range of taxonomically distinct species, yet even these two plasmids achieve a broad host range by different means. Two oligomeric forms of a Rep protein is thought to be the strategy used by RK2/RP4 to allow for interactions with the replication proteins of different hosts. For instance, there may be a particular TrfA oligomer requirement for DnaB ortholog loading and activation in specific hosts. RSF1010 is probably the plasmid replicon which is most independent of host replication factors. Its independence from DnaA, DnaB and DnaG for replication allows for an exceptionally broad host range that is not limited to a subdivision of the proteobacteria, but includes representatives of both Gram-negative and Gram-positive bacteria, and cyanobacteria (for a comprehensive host range list see Rawlings and Tietze, 2001). However, to achieve this independence from host replication factors, it infers that the promoters of the RSF1010 replication genes must be recognized and expressed in these different hosts.

#### **1.5) Conclusion**

The classification of plasmids into distinct groups is hampered by the mosaic composition of plasmids. Without a unique and ubiquitous gene, such as the 16S rRNA gene, the classification of plasmids is restricted to the one phenotypic characteristic shared by all plasmids: the ability to replicate (Osborn *et al.*, 2000). Therefore, the plasmid replicon is the defining feature of a plasmid and the single unifying system which can be used to define evolutionary relationships between



plasmids (Osborn *et al.*, 2000). It is the plasmid replicon upon which selective pressure acts to drive replicon evolution towards a system that achieves a balance between stable plasmid carriage in a particular host and an acceptable metabolic burden imposed by the plasmid on the host. Stable plasmid carriage implies a need for communication between host and replicon that is achieved through coevolution (Lilley *et al.*, 2000). Thus, it is the plasmid-host history which would optimize plasmid-host molecular interactions allowing for stable replication and segregation of the replicon in the host population.

The specificity of ColE1 for DNA Pol I of taxonomically related species (del Solar *et al.*, 1996) is most likely the consequence of plasmid-host evolutionary history. The species-dependent requirement for DnaA-boxes in the initiation of RK2/RP4 replication (Doran *et al.*, 1999[b]) may also reflect the history of this plasmid with its associated hosts.

The reference to the broad host range IncP $\alpha$  plasmid RK2/RP4 introduces a further complication in replicon evolution. Should the replicon acquire a transfer function through a horizontal event, conjugation or mobilization would increase the selective pressure on the replicon by introducing the replicon to a variety of host genetic backgrounds. Different genetic backgrounds imply different host-plasmid interactions with serious consequences for plasmid replication initiation and control. A successful replication strategy to overcome plasmid-host barriers is presented by the IncQ plasmid RSF1010. The reliance of RSF1010 on host replication machinery appears to be limited to only DNA gyrase and DNA Pol III; with all the replication requirements being plasmid-encoded and plasmid template specific (Scherzinger *et al.*, 1991).

The direction in which replicon evolution proceeds, either towards greater specificity for host replication factors, or towards independence from the host replication machinery, will dictate whether a plasmid is restricted to taxonomically related species, or will be able to colonize a wide range of bacterial species.



Micro- and macromutation are the two driving forces behind the evolution of the array of replication initiation and copy control mechanisms that have been documented. Micromutation, through the introduction of point mutations and the insertion/deletion of small DNA fragments, is most probably behind the evolution of copy number control mechanisms and incompatibility determinants. Micromutation events have been found to occur in many high copy number plasmids (Sýkora 1992; Osborn *et al.*, 2000). Macromutation, however, refers to insertions/deletions and rearrangements of large segments of DNA, and is most likely responsible for gross replicon changes, but at the highest risk. Sýkora (1992) introduced the concept of plasmid speciation through cointegrate formation in an attempt to address macromutation. In brief, the model for plasmid speciation by cointegrate formation between two different replicons suggests that one replicon of the cointegrate is released from conservative selection pressure allowing it to freely evolve into a new replicon. Resolution of the cointegrate after macromutation will release the new replicon variant.

In summary, evolution of plasmid replicons has given rise to a diverse range of replication and copy control mechanisms. An understanding of the replication mechanism employed by a plasmid provides insights into its host-dependence, which ultimately determines the replicon's host range.

## 1.6) Aims of this study

A ±14-kb plasmid isolated from the chemolithoautotrophic bacterium *Acidithiobacillus caldus* strain “f” was cloned into a ColE1-vector, and subsequently shown to replicate in an *E. coli polA* mutant. This suggested that the plasmid isolate had a broad host range replicon. Analysis of the replicon DNA sequence, indicated that this plasmid had a replicon with significant similarity to the IncQ family of plasmids. Furthermore, the replicon of this plasmid was found to be more closely related to the broad host range IncQ-like plasmid pTF-FC2, which had been isolated from *Acidithiobacillus ferrooxidans* strain FC (Rawlings *et al.*, 1984), than to other members of the IncQ plasmid family. Since the *At. caldus* plasmid isolate was now the second IncQ-like replicon isolated from the biomining environment, we were in the unique position to investigate not only the molecular biology of pTC-F14, but its interaction with the related pTF-FC2 plasmid, and with other members of the IncQ family of plasmids. Therefore, the overall aim of this study was the molecular characterization and comparative analysis of the replicon of this new IncQ-like plasmid – pTC-F14.

Since the biomining environment is thought to be genetically isolated from animal- and human-associated environments from which IncQ plasmids have been isolated, it raised the question as to how the two IncQ-like plasmids pTC-F14 and pTF-FC2 evolved with respect to each other. Furthermore, we wished to establish how the two plasmids isolated from biomining bacteria are related to the rest of the IncQ plasmid family. This study was narrowed down to focus mainly on the pTC-F14 replicon region, as this could provide insights into the evolution and relatedness of IncQ replicons found in diverse environments.

The discovery that pTC-F14 and pTF-C2 were compatible when co-resident in an *E. coli* host cell line suggested that pTC-F14 belonged to a unique incompatibility group within the IncQ plasmid family. Since the two plasmids presumably originated from the same ancestor plasmid, it was hoped that an investigation of the pTC-F14 replicon would uncover the molecular events that might lead to the evolution of a new incompatibility group. In light of this, we determined to what extent the replication proteins of pTF-FC2 and related IncQ plasmids were able to complement those of



pTC-F14. This would provide information on the functional interaction between replication proteins of the IncQ plasmids. By correlating the results from the cross-complementation studies to the incompatibility phenotype expressed by plasmid pairs, it was hoped that this might aid in determining the events involved in the evolution of a compatible replicon.

Since cross-complementation was observed, cross-regulation of pTC-F14 replication gene expression by pTF-FC2 and other related IncQ plasmids was also examined. Northern blot and reverse-transcriptase polymerase chain reaction (RT-PCR) analysis were used to determine the size and number of transcripts initiated from the pTC-F14 replicon. Such information would aid in the identification of critical promoter regions of the pTC-F14 replicon, and this in turn would assist in promoter-reporter gene (*lacZ*) fusion-construct cloning for  $\beta$ -galactosidase assays. It was hoped that the  $\beta$ -galactosidase assays would show both how the pTC-F14 replication genes were regulated, and whether any cross-regulation by related IncQ plasmids occurred. Cross-regulation studies of replicon gene expression may also help to explain how replicon evolution overcomes plasmid incompatibility.

## Chapter Two

### Replicon analysis of the IncQ-like plasmid pTC-F14 isolated from *Acidithiobacillus caldus* strain “f”

#### Contents

2.1)	Introduction	72
2.2)	Materials and Methods	85
	2.2.1) Bacterial strains, media and growth conditions	85
	2.2.2) DNA techniques, sequencing and analysis	85
	2.2.3) Harvesting bacteria and preparation of chromosomal DNA	86
	2.2.4) Trans-Alternating Field Electrophoresis (TAFE)	86
	2.2.5) Host range determination	87
	2.2.6) Plasmid copy number determination	87
	2.2.7) Nucleotide sequence accession number	88
2.3)	Results	89
	2.3.1) Isolation of pTC-F14 and identification of the plasmid replicon	89
	2.3.2) Comparative analysis of the replication region	90
	2.3.3) Plasmid pTC-F14 appears to have a broad host range	95
	2.3.4) Copy number of pTC-F14	96
2.4)	Discussion	97



## 2.1) Introduction

Microbial biooxidation of refractory metal-sulfide ores and concentrates has become a technically and economically viable alternative to the historically environmentally polluting pyrometallurgical and hydrometallurgical processes of roasting and pressure-leaching (Morin, 1995; Rawlings *et al.*, 2003). The first biohydrometallurgical process to pre-treat refractory arsenopyrite (FeAsS) ores and concentrates by microbial biooxidation was patented in the 1980's by Gencor S.A. Ltd. (South Africa) as the Biox<sup>®</sup> process (Rawlings, 1998). This was followed by the competing BACOX process, developed by BacTech (Australia), to pre-treat gold-bearing concentrates, and this process operates at 50°C as opposed to 40°C for the Biox<sup>®</sup> process (Miller, 1997). Both the Biox<sup>®</sup> and BACOX processes are examples of commercial biomining processes utilizing continuously-operating, highly aerated, stirred-tank bioreactors. However, bioleaching by percolation of leaching solutions through crushed ore or concentrates stacked in columns, heaps, or dumps is also an economically viable alternative for biomining operations (Rawlings, 1998; Rawlings *et al.*, 1999[b]). With the success of arsenopyrite (FeAsS) and pyrite (FeS<sub>2</sub>) biooxidation operations, investigations in the recovery of base metal-sulfides from pentlandite ([FeNi]<sub>9</sub>S<sub>8</sub>), spalerite (ZnS), covellite (CuS) and chalcocite (Cu<sub>2</sub>S) have been undertaken (Rawlings *et al.*, 2003). Successful development of biomining processes for nickel and copper recovery has led to the registration of the BioNic<sup>®</sup> and BioCOP<sup>™</sup> processes for the pre-treatment of nickel-sulfide and copper-sulfide ores, respectively (Dew and Miller, 1997; Rawlings *et al.*, 2003).

Microbial bioleaching is the solubilization of water insoluble metal-sulfides by oxidation of these minerals to their metal-sulfates. The reaction describing mineral dissolution is not identical for all metal-sulfides, as the acid-solubility of the metal sulfide determines which of two reaction mechanisms is utilized. The distinction between these two bioleaching mechanisms lies in their different reaction intermediates. The thiosulfate mechanism for the oxidation of acid-insoluble metal-sulfides such as pyrite, involves the ferric (Fe<sup>3+</sup>)-iron attack on the metal-sulfide, with thiosulfate being the intermediate and sulfate the end-product. The polysulfide mechanism involves the combined attack on acid-soluble metal-sulfides by ferric-iron and protons (provided by sulfuric acid), with the elemental sulfur intermediate being



oxidized to sulfate by sulfur-oxidizing bacteria. The ferrous ( $\text{Fe}^{2+}$ )-iron end product of both of these reaction mechanisms is recycled by iron-oxidizing bacteria back to ferric-iron. Thus, the microbial contribution to solubilization of metal-sulfide ores is the production of sulfuric acid for proton attack, and the oxidation of ferrous-iron to produce ferric-iron (Schipper and Sand, 1999).

Therefore, mineral dissolution is accomplished through the concerted action of a consortium of bacteria which are responsible for the oxidation of metal-sulfide ores and concentrates. This consortium composition is dependent on the operational temperature of the biomining process, but generally consists of chemolithoautotrophic and heterotrophic eubacteria, and Archaea (Rawlings, 2002). The predominant eubacteria found in stirred-tank bioreactors operating at 40°C or less are the highly motile Gram-negative acidophilic chemolithoautotrophs *Acidithiobacillus ferrooxidans*, *Acidithiobacillus thiooxidans*, *Acidithiobacillus caldus* and *Leptospirillum ferrooxidans* (Rawlings *et al.*, 2003).

These acidophilic bacteria are highly adaptable to the extreme environment in which they are found (Rawlings, 1997). The acidic, inorganic environment in which biooxidation occurs is very nutrient poor, but the chemolithoautotrophs are able to meet their energy requirements by utilizing an inorganic energy source ( $\text{Fe}^{2+}$  and/or reduced inorganic sulfur) as electron donor, and air which provides the preferred electron acceptor ( $\text{O}_2$ ). Air also provides carbon ( $\text{CO}_2$ ) and nitrogen ( $\text{N}_2$ ), while water and ore provide trace elements in the form of impurities. The Acidithiobacilli are responsible for the oxidation of reduced inorganic sulfur compounds, whereas *L. ferrooxidans* can only oxidize ferrous iron (Brierley, 1982; Rawlings and Silver, 1995, Rawlings, 2002).

For many years *At. ferrooxidans* was considered the most important bacterium in commercial biomining processes that operate at 40°C or less. This conclusion had been reached from inherently biased pour-plate assays that skewed growth towards bacteria that can be readily cultured on solid medium, and the ability of *At. ferrooxidans* in batch culture to initially out-grow competitors in liquid ferrous iron, or pyrite medium (Rawlings *et al.*, 1999[b]; Rawlings, 2002). *At. ferrooxidans* is an obligately autotrophic mesophilic chemolithotrophic bacterium which grows



optimally at 20-35°C, pH 1.8-2.0, and is capable of obtaining its energy by oxidation of either reduced iron or sulfur compounds (Rawlings, 2002). However, bioleaching processes operating in the 35-50°C range are dominated by a combination of iron-oxidizing Leptospirilli and the acidophilic sulfur-oxidizing *At. caldus* (Rawlings *et al.*, 1999[a]). *At. caldus* is a moderately thermophilic chemolithoautotroph which has an optimum growth temperature of 45°C, and a growth pH optimum of 1.3-2.5 (Hallberg and Lindström, 1994).

A number of surveys have been undertaken in an attempt to isolate natural plasmids in *At. ferrooxidans* (Rawlings and Kusano, 1994). The carriage of such traits as metal ion resistance, and the possibility of developing genetic systems for the improvement of biomining strains has prompted these investigations. Thus, plasmid pTF-FC2 was discovered and isolated from the arsenic-resistant *At. ferrooxidans* strain FC (Rawlings *et al.*, 1984). Sequencing of the entire pTF-FC2 plasmid revealed three regions: a replicon region, a mobilization region, and a transposon-like element (Rawlings *et al.*, 1993). This 12.2-kb cryptic plasmid was initially found to replicate in *E. coli*, *P. aeruginosa*, *Klebsiella pneumoniae*, and *Thiobacillus novellus*; with this range subsequently being extended to at least one representative of each proteobacteria subdivision (Dorrington and Rawlings, 1989; Rawlings and Kusano, 1994). Such a host range of taxonomically distinct eubacteria is indicative of a broad host range plasmid, and this was further substantiated by the finding that pTF-FC2 had a replicon similar to the *E. coli* incompatibility group Q (IncQ) family of plasmids (Dorrington and Rawlings, 1990).

Plasmids of the IncQ family have a very broad host range, being capable of replication in a wide variety of Gram-negative and also Gram-positive bacteria (Frey and Bagdasarian, 1989). Although none of the IncQ and related plasmids are self-transmissible they are efficiently mobilized by IncP $\alpha$  (RK2/RP4/R68) and IncP $\beta$  (R751) plasmids (Derbyshire and Willets, 1987). As a result of their host range and mobility, these plasmids are highly promiscuous. The best-studied IncQ plasmids are RSF1010 (Guerry *et al.*, 1974), R1162 (Meyer *et al.*, 1982) and R300B (Barth and Grinter, 1974), which are identical, or nearly identical, in spite of RSF1010 being isolated from *E. coli*, R1162 from *P. aeruginosa* and R300B from *Salmonella*



*enterica* serovar Typhimurium (Rawlings and Tietze, 2001). The prototype IncQ plasmid is RSF1010, and this 8684-bp multicopy plasmid carries resistance to streptomycin (Sm) and sulfonamides (Su) (Rawlings and Tietze, 2001).

Plasmid NTP2 was recognized as the prototype plasmid of a group of R-plasmids described as small non-conjugative plasmids conferring resistance to sulfonamides (Su) and streptomycin (Sm). This group of plasmids was found to be relatively common in clinical isolates of enteric bacteria (De Graaff *et al.*, 1978). A survey by Barth and Grinter (1974) of Su-Sm resistant R-plasmids isolated from both broad host and geographical range showed that these plasmids were very similar, and led these authors to postulate that an ancestral plasmid had spread efficiently, with relatively few molecular alterations, around the world. The Su-Sm R-plasmid RSF1010 was later found to be identical to the NTP2 plasmid isolate (De Graaff *et al.*, 1978), and RSF1010 has since been widely used in laboratory studies.

Early studies of RSF1010 biology revealed that this R-plasmid was genetically stable in *polA1* mutants, mobilizable, and had a copy number of 10-12 copies per chromosome (Barth and Grinter, 1974, Guerry *et al.*, 1974).

Further investigations into identifying the region encoding replication functions began to reveal the replicon and replication mechanism. It was the work of De Graaff *et al.*, (1978) that revealed the unidirectional or bidirectional replication mode of RSF1010. Analysis of *EcoRI*-cleaved replicative intermediates of RSF1010 containing the transposon *TnA* allowed for the assignment in these early studies of the *oriV* relative to the Su and Sm genes on the plasmid, and the discovery that unidirectional or bidirectional replication of RSF1010 was initiated with about equal probability (De Graaff *et al.*, 1978; Frey and Bagdasarian, 1989).

The replication requirements of RSF1010 were found to be very different from that of ColE1 plasmids as Diaz and Staudenbauer (1982) discovered that RSF1010 was able to initiate replication independently of RNAP. Using antibiotics in an *in vitro* cell extract system, prepared from both *E. coli* and *P. aeruginosa* RSF1010 transformants, these authors were able to demonstrate RNAP-independent replication initiation, but found that RSF1010 had a requirement for DNA gyrase and DNA Pol III. By using



the protein biosynthesis inhibitors chloramphenicol and streptomycin in this *in vitro* replication assay it was found that these antibiotics had no significant effect on RSF1010 replication. This indicated that RSF1010 replication was not coupled to the concomitant synthesis of an essential host replication protein. With the addition of rifampicin and streptolydigin to this replication assay, no inhibitory effect was demonstrated for RSF1010 replication showing that bacterial host RNAP was not required for replication. The DNA Pol III inhibitor arabinosylcytosine triphosphate, and the DNA gyrase specific inhibitors novobiocin and oxolinic acid, were found to inhibit RSF1010 replication, inferring a requirement for these replication proteins (Diaz and Staudenbauer, 1982).

Diaz and Staudenbauer (1982) were thus led to conclude that RSF1010 encoded a *trans*-acting replication function enabling RSF1010 to initiate replication independently of specific host replication factors, and that this was possibly the clue to the remarkable host range of this plasmid.

Expanding on this work, Scholz *et al.* (1985) used *E. coli* strains with conditional mutations in the *dnaB*, *dnaC* and *dnaG* genes to prove that under conditions where either DnaB, DnaC or DnaG function was inhibited, the replication of RSF1010 was unaffected. A radioactive nucleotide assay showed that under non-permissive conditions (41°C) the incorporation of <sup>3</sup>H-thymine was drastically reduced in chromosomal DNA, but not for supercoiled RSF1010 plasmid DNA. Studies with a DNA Pol III  $\gamma$  subunit mutant, however, suggested that DNA polymerase III was required for the replication of RSF1010 in *E. coli* cells (Scholz *et al.*, 1985). Scherzinger *et al.* (1991) went on to determine that RSF1010 replication was also independent of host encoded DnaA, DnaT and DNA Pol II, but was dependent on SSB protein, and they confirmed the requirement for the DNA Pol III  $\gamma$  subunit. Integrative suppression of a *dnaA46* temperature-sensitive mutant *E. coli* by R1162 also demonstrated DnaA-replication independence (Brasch and Meyer, 1988). Like the broad host range plasmid RK2/RP4, which was known to encode replication functions rendering it independent of host replication factors, these studies gave an indication as to how RSF1010 is able to overcome host-dependent replication barriers that restrict narrow host range plasmids.

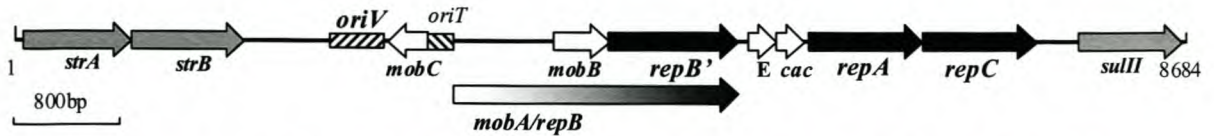


Reconstitution of RSF1010-DNA containing cell extracts with partially purified RSF1010 protein fractions identified the 3 proteins essential for RSF1010 replication (Scherzinger *et al.*, 1984). A plasmid rescue technique involving the circularized *oriV* region of RSF1010 (plasmid pMMB12) was used to identify a RSF1010 fragment containing all the essential replication genes. This recombinant plasmid, designated pMMB2, which contained a 4.7-kb RSF1010 fragment of RSF1010 inserted into a ColD-vector plasmid, was then used to purify three proteins that were shown by means of an *in vitro* cell extract system to initiate RSF1010 replication when all three protein fractions were supplied. The results from this assay system showed the requirement for fraction B in the assay, and that no DNA synthesis was detected in the absence of fractions A and C (Scherzinger *et al.*, 1984). Deletion constructs used in plasmid rescue studies, and use of these constructs to produce protein fractions for the *in vitro* cell extract system, confined RSF1010 replication genes to an approximate 3.1-kb fragment, and again demonstrated the requirement for all three proteins to initiate DNA synthesis. The RSF1010 replication genes were subsequently designated (in accordance with the protein fractions described above): *repB*, *repA*, and *repC* (Scherzinger *et al.*, 1984). These authors noted the similarity of RepC to phage  $\lambda$  O protein and *E. coli* DnaA protein, both of which play a crucial role in DNA replication by binding to their analogous replication origins to initiate replication.

Having purified these proteins essential for RSF1010 replication, cell extracts of transformed cells containing the pMMB2 plasmid deletion constructs were made. These cell extracts were then screened for deficiencies in either of the three protein activities, and were then reconstituted with the purified proteins. This, together with analysis of the deletion constructs, allowed for the gene order and their location on the restriction enzyme map of RSF1010 to be determined (Scherzinger *et al.*, 1984).

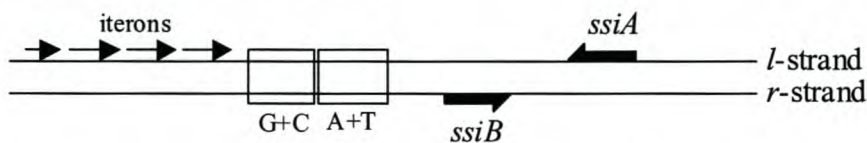
It was the publication in 1989 of the complete nucleotide sequence of RSF1010 that confirmed the findings of Scherzinger *et al.* (1984), and gave greater detail to the replicon structure and potential regulatory regions (Fig 2.1.1) (Scholz *et al.*, 1989).





**Fig 2.1.1:** Genetic map of RSF1010 showing location of *oriV* relative to replication genes *mobA/repB* (*repB'*), *repA* and *repC*. Direction of arrows indicates direction of gene transcription. The sulfonamide resistance gene, *sullI*, and streptomycin resistance genes, *strA* and *strB*, are represented by grey arrows. 'E' represents ORF E with no known function, and *cac* refers to ORF F (Control of A and C genes). Map constructed from published sequence data (accession numbers M28829 and M21475), and Scholz *et al.* (1989).

Analysis of the RSF1010 *oriV* has revealed three major features (Fig 2.1.2): (i) three and a half perfect direct repeats of 20-bp separated by 2-bp spacers, (ii) adjacent A+T-rich and G+C-rich regions of 31-bp and 28-bp, respectively, (iii) and a large extended inverted repeat (40-bp stem, 40-bp loop) containing the two plasmid-specific single-stranded DNA priming sites, *ssiA* and *ssiB*, which are in the *l*- and *r*-strands, respectively. These *ssi* sites are the target sites for the independent priming of their complementary strand syntheses (Lin and Meyer, 1987; Honda *et al.*, 1988; Honda *et al.*, 1991; Rawlings and Tietze, 2001).



**Fig 2.1.2:** Diagrammatic representation of the *oriV* region of RSF1010 showing the 3½ iterons (arrows), the location of the G+C-rich and A+T-rich regions (boxed), and the orientation and location of *ssiA* and *ssiB* on the *l*-strand and *r*-strand, respectively.

The 31 kDa RepC protein acts as a positive specific initiator of RSF1010 replication by regulating the frequency of replication initiation (Haring *et al.*, 1985; Scherzinger *et al.*, 1991). A DNA fragment retardation experiment using purified RepC and restriction endonuclease digested fragments of the RSF1010 *oriV* region showed the unequivocal binding of RepC to the *oriV* (Haring *et al.*, 1985). The RepC (previously RepIB) of R1162 was found to also bind to a second region in the *oriV*. This site was mapped to the A+T-rich region and lies approximately 60-bp from the iterons (Kim and Meyer, 1991). It was in this A+T-rich region that two 10-bp direct repeats were identified that were speculated to be functionally equivalent to the *oriC* 13-mer repeats (Kim and Meyer, 1991). It was later demonstrated that RepC binding to the iterons induces DNA bending and that even a single iteron copy is bent upon RepC binding (Miao *et al.*, 1995[b]). This RepC-iteron complex is thought to introduce



structural changes in the *oriV* that bring about localized melting and open complex formation. Miao *et al.* (1995[b]) also demonstrated that three iterons were functionally optimal for RSF1010 replication, and that mutagenesis of iterons away from the consensus resulted in the replication ability of RSF1010 being severely hampered. This confirmed the finding of Scherzinger *et al.* (1991) who earlier had demonstrated that deletion of a single iteron resulted in loss of replication ability in a RSF1010 *in vitro* replication assay. The RepC protein was also found not to form a physical complex with RepA, even in the presence of *oriV* DNA, and that RepA did not affect the DNA binding property of RepC (Haring *et al.*, 1985).

The sequence data obtained by Scholz *et al.* (1985) gave the first suggestion that the *repA* and *repC* genes were transcribed as one unit, since these authors could find no discernible *repA* transcriptional termination structure. Haring *et al.* (1985) also suggested that RepA and RepC were synthesized from a polycistronic mRNA, and *repA* translation was a prerequisite for optimal expression of *repC*. Since the 5'-end of the *repC* gene overlaps 14 nucleotides of the 3'-end of *repA*, and the polycistronic mRNA in this region is thought to be able to form a stable loop which encloses the RBS and the start codon of *repC*, the *repC* gene is believed to be regulated at both the transcriptional and translational level (Scholz *et al.*, 1989).

The *repA* sequence of RSF1010 coding for a 278 amino acid helicase was first published by Scholz *et al.* (1985). It was from this sequence that these authors were able to identify the possible RBS of *repA*. The purified 30 kDa RSF1010 RepA protein is one of the smallest known helicases, has an annular shape with 6-fold rotational symmetry, and was found to be active as a hexamer (Scherzinger *et al.*, 1997; Niedenzu *et al.*, 2001). The diameter of the central hole in the RepA hexamer was determined to be ~17 Å, which is too small to accommodate dsDNA, suggesting that only ssDNA can thread through the central hole in the RepA hexamer (Niedenzu *et al.*, 2001). This helicase belongs to the DnaB family of helicases, and is thus a member of the helicase superfamily 4 (SF4) (Patel and Picha, 2000; Niedenzu *et al.*, 2001).

Features of the RepA helicase include a type A nucleotide-binding motif (Walker A box) which explains the ATP-dependent ssDNA-binding activity, and ATPase



activity, of this protein (Scherzinger *et al.*, 1991; Scherzinger *et al.*, 1997). The RepA helicase also displays a 5'→3' polarity, with optimal unwinding activity at pH 5.5-6.0 (Scherzinger *et al.*, 1997). In the presence of Mg<sup>2+</sup> and Mn<sup>2+</sup> ions, RepA can be powered by either ATP, dATP, GTP or dGTP hydrolysis, and to a lesser extent by CTP or dCTP (Scherzinger *et al.*, 1997). This is unusual as some helicases show a preference for a single triphosphate (Scherzinger *et al.*, 1997). The ATP binding sites of the homohexamer are each located at the interfaces between two adjacent monomers. Four amino acids are catalytically active in binding ATP: Lys-43, Glu-77, Asp-140, and His-179 (Niedenzu *et al.*, 2001). Therefore, the ATP-binding site is defined by the consensus sequence for the phosphate-binding loop (P-loop) <sup>40</sup>GAGKS<sup>44</sup>, and residues Glu-77, Asp-140, His-179 (belonging to the same monomer) and Arg-207 (from the adjacent monomer). Residues from the P-loop contact all three phosphates of ATP, while the adenine base is locked between Arg-86 of a monomer and Tyr-243 of the adjacent monomer (Xu *et al.*, 2001). The Arg-207 residue is thought to act in *trans* from the adjacent monomer by contributing its functional side-chain to the ATPase site (Niedenzu *et al.*, 2001). Niedenzu *et al.* (2001) suggest that Arg-207 is suitably located to play a role similar to the “arginine finger” found in GTPase-activating proteins, and may modulate the ATPase activity of the adjacent monomer dependent on whether DNA is bound to its monomer. DNA binding is thought to occur at motif H4 where Arg-207 is located (Niedenzu *et al.*, 2001). This may have relevance to the three-site translocation model involving bind-change conformational states at each site that drives hexameric helicase translocation dependent on ATP binding and hydrolysis (see section 1.2.2).

However, unlike the physical interaction between DnaB and DnaG, there is no specific interaction between the RSF1010 RepA helicase and the RepB primase, although SSB protein does appear to stimulate RepA (Scherzinger *et al.*, 1997).

DNA synthesis is initiated at the oppositely orientated *ssi* sites in the *oriV* region by the RepB (RepB') primase of RSF1010 (Fig 2.1.2) (Scholz *et al.*, 1989). Lin and Meyer (1987) showed that R1162 DNA synthesis was initiated from these two sites found in the palindromic region of *oriV*. Honda *et al.* (1988) demonstrated that *ssiA* and *ssiB* were functionally independent, and this was further substantiated by the



findings of Zhou and Meyer (1990) who demonstrated that either *ssi* signal could substitute for the other on its particular strand. No apparent requirement was found for the distance between the two *ssi* signals, although deletion of a *ssi* signal resulted in plasmid instability and lowering of copy number, indicating that both *ssi* signals are required for optimal replication of R1162 (Zhou and Meyer, 1990). These sites were also found to contain a 40-bp homologous region in which a conserved part of the PriA recognition sequence was identified (Honda *et al.*, 1988). This 40-bp homologous region appears to form a stem-loop structure, and this secondary structure has been found to play a crucial role in RepB-recognition of the *ssi* signal (Honda *et al.*, 1993). Mutational analysis of the stem-loop subsections has revealed that the nucleotide sequence in the loop subsection is critical for *ssi* function, but that specific base-pairing interactions in the stem subsection was more important than nucleotide sequence. The distance between the stem substructure and the proposed DNA initiation site on the 3'-flanking end of the stem-loop was also found to be critical for efficient DNA synthesis initiation (Honda *et al.*, 1993; Miao *et al.*, 1993; Miao *et al.*, 1995[a]).

The RSF1010 primase occurs in two forms: a 78 kDa MobA-RepB fusion protein, and a smaller 36 kDa RepB' protein (Fig 2.1.1). The RepB' protein is translated from within the same reading frame as the fusion protein (Sakai and Komano, 1996; Rawlings and Tietze, 2001). Both RepB and RepB' are capable of primer synthesis at the RSF1010 origin, as deletions in the *repB* gene that left the open reading frame of *repB'* intact did not affect the ability of RSF1010 to replicate in *E. coli* (Haring and Scherzinger, 1988). A replication deficient M13 bacteriophage construct containing the RSF1010 *ssi* signals also showed this by demonstrating enhanced replication when RepB' was supplied *in trans* (Honda *et al.*, 1989). As the open reading frame specifying the MobA protein was found to be identical to that of the *repB* gene, it was proposed that RepB has two functional domains: a C-terminal primase domain, and a N-terminal domain active in plasmid mobilization. The MobA protein of RSF1010, in conjunction with MobC, forms a complex with the RSF1010 *oriT* region (Fig 2.1.1), and these two mobilization proteins are responsible for site and strand-specific nicking of the *oriT* (Henderson and Meyer, 1996). The MobA-linked primase is thought to be covalently bound to the 5'-end of the cleaved strand, and may therefore

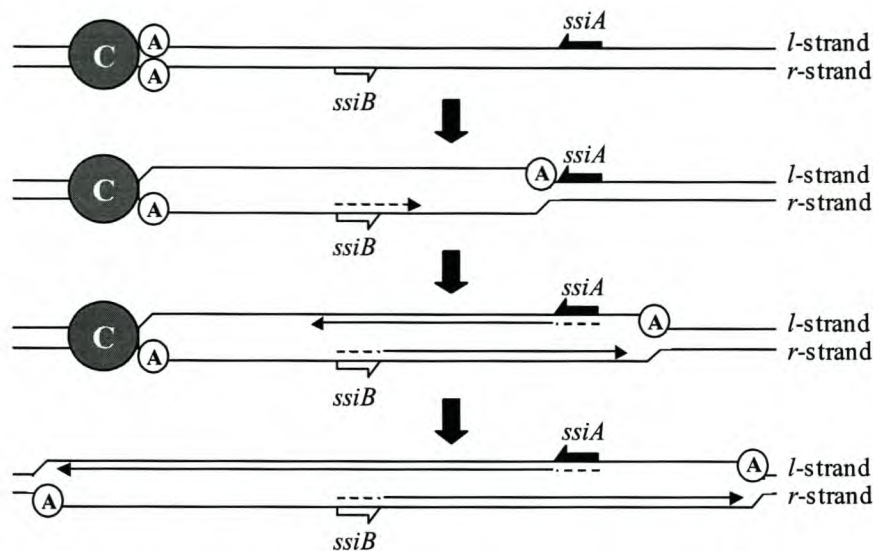


assist in conjugation by priming the complementary strand during strand transfer in the conjugation pore (Henderson and Meyer, 1999).

The priming reaction was found to be independent of the pre-priming events in which RepC brings about localized melting of the A+T-rich region for RepA entry. It is only after RepA helicase translocation has exposed the *ssi* sites that priming for complementary strand synthesis at *ssiA* and *ssiB* occurs (Honda *et al.*, 1991). The priming reaction is also unique in that the two oligomeric forms of the RSF1010 primase do not require ribonucleotide triphosphates for the priming reaction (Honda *et al.*, 1991).

The *in vitro* replication system employed by Scherzinger *et al.* (1991) appeared to show that RSF1010 DNA synthesis during replication occurred in a continuous manner. These authors used electron micrographs to demonstrate that replication initiation at each *ssi* site began with the formation of a D-loop, resulting from the synthesis of one complementary strand, and that the two displacing complementary strands converge on each other (Fig 2.1.3). It was the replacement of either one or both of the RSF1010 *ssi* sites with two distinct G4-type *ssi* signals that suggested the strand displacement mode of replication used by RSF1010 and related plasmids. The G4-type *ssi* signal of bacteriophage G4, and the *ssi* site from plasmid pSY343 require only *E. coli* DnaG (without the requirement of primosome assembly), and direct only the initiation of leading strand synthesis. Since either *ssiA* or *ssiB* could be functionally replaced with these G4-type priming signals, it was suggested that RSF1010 did not have a system for lagging-strand synthesis, and DNA synthesis from *ssiA* and *ssiB* occurred in a continuous manner independent of each other (Honda *et al.*, 1992). Tanaka *et al.* (1994) came to the conclusion, after deletion, inversion, or substitution of either *ssiA* or *ssiB* with G4-type and  $\phi$ X174 (primosome assembly site; PAS) *ssi* signals, that irrespective of the type of priming signal, two convergently orientated priming events for the initiation of leading-strand synthesis are essential for normal RSF1010 replication.





**Fig 2.1.3:** Model for the initiation of strand synthesis in the RSF1010 strand displacement replication mechanism. DNA synthesis from the two *ssf* sites (indicated) within RSF1010 *oriV* is initially primed at *ssiB* by RepB, and subsequently the two displacing leading strands (arrows) converge on each other. RepA proteins are represented by small open circles with letter A, and RepC proteins represented by large solid circles with letter C. Reproduced from Scherzinger *et al.* (1991).

There appears to be a difference in the functional significance of *ssiA* and *ssiB*. A mini-plasmid lacking *ssiA*, but still retaining a functional *ssiB*, was found to replicate in *P. aeruginosa*, whereas a functional *ssiA* could not substitute for the loss of *ssiB*. This suggested that *ssiB* was predominantly used by RSF1010 to replicate in this host (Higashi *et al.*, 1994). In other experiments, the introduction of a *Ter* site between the iterons and the *ssiB* site (Fig 2.1.2) inhibited R1162 replication by blocking RepA translocation towards the *ssi* sites. This showed that the replication fork moved preferentially towards the *ssi* sites (Zhou *et al.*, 1991). As replication fork movement away from the iteron region would expose the *ssiB* site first, these results were interpreted as the predominance of the *ssiB* site in the strand displacement mechanism to initially direct DNA synthesis away from this region (Fig 2.1.3).

Many years of investigation of the plasmid RSF1010/R1162/R300B have revealed the replication mechanism of this plasmid, and how this has contributed to the promiscuous ability of this plasmid. By encoding replication factors specific for the plasmid template the plasmid has evolved away from the reliance on host replication proteins, thereby overcoming host barriers and achieving a broad host range. An early observation was that RSF1010 had a greater number of essential genes than narrow host range plasmids, and at the time it was calculated that despite its relatively



small size, about 50% of the RSF1010 genome comprised essential genes (Scholz *et al.*, 1985).

This chapter describes the isolation of plasmid pTC-F14 from the biomining bacterium *At. caldus* strain “F”, and analysis of the plasmid replicon. The analysis involved the sequencing of the replicon and comparison of the replication proteins with known members of the IncQ family of plasmids. The copy number of pTC-F14 was also determined. As the IncQ family characteristically have a broad host range, the host range of this new replicon was also investigated.

## 2.2) Materials and Methods

**2.2.1) Bacterial strains, media and growth conditions.** The bacterial strains and plasmids used in this study are shown in Appendix B. *E. coli* DH5 $\alpha$  cells were grown at 37°C in Luria-Bertani medium (Appendix A), and ampicillin (100  $\mu\text{g}\cdot\text{ml}^{-1}$ ); kanamycin, (30  $\mu\text{g}\cdot\text{ml}^{-1}$ ); tetracycline (20  $\mu\text{g}\cdot\text{ml}^{-1}$ ) were added as required. *P. putida* was grown at 30°C in Luria-Bertani medium, and kanamycin (50  $\mu\text{g}\cdot\text{ml}^{-1}$ ) was added when appropriate. *Agrobacterium tumefaciens* LBA 4404 was grown at 28°C in Luria medium which included rifampicin (5  $\mu\text{g}\cdot\text{ml}^{-1}$ ), and when required, kanamycin (30  $\mu\text{g}\cdot\text{ml}^{-1}$ ). Tetrathionate medium used to culture *At. caldus* strain “f” was made from mineral salts solution ( $\text{g}\cdot\text{l}^{-1}$ ):  $(\text{NH}_4)_2\text{SO}_4$ , 3.0; KCl, 0.1;  $\text{K}_2\text{HPO}_4$ , 0.5;  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ , 0.5;  $\text{Ca}(\text{NO}_3)_2\cdot 4\text{H}_2\text{O}$ , 0.014;  $\text{Na}_2\text{SO}_4$ , 1.45. The pH adjusted to 2.5 with  $\text{H}_2\text{SO}_4$  and autoclaved. Trace element solution ( $\text{mg}\cdot\text{l}^{-1}$ ):  $\text{ZnSO}_4\cdot 7\text{H}_2\text{O}$ , 10.0;  $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ , 1.0;  $\text{MnSO}_4\cdot 4\text{H}_2\text{O}$ , 1.0;  $\text{CoCl}_2\cdot 6\text{H}_2\text{O}$ , 0.5;  $\text{Cr}_2(\text{SO}_4)_3\cdot 15\text{H}_2\text{O}$ , 0.5;  $\text{Na}_2\text{B}_4\text{O}_7\cdot 10\text{H}_2\text{O}$ , 0.5;  $\text{NaMoO}_4\cdot 2\text{H}_2\text{O}$ , 0.5, was acidified, autoclaved and 1 ml added per liter. Filter-sterilized  $\text{K}_2\text{S}_2\text{O}_6$  was added to a final concentration of 10 mM, and the pH adjusted to 2.5 with  $\text{H}_2\text{SO}_4$ . *At. caldus* was grown at 37°C with constant shaking.

**2.2.2) DNA techniques, sequencing and analysis.** Plasmid preparation, restriction endonuclease digestions, gel electrophoresis, ligations and Southern blot hybridization were carried out using standard methods (Sambrook *et al.*, 1989; Ausubel *et al.*, 1993). The polymerase chain reaction was performed in a PCR Sprint Temperature Cycling System (Hybaid) using the Expand High Fidelity PCR System DNA polymerase (Roche Molecular Biochemicals). After an initial denaturation of 60 s at 94°C, 25 cycles of 30 s at 94°C, 30 s at 63°C (for primers TACREPA and TACREPAE) and 90 s at 72°C were performed. A final extension step of 120 s at 72°C before cooling to 4°C completed the reaction. Sequencing was done by the dideoxy chain termination method, using an ABI PRISM™ 377 automated DNA sequencer, and the sequence was analyzed using a variety of software programs, but mainly the PC-based DNAMAN (version 4.1) package from Lynnon BioSoft. Comparison searches were performed using the gapped-BLAST program at the National Center for Biotechnology Information (NCBI) [<http://www.ncbi.nlm.nih.gov/BLAST/>] (Altschul *et al.*, 1997). Homology trees were



constructed using the Multiple Sequence Alignment tool in DNAMAN. The homology trees constructed by DNAMAN are setup with the distance matrix using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method. Protein motif signatures were detected using programs accessed through the Pôle Bio-Informatique Lyonnais (PBIL) Network Protein Sequence Analysis (NPS@) server [<http://npsa-pbil.ibcp.fr/>] (Combet *et al.*, 2000). The PBIL PROSCAN program was used to search a protein sequence for motif signatures against the PROSITE database, while searches for helix-turn-helix motifs were made using the PBIL HTH detection program (Dodd and Egan, 1990; Bairoch *et al.*, 1997; Combet *et al.*, 2000). The alignment of IncQ RepA sequences with the *E. coli* DnaB sequence was compiled with CLUSTALW (version 1.82) [<http://clustalw.genome.ad.jp/>] (Thompson *et al.*, 1994), and then edited with the PC-based Genedoc (version 2.6.002) software package [<http://www.psc.edu/biomed/genedoc/>]. The RSF1010 RepA and *E. coli* DnaB helicase alignment constructed by Patel and Picha (2000) was used as the reference against which the helicase alignment was edited.

**2.2.3) Harvesting bacteria and preparation of chromosomal DNA.** *At. caldus* cells were recovered by centrifugation, washed three times in acidified water (pH 1.8) and extraneous sulfur species were removed by a process of low- and high-speed centrifugation. Washed cell pellets were resuspended in TE (0.01 M Tris, 0.001 M EDTA)-0.15 M NaCl (pH 7.6) buffer. Cells resuspended in TE -NaCl (pH 7.6) buffer were used for the preparation of chromosomal DNA as described in Breed *et al.* (1999).

**2.2.4) Trans-Alternating Field Electrophoresis (TAFE).** Tetrathionate-grown cells were washed twice and resuspended in SET buffer (25 % sucrose, 2 mM EDTA, 50 mM Tris pH 8) to give an optical density at 600 nm of 1. The cells were set in agarose using an equal volume of 2 % LMP agarose (Seaplaque, FMC Bioproducts) in the presence of 1 mg.ml<sup>-1</sup> proteinase K. The plugs were incubated in ESP buffer (0.5 M EDTA pH 8, 1 % sodium lauroyl sarcosine, 1 mg.ml<sup>-1</sup> proteinase K) for 16 h at 50°C, repeated twice, to lyse the cells. Proteinase K was inactivated by incubation of the plugs in TE buffer containing 5 mM Pefabloc<sup>®</sup> (Roche Molecular Biochemicals) for 16 h at 4°C. The DNA-containing plugs were washed for 30 min at 4°C in 5 volumes of distilled water, pre-equilibrated in restriction buffer for 1 hour at



4°C, and digested according to the supplier's instructions, in three volumes of fresh restriction buffer containing restriction enzyme. TAFE was performed using a Beckman GeneLine™ apparatus. DNA fragments were separated in a 1 % agarose (SeaKem LE, FMC Bioproducts) gel at 150 mA and 12°C for 16 hours with a pulse interval of 13 s.

**2.2.5) Host range determination.** Electroporation of *P. putida* was performed with a Gene Pulser® electroporation apparatus (Bio-Rad laboratories) using the protocol described by Iwasaki *et al.* (1994). *P. putida* was grown to mid-log phase (Optical density at 600 nm, 0.4) at 30°C, harvested by centrifugation, and washed once in sterile cold water. The cells were then washed twice in sterile cold 300 mM sucrose (electroporation buffer). The washed cells were resuspended in the electroporation buffer, and 100 µl of this sample was placed in a pre-chilled electroporation cuvette with pTC-F101 plasmid DNA. The electrical settings were: voltage, 12.5 kV/cm; capacitance, 25 µF; and pulse controller parallel resistor, 200 Ω. Immediately after discharge, 900 µl Luria-Bertani medium was added to the electroporated cells, which were then incubated at 30°C for 2-3 hours prior to plating on selective medium. Tri-parental mating (Rawlings and Woods, 1985) was used to transform *A. tumefaciens* LBA 4404. Plasmid stability was assayed for by diluting an overnight culture to 10<sup>-6</sup>, and then growing it to saturation in 10 ml of Luria-Bertani medium without selection. A dilution series of the culture was then plated on selective and non-selective medium, and after the appropriate incubation time the percentage stability after 20 generations of growth was calculated. This procedure was repeated for 100 generations of growth.

**2.2.6) Plasmid copy number determination.** Estimation of plasmid copy number was performed in a slot-blot experiment by hybridizing plasmid DNA to known concentrations of total DNA isolated from plasmid containing cells and known amounts of purified plasmid DNA. By comparing the concentrations of total DNA and plasmid giving equivalent hybridization signals, the amount of plasmid in a sample of total DNA could be estimated. As the approximate sizes of the plasmid and chromosomal DNA are known it was possible to calculate the number of copies of plasmid per chromosome (Eqn 2).



$$\frac{\text{Plasmid size (kb) X Copy number}}{\text{Chromosome size (kb)}} = \frac{\text{Amount of plasmid DNA with signal equivalent to total DNA}}{\text{Amount of total DNA with signal equivalent to plasmid DNA}} \quad (2)$$

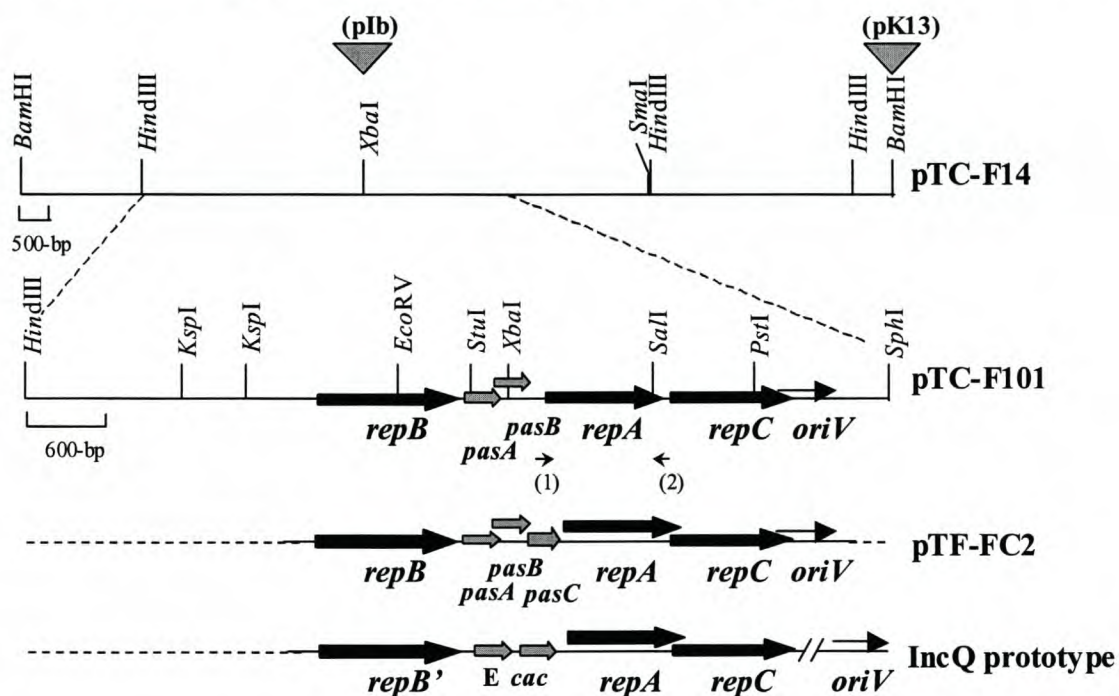
The PCR amplified *repA* fragment was labeled according to the digoxigenin-11-dUTP (DIG) random primed DNA labeling method using the DIG DNA labeling kit (Roche Molecular Biochemicals). A 1:2 serial dilution of DNA sample was vacuum-transferred onto Hybond<sup>TM</sup>-N<sup>+</sup> nucleic acid transfer membrane (AEC-Amersham) using a Schleicher and Schuell Minifold II micro-sample filtration manifold (Dassel, Germany). The starting concentration of DNA, prior to serial dilution, was based on the amount of DNA of each total genomic isolation estimated to give an equivalent signal to one plasmid copy. The DNA from the total genomic isolation and the plasmid isolation were diluted in denaturation solution (0.4 N NaOH, 0.01 M EDTA pH 8), heat denatured, and snap cooled on ice. The slot blot manifold was assembled after pre-wetting the membrane in 0.4 N NaOH, a vacuum applied, and the DNA serial dilution pipetted into the appropriate wells. With the vacuum still applied, two aliquots of 300 µl of 0.4 N NaOH were sequentially added to each well. The manifold was then dismantled and the membrane rinsed in 2X SSC (300mM NaCl, 30mM sodium citrate pH 7.0). Following this, the membrane was prehybridized with DIG Easy Hyb (Roche Molecular Biochemicals) solution for one hour at 40°C. After prehybridization, the denatured probe was added to the DIG Easy Hyb solution and hybridization was carried out at 40°C overnight. The membrane was then washed twice for ten minutes with 2X SSC containing 0.1 % SDS at room temperature. A second set of higher stringency washes using 0.1X SSC (15 mM NaCl, 1.5 mM sodium citrate pH 7.0) containing 0.1 % SDS were carried out at 65°C for ten minutes. After these washes the probe was detected using the DIG luminescent detection kit and protocol of Roche Molecular Biochemicals. The chemiluminescent substrate used was CSPD® (Roche Molecular Biochemicals) and the signal was detected by autoradiography using MG-SR X-ray film (Konica).

**2.2.7) Nucleotide sequence accession number.** The nucleotide sequence determined in this study has been deposited in the GenBank database under accession number AF325537.



## 2.3) Results

**2.3.1) Isolation of pTC-F14 and identification of the plasmid replicon.** *At. caldus* strain “f” was isolated from a pilot plant operating at 45°C used to treat a nickel concentrate situated at the Billiton Process Engineering Laboratory (Randburg, South Africa). Trans-alternating field gel electrophoresis (TAFE) of total *At. caldus* DNA indicated that strain “f” contained at least two plasmids; one of approximately 14-kb and one about 45-kb (not shown; S. Deane, per comm.). Restriction endonuclease mapping of the mixed plasmid preparation suggested that the smaller plasmid contained unique *Xba*I and *Bam*HI sites. These sites were used to clone the 14-kb plasmid, called pTC-F14, into the *E. coli* pBluescript(KS) cloning vector, using both the *Xba*I (plasmid pIb) and *Bam*HI (plasmid pK13) sites (Fig 2.3.1).



**Fig 2.3.1:** Restriction enzyme and genetic maps of the pTC-F14 replicon and subclones. Grey inverted triangles represent the sites into which pBluescript(KS) was cloned to construct pIb and pK13. Genes are shown as solid black or grey arrows, and the *oriV* region is indicated as a long thin arrow. The small numbered arrows represent the primer pair (Appendix C) used to amplify the *repA* gene ([1] TACREPA and [2] TACREPAE). The genetic map of pTF-FC2 (Dorrington *et al.*, 1991; Smith and Rawlings, 1997) and the IncQ prototype plasmid (RSF1010/R1162/R300B) (Scholz *et al.*, 1989) are provided for replicon structure comparison. Diagonal lines in the IncQ prototype map represent the distal location of the *oriV* from the *rep* genes.

The resulting clones were used to transform an *E. coli polA* mutant (GW125a), but only pK13 produced transformants. As ColE1-based cloning vectors are unable to replicate in an *E. coli polA* mutant strain, this result was interpreted as indicating that



pTC-F14 had a replicon which was capable of independent replication in *E. coli* and that disruption of the *XbaI* site had inactivated this replicon. This interpretation was confirmed by deleting the cloning vector inserted at the *BamHI* site and replacing it with a chloramphenicol resistance gene. The resulting construct (pTC-F14Cm) retained the ability to replicate in both the *E. coli polA* mutant and *polA* wild-type cells (S. Deane, per comm.). A 6.4-kb *HindIII-SphI* fragment spanning the *XbaI* site was ligated to a kanamycin resistance gene and the resulting construct (pTC-F101; Fig 2.3.1) was able to replicate in *E. coli*, confirming that cloning at the *XbaI* site had disrupted the pTC-F14 replicon.

**2.3.2) Comparative analysis of the replication region.** The 4-kb region incorporating the *rep* genes and *oriV* was sequenced in both directions (Appendix E). The G+C mole % ratio of this region is 60 % which is typical for IncQ and IncQ-like plasmids (59-62 %) (Rawlings and Tietze, 2001). It contains three open reading frames with high predicted amino acid sequence identity to the RepA, RepB and RepC proteins of other IncQ-like plasmids (Fig 2.3.2).

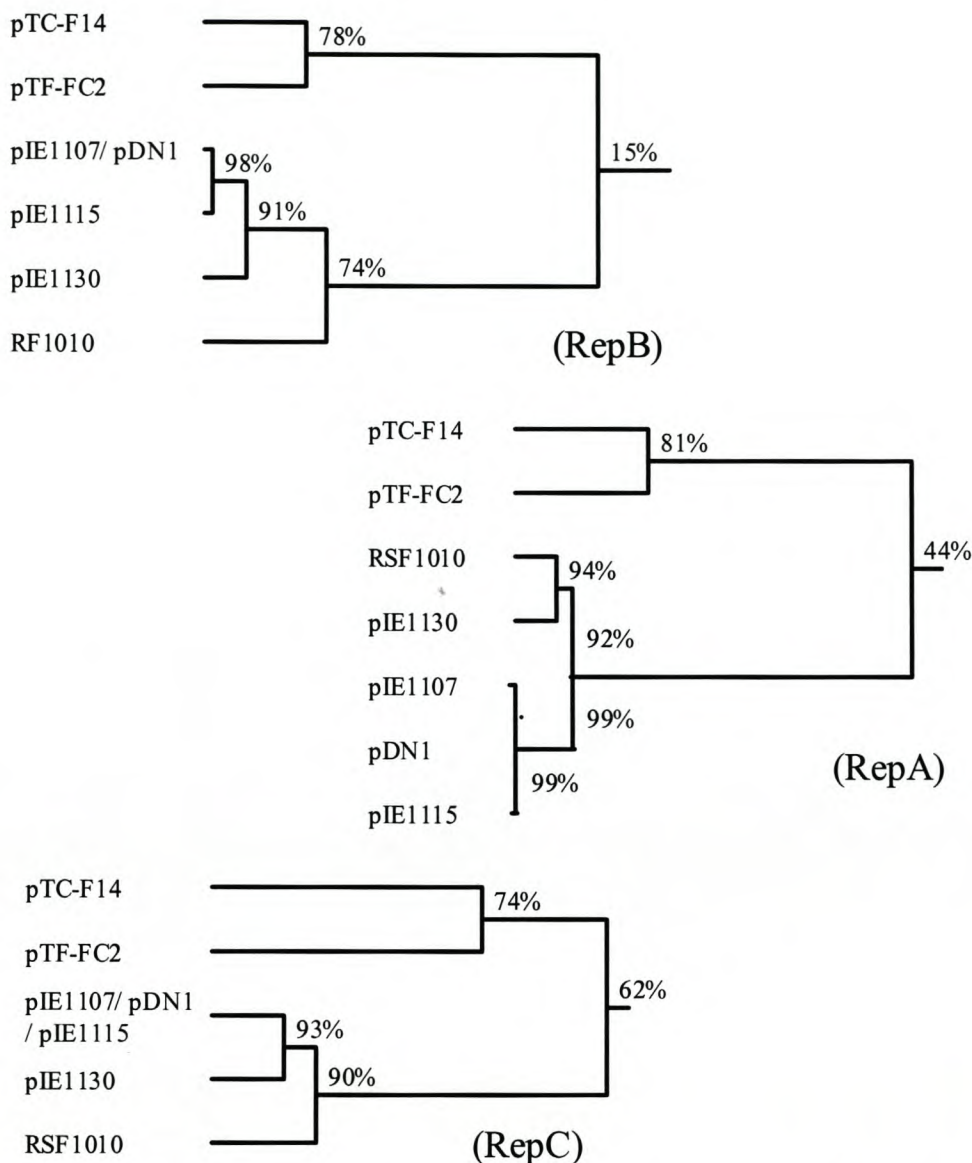
The highest similarity was to the RepA, RepB and RepC proteins of the *At. ferrooxidans* plasmid, pTF-FC2 (Fig 2.3.2 and Table 2.3.1). Two small open reading frames encoding proteins with high amino acid sequence identity to the PasA and PasB proteins of pTF-FC2 (Smith and Rawlings, 1997) were located between the *repB* and *repA* genes (Fig 2.3.1 and Table 2.3.1). The *pas* genes of pTF-FC2 have been shown to function as a plasmid addiction system which enhances plasmid stability through post-segregational killing of plasmid-free cells (Smith and Rawlings, 1997).

**Table 2.3.1:** Comparison of replicon-associated proteins of pTC-F14 and pTF-FC2.

Protein	Translational initiation sites*	pTC-F14			pTF-FC2			% amino acid identity
		Amino acids	Mol mass (Da)	pI	Amino acids	Mol mass (Da)	pI	
RepA	UCUGGAA <b>AGG</b> GAGAACAGCAUG	291	31289	5.92	290	31227	6.21	81.0
RepB	GC <b>AGG</b> AGAGGGCACAGCGAUG	352	40622	9.84	352	40111	9.77	78.2
RepC	UACCCAG <b>GG</b> AGGCAAGCCAUG	303	33811	9.46	299	33740	8.99	74.2
PasA-like	UUUGAGC <b>AGG</b> AGCUAAACAUG	74	8523	4.46	74	8453	4.71	81.1
PasB-like	AGGAAGUG <b>G</b> AGCGCAUCUUG	90	10483	10.36	90	10307	10.4	72.2

\*Putative ribosome binding sites are shown in bold and start codons in italics





**Figure 2.3.2:** Relationships between the replicons of plasmids of the IncQ-like family based on percent amino acid identities of the replication proteins. Accession numbers are as follows: pTC-F14, AF325537; pTF-FC2, M35249 and M64981; pIE1107, ACZ74787; pIE1115, AJ293027; pIE1130, AJ271879; RSF1010, M28829; pDN1, ACY19120.

The region of pTC-F14 which was required in *cis* for the plasmid to replicate was situated on a 716-bp fragment between the *Pst*I and *Nco*I sites (Fig 2.3.3). This *oriV* region contained five iterons, with the central three 22-bp iterons being perfectly conserved, while flanking iterons had either a single base pair insertion or a single base pair deletion (Fig 2.3.3). The iterons overlapped with the predicted C-terminus of the RepC protein. IncQ plasmids are characterized by having a G+C-rich region (24/28-bp) about 25-bp downstream of the iterons (Rawlings and Tietze, 2001). This is followed by an A+T-rich region (23/31-bp) which is believed to be the site of DNA

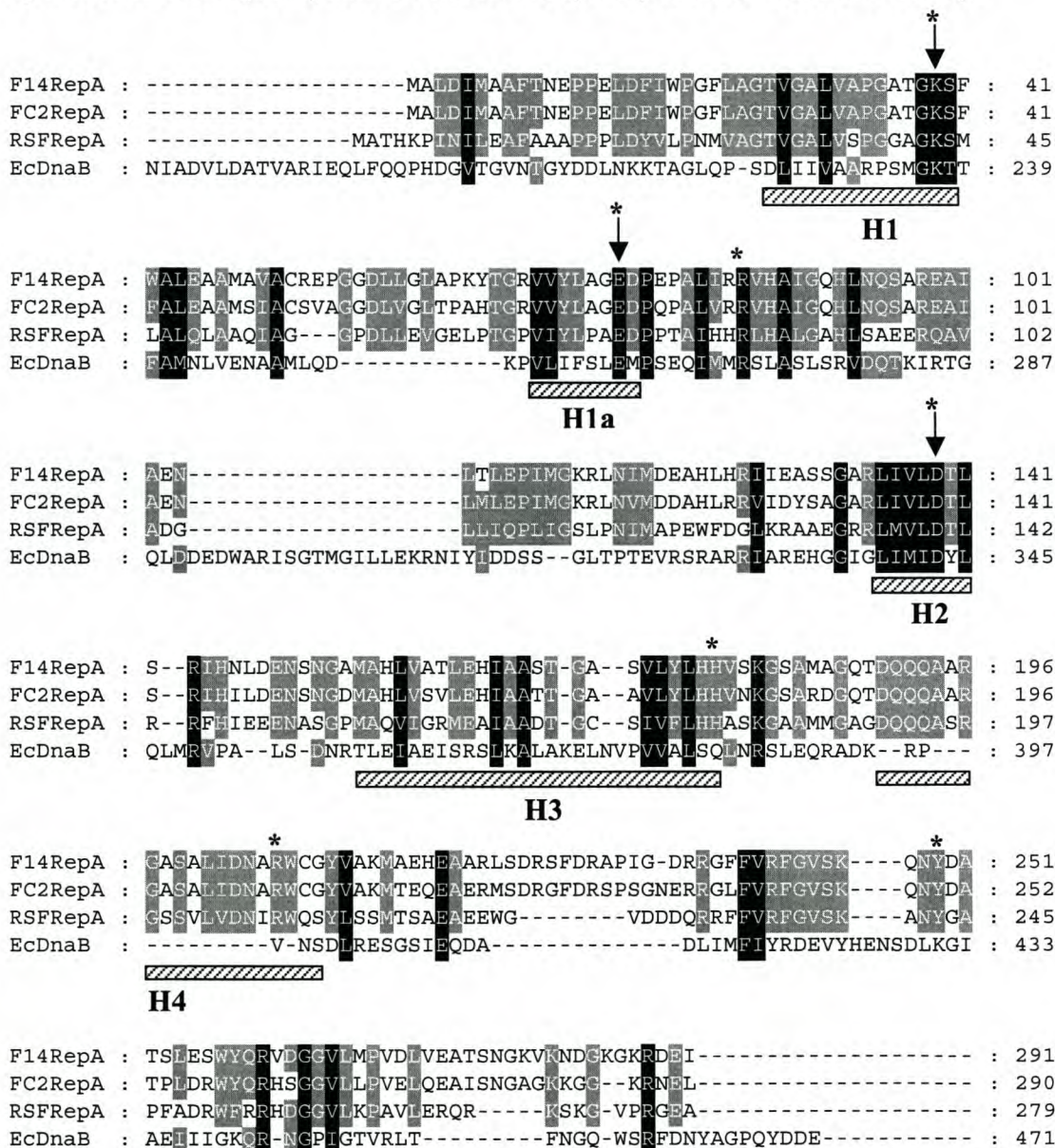






inverted or direct repeats plays a role in the initiation of pTC-F14 replication is unknown.

Alignment of the IncQ and IncQ-like RepA proteins with the *E. coli* DnaB helicase amino acid sequence (Fig 2.3.4) revealed significant similarity to features identified by Patel and Picha (2000) and Niedenzu *et al.* (2001) in the SF4 helicase family.



**Fig 2.3.4:** Alignment of pTC-F14 (F14RepA; AF325537), pTF-FC2 (FC2RepA; M35249), and RSF1010 (RSFRepA; M28829) RepA primary sequences with the *E. coli* DnaB helicase (EcDnaB; P03005). The five conserved regions of SF4 helicases as determined by Niedenzu *et al.* (2001) are shown by labelled hatched rectangles. Residues shaded in black are conserved, while those in grey represent an approximate 80% similarity. Arrows above the H1, H1a and H2 motifs indicate highly conserved amino acids as identified by Patel and Picha (2000). Stars above amino acids indicate catalytically important residues identified in RSF1010 RepA (Niedenzu *et al.*, 2001; see section 2.1).



The apparent H1 motif in pTC-F14 contained the highly conserved and essential lysine of the Walker A box, and the conserved aspartic acid for the Walker B box is present in the H2 motif. The four amino acids found in motifs H1, H1a, H2 and H4 of RSF1010 RepA (Niedenzu *et al.*, 2001), thought to bind the three phosphates of ATP, were identified in pTC-F14 RepA. The arginine and tyrosine residue thought to bind the adenine nucleoside of ATP (Arg-86 and Tyr-243 of RSF1010 RepA) are also identifiable in the pTC-F14 repA sequence. The counterpart to RSF1010 RepA Arg-207 is present in pTC-F14 RepA.

**Table 2.3.2:** Motif signatures identified in pTC-F14 replication proteins.

Protein	PROSITE Entry		PROSCAN result		
	Motif	Pattern signature	Query sequence signal*	Randomized probability	% pattern Similarity
RepB	Copper/Zinc superoxide dismutase signature 1	[GA]-[IMFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGDE]	<sup>151</sup> AIHpHRAPGYE <sup>161</sup>	4.392x10 <sup>-8</sup>	91
	Zinc carboxy-peptidases, zinc-binding region 1 signature	[PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-[LIVMFYTA]	<sup>158</sup> PGYEnRKPkHrR EDgSYPEVRLL <sup>180</sup>	1.141x10 <sup>-7</sup>	75
RepA	ATP/GTP-binding site motif A (P-loop)	[AG]-x(4)-G-K-[ST]	<sup>33</sup> APGATGKS <sup>40</sup>	7.781x10 <sup>-5</sup>	100
	DEAH-box subfamily ATP-dependent helicases signature	[GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR]	<sup>114</sup> rLnIMDEAHI <sup>123</sup>	1.916x10 <sup>-8</sup>	75
RepC	ATP/GTP-binding site motif A (P-loop)	[AG]-x(4)-G-K-[ST]	<sup>221</sup> GWIDPGKS <sup>228</sup>	7.781x10 <sup>-5</sup>	100
	Bacterial histone-like DNA-binding proteins signature	[GSK]-F-x(2)-[LIVMF]-x(4)-[RKEQA]-x(2)-[RST]-x-[GA]-x-[KN]-P-x-T	<sup>178</sup> IFVALNPRLAEAi IGArPHT <sup>197</sup>	5.069x10 <sup>-9</sup>	70
	DNA polymerase family A signature	R-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA]	<sup>32</sup> RGErKKLKLdVTY TYGKDRV <sup>51</sup>	7.028x10 <sup>-8</sup>	82

\*Letters in lower case do not match the PROSITE pattern signature, and superscript numbers indicate the amino acid sequence relative to the query input sequence

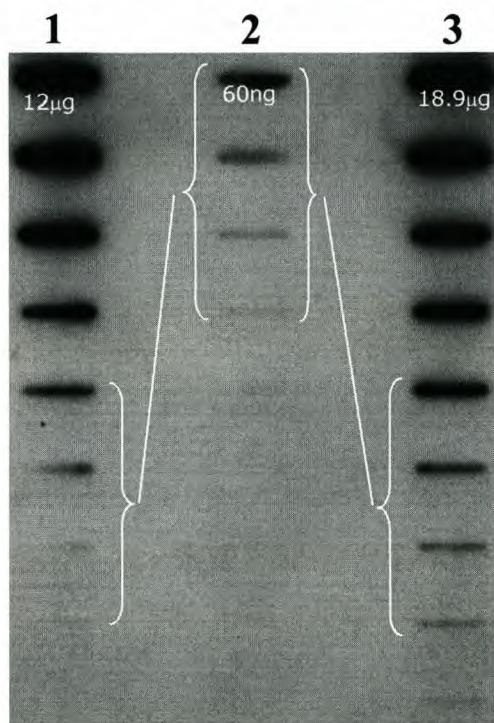


The results for motif signature searches of pTC-F14 replication proteins using the PROSITE database are shown in Table 2.3.2. The most significant signal was the Walker A box (P-loop) found in the RepA and RepC proteins. The Walker A box suggests the presence of a nucleoside-5'-triphosphate binding site. The DEAH-box signature was also identified in pTC-F14 RepA. The DEAH-box signature is usually found in SF4 helicases, and is typically associated with a Walker A box. It is interesting that Patel and Picha (2000) do not recognise this amino acid pattern as being part of the Walker B signal. Found in the RepB sequence were signatures for two zinc-binding proteins, with the histidine residue in each pattern being zinc ligands. The possible ZF-motif in DnaG is thought to prevent disulfide formation, and possibly play a role in priming site recognition (Frick and Richardson, 2001). Two DNA-binding protein motif signatures were detected for the pTC-F14 RepC protein (Table 2.3.2), but no HTH signatures were identified in any of the pTC-F14 replication proteins by bioinformatic analysis.

**2.3.3) Plasmid pTC-F14 appears to have a broad host range.** As plasmid pTC-F14 was isolated from *At. caldus*, and pTC-F14 with a chloramphenicol marker (pTC-F14Cm) was capable of replication in *E. coli*, it appears that like its IncQ relatives, pTC-F14 has a broad host range replicon. To obtain additional evidence for the broad host range property of the pTC-F14 replicon, pTC-F101 (Fig. 2.3.1) was transformed into *P. putida* and *A. tumefaciens* LBA 4404 using electroporation and triparental mating, respectively. The presence of the pTC-F14 replicon was confirmed by using the polymerase chain reaction (PCR) and primers specific for the *repA* gene of pTC-F14 (Appendix C; not shown). The identity of the transformants was confirmed by API 20E (bioMérieux sa, Marcy-l'Etoile, France) strips. Based on the premise that the stability of a plasmid in the absence of selection is a more rigorous test for the suitability of a plasmid to replicate in a host cell, the presence of pTC-F101 was tested for after 100 generations of *E. coli* DH5 $\alpha$ , *P. putida* and *A. tumefaciens* growth in the absence of selection. No plasmid loss was detected in *E. coli* DH5 $\alpha$  or *P. putida*, but after 60 generations of *A. tumefaciens* LBA 4404 growth, the plasmid had been lost from the population.



**2.3.4) Copy number of pTC-F14.** We wished to determine the copy number of pTC-F14 in both *E. coli* and the original *At. caldus* strain “f” from which it was isolated. Plasmid copy number was determined by the Southern hybridization technique using a genome size of 4.7-mb for *E. coli* (<http://www.tigr.org>) and 2.8-mb for *At. caldus*. The genome size of *At. caldus* was estimated by digesting chromosomal DNA from several strains with the relatively rare cutting restriction endonucleases *Xba*I and *Dra*I, followed by separation of the restriction fragments using a TAFE pulse field gel apparatus and summation of the sizes (not shown, S. Deane, per comm.). By using the relative sizes of the plasmid and chromosomes as well as the quantities of plasmid and total DNA loaded onto the hybridization membrane, the copy number of pTC-F14 was calculated (Eqn 2) to be between 12-17 copies per chromosome for both *E. coli* and *At. caldus* (Fig 2.3.5). This is approximately the same as the 12-15 copies (Dorrington and Rawlings, 1989) and 10-14 copies (Smith and Rawlings, 1997) estimated for plasmid pTF-FC2 in *E. coli*.



**Fig 2.3.5 :** Slot blot for pTC-F14 copy number calculation in *At. caldus* strain “f” ( lane 1) and *E. coli* DH5α (lane 3). A 1:2 dilution series using a known amount of total genomic DNA (lanes 1 & 3) and pTC-F14Cm plasmid DNA (lane 2) was hybridized to the membrane, and probed with pTC-F14 *repA* DNA. The values below the signals in the first row represent the amount of DNA in that slot.



## 2.4) Discussion

The identical, or nearly identical, IncQ prototype plasmids RSF1010, R1162, and R300B were discovered in different hosts that are commonly associated with clinical environments. The IncQ-like plasmids pIE1107, pIE1115 and pIE1130 were isolated from bacteria present in pig manure (Tietze, 1998; Smalla *et al.*, 2000[a]), while pDN1 was isolated from the sheep foot rot-causing pathogen *Dichelobacter nodosus* (Whittle *et al.*, 2000). Therefore, the discovery of a 14-kb plasmid maintained in the chemolithoautotrophic *At. caldus* strain “f” represents only the second IncQ-like plasmid isolated from non-human or non-animal associated environments; the other being the IncQ-like plasmid pTF-FC2 isolated from the chemolithoautotrophic *At. ferrooxidans* (Dorrington and Rawlings, 1989).

The demonstration that pTC-F14 was able to replicate in an *E. coli polA* mutant (GW125a), and sequencing data that revealed replication proteins with significant amino acid identity to those of the IncQ family of plasmids (Fig 2.3.2), indicated that pTC-F14 had an IncQ-like replicon. The highest amino acid identity of all three pTC-F14 replication proteins was to the replication proteins of the IncQ-like plasmid pTF-FC2, but they are also clearly related to the IncQ prototype replication proteins. Hence, they were designated accordingly: RepB, RepA, and RepC.

Amino acid alignments of IncQ and IncQ-like replication proteins appear to divide the IncQ family into two major groups (Fig 2.3.2). The plasmids isolated from chemolithotrophic bacteria form one group, while plasmids isolated from human or animal associated environments cluster in the other. Even though pTC-F14 and pTF-FC2 were isolated from the same environment, and pTF-FC2 is the closest relative of pTC-F14, the amino acid homology trees show that the amino acid sequence of the pTC-F14 and pTF-FC2 replication proteins are, on average, more dissimilar than the replication proteins of human or animal associated IncQ plasmids which cluster in the second branch (Fig 2.3.2). With reference to Fig 2.3.2, the RepC DNA-binding proteins are the most highly conserved of the replication proteins, with a 62 % amino acid identity between the two groups. The RepB' primases are the least conserved (15 %) between the two groups, but have a  $\geq 74$  % conservation within the groupings.



The sequence of pTC-F14 has been completed (G. Goldschmidt and L. van Zyl, per comm.; Appendix D [Fig C]) and it has been found that like RSF1010, the RepB protein sequence is in the same reading frame as MobA, and is likely to be translated as a larger MobA-RepB fusion protein. An internal RBS for the RepB protein has been identified within the sequence for the fusion protein (Table 2.3.1; Appendix E).

The replicon structure of pTC-F14 was found to more closely resemble the replicon of pTF-FC2 than the RSF1010/R1162/R300B IncQ prototype plasmids (Fig 2.3.1). Like pTF-FC2, but unlike the IncQ prototype plasmids, the C-terminus of the RepC coding region terminates in the *oriV* region, and the *oriV* region is proximally located to the *repC* gene (Fig 2.3.1). Plasmid pTF-FC2 is unique in having three open reading frames, found between the *repB* and *repA* genes, coding for a plasmid addiction system (*pas*) that enhances plasmid stability through postsegregational killing of plasmid-free cells (Smith and Rawlings, 1997). Located between *repB* and *repA* of pTC-F14 are two open reading frames whose proteins have clear amino acid identity to the PasA (81 %) anti-toxin and PasB (72 %) toxin, but pTC-F14 lacks the *pasC* gene found in the pTF-FC2 *pas* operon. The product of the *pasC* is thought to enhance the ability of the anti-toxin (PasA) to neutralize the toxin (Smith and Rawlings, 1997). The *pas* operon is at the same locus on the IncQ prototype replicon that the two open reading frames ORF E and ORF F are found (Scholz *et al.*, 1989). The function of the product of ORF E is unknown, but ORF F appears to encode a repressor that autoregulates the *repAC* operon (Maeser *et al.*, 1990). A mutant pTF-FC2 replicon was reported by Smith and Rawlings (1997) in which a spontaneous deletion of the pTF-FC2 *pas* operon had occurred. This spontaneous deletion apparently did not affect copy number suggesting that the pTF-FC2 *pas* operon was not involved in replicon regulation. Although not a focus of this study, a possible regulatory role of the pTC-F14 *pas* operon was, nevertheless, briefly investigated (see Chapter 4).

Motif searches of the pTC-F14 replication protein sequences revealed motif signatures for the RepC and RepB proteins that may have some relevance to the function of these proteins (Table 2.3.2). As the RepC protein of IncQ plasmids binds the *oriV* iterons, motif signatures of proteins that bind DNA were not unexpected.



However, the relevance of these signatures to the known function of RepC is questionable. Reasonable correspondence to motifs that bind zinc were obtained within the RepB protein, but without experimental evidence, the suggestion that these are part of a ZF-like motif is purely speculative. The more in-depth analysis of RSF1010 RepA by Scherzinger *et al.* (1997) and Niedenzu *et al.* (2001) has provided information on the secondary and tertiary structure of this helicase. The RSF1010 RepA helicase was classified as a member of the DnaB family of helicases, and therefore belongs to the SF4 group of hexameric helicases (Niedenzu *et al.*, 2001). Alignment with the *E. coli* DnaB and RSF1010 RepA helicase has highlighted critical residues conserved in RepA of pTC-F14 and pTF-FC2 (Fig 2.3.4). These conserved residues appear to be catalytically essential to the RepA helicases, and thus less likely to be mutated during protein evolution. This may be reflected in the 81 % amino acid identity between pTC-F14 and pTF-FC2 RepA, which makes RepA the most highly related replication protein of these two plasmids (Fig 2.3.2; Table 2.3.1).

Analysis of the *oriV* region of pTC-F14 revealed three perfectly conserved 22-bp direct repeats flanked by two imperfect repeats having either a base insertion or deletion (Fig 2.3.3). This is unlike the pTF-FC2 iteron region which only has three 22-bp iterons (Dorrington and Rawlings, 1989), and RSF1010 which has 3½ 20-bp iterons separated by 2-bp spacers (Scholz *et al.*, 1989). Thus, the pTC-F14 iterons conform to the notion that these repeated sequences generally occur as multiples of 11 which, being close to the helical periodicity of the DNA double helix, may align the RepC proteins on the same face of the DNA helix (del Solar *et al.*, 1998). The copy number of some iteron-containing replicons has been shown to be increased or decreased by the deletion or addition of iterons, respectively (Thomas *et al.*, 1981; Tsutsui *et al.*, 1983; Thomas *et al.*, 1984). As pTC-F14 has two iterons more than pTF-FC2, the copy number of pTC-F14 was determined. The estimated copy number of plasmids pTC-F14 and pTF-FC2 appears to be similar and within the range of 12-15 copies per chromosome reported for IncQ plasmids (Frey and Bagdasarian, 1989). In a study of the effect of mutations on the functionality of RSF1010 iterons it was found that even a single base pair replacement in one of the iterons could result in a non-functional iteron and the inability of RSF1010 to replicate (Miao *et al.*, 1995). It is, therefore, possible that the 23 or 21-bp iterons which flank the 22-bp iterons may be non-functional and not active in copy number control.



Although sequence analysis of the pTC-F14 replicon did not show any palindromic sequences with clear similarity to the *ssi* signals found in the RSF1010 *oriV* region, a 12-bp inverted repeat that could form a putative stem-loop structure was identified (Fig 2.3.3). It is unknown as to whether this is the site of complementary strand initiation.

Plasmid pTC-F14 does have an apparent broad host range, being able to replicate in representatives of the  $\alpha$ - (*A. tumefaciens*),  $\beta$ - (*At. caldus*), and  $\gamma$ - (*E. coli* and *P. putida*) subdivisions of the proteobacteria. Although the tested host range of pTC-F14 is limited, there is no reason to believe that it would be any different from its closest relative, pTF-FC2. The IncQ-like plasmid pTF-FC2 has been found to replicate in *K. pneumoniae*, *P. aeruginosa*, *P. putida*, *S. enterica* serovar Typhimurium, *Sinorhizobium meliloti*, *T. novellus*, *Acidiphilum facilis*, *Myxococcus xanthus*, and *Pseudomonas fluorescens*, in addition to those tested as part of this study (Rawlings and Tietze, 2001). However, pTC-F14 was lost from *A. tumefaciens* LBA 4404 after 60 generations. This maintenance instability has also been reported for the IncQ-like plasmid pIE1107 which was lost from *A. tumefaciens* DSM 30150 after 30 generations (Smalla *et al.*, 2000[a]).

Therefore, the isolation of plasmid pTC-F14 from the chemolithoautotroph *At. caldus* strain “F”, with significant similarities to the RSF1010 replicon and its replicative ability, suggests that pTC-F14 is a new member of the broad host range IncQ family of plasmids, and this warranted further investigation into its biology.



## Chapter Three

### Incompatibility and complementation of the pTC-F14 replicon

#### Contents

3.1)	Introduction	102
3.2)	Materials and Methods	108
	3.2.1) Bacterial strains, bacteriophage, media and growth conditions	108
	3.2.2) DNA techniques, sequencing and analysis	108
	3.2.3) Incompatibility assay	108
	3.2.4) Complementation assay	109
	3.2.5) Primase assay	109
3.3)	Results	111
	3.3.1) IncQ and IncQ-like plasmids share iteron sequence similarity	111
	3.3.2) Incompatibility among IncQ and IncQ-like plasmids	111
	3.3.3) Identifying the IncQ plasmid locus expressing incompatibility to pTC-F14	113
	3.3.4) Complementation of pTC-F14 replication proteins by IncQ and IncQ-like plasmids	114
	3.3.5) pTC-F14 RepB can prime <i>oriV</i> <sub>pTC-F14</sub>	117
3.4)	Discussion	119
	3.4.1) Division of the IncQ plasmid group	119
	3.4.2) Evolution of new incompatibility groups	120
	3.4.3) Interference with pTC-F14 replication by Rep proteins of the IncQ prototype plasmid	123



### 3.1) Introduction

The term 'plasmid backbone' delineates those genes typically required for plasmid transfer, maintenance and replication functions, from non-essential accessory genes which are generally considered to contribute to the phenotype of the host, and may encode genes for resistance, or catabolic products (Osborn *et al.*, 2000; Smalla *et al.*, 2000[b]). The plasmid backbone is not a rigid defining term as not all plasmids encode transfer and/or maintenance functions. Multiple horizontal DNA transfer events such as recombination and transposition have contributed to the usually different lineage of accessory DNA from the plasmid backbone, but this is not to say that functions within the plasmid backbone themselves do not have differing lineages. It is this mosaic composition of plasmids that has hampered the establishment of a satisfactory plasmid classification system. Therefore, it has been argued that plasmid classification should be based on genetic traits that are universally present. Hence, the replicon becomes the defining feature of a plasmid upon which a classification system will have to be based.

Early attempts at classifying plasmids were based on a plasmid's ability to inhibit F plasmid transfer, and on the plasmid's conjugative ability (Smalla *et al.*, 2000[b]). It had been found that some R plasmids, when present in the same host cell as the F plasmid, were able to inhibit its transfer. This difference was used as a means to divide plasmids into those that were classified  $fi^+$  (fertility inhibition plus) and those that were  $fi^-$  (Couturier *et al.*, 1988). A correlation was also found between the  $fi$  status of a plasmid and the type of sex pili produced (Meynell *et al.*, 1968). The discovery of plasmids that were not self-transmissible, and the identification of new types of pili, proved the inadequacy of this classification system.

Incompatibility group assignment of plasmids was introduced by Datta and Hedges (1971) as a formal scheme to try and address the problem of plasmid classification. It had been recognized that plasmids with the same, or significantly similar, replicons were incompatible with each other, while plasmids with dissimilar replicons were usually compatible. Therefore, plasmid incompatibility was defined as the failure of two co-resident plasmids to be stably inherited in a cell line in the absence of external selection (Novick, 1987). Plasmid incompatibility results in the segregational



instability of a co-resident replicon in a cell line as direct consequence of one of two properties: (i) interference with copy number self-correction by an isologous replicon (homologous plasmid backbone) competing for replication factors, or (ii) the inability to distinguish between two plasmids on partitioning into progeny cells. Plasmids which are incompatible are therefore assigned to the same incompatibility (Inc) group.

The plasmid copy pool in a host cell containing two isologous replicons is considered to be the same size as that of either replicon alone (Novick, 1987). The two plasmids, therefore, share the same pool, and the model to explain replicon-associated incompatibility is that either plasmid may be chosen at random from the pool for replication. This random choice may result in one plasmid being replicated more frequently than the other. If the copy number of either plasmid falls to one, it will be lost on cell division, irrespective of whether it has a unique partitioning system or not. Furthermore, interference with a replicon's copy number control mechanism may allow one of the competing replicons to achieve a numerical copy advantage that ensures its segregational inheritance over that of the inhibited replicon. A correlation has frequently been found between DNA sequence homology in copy number regulators, and plasmid incompatibility (Smalla *et al.*, 2000[b]).

However, plasmid classification by incompatibility testing is not without technical and methodological complications that have led to suggestions of alternate classification methods. Some technical problems that complicate incompatibility testing include the lack of a suitable selectable marker on a plasmid isolate, the narrow host range of some plasmids, and partitioning functions encoded by the plasmid which could assist inheritance (Smalla *et al.*, 2000[b]).

Methodological problems for incompatibility testing arise from the kinds and number of copy number control mechanisms, and other incompatibility determinants present on a plasmid (Couturier *et al.*, 1988). In inhibitor-target copy number control mechanisms used, for example, by plasmids ColE1, pMV158 and R1 (see Chapter 1), a high potential for genetic variation exists that could lead to new incompatibility groups. As the ctRNA species and the target are transcribed from the same DNA segment (in opposite directions), any base-pair changes in this DNA sequence will lead to complementary changes in both the ctRNA and the target RNA. Of particular



importance are the mutations which occur in the loop of the stem-loop structure where initiation of sense-antisense RNA pairing occurs. These complementary mutations preserve the pairing of the inhibitor and target RNA, thus maintaining the specificity of the inhibitor for its target, but releases the replicon from replication inhibition by the ctRNA of the parental plasmid. Consequently, evolution of new incompatibility groups can rapidly occur by point mutations in the replication control circuit (Novick, 1987; Couturier *et al.*, 1988, Osborn *et al.*, 2000). Couturier *et al.* (1988) aligned the nucleotide sequence of ctRNA species from plasmids of different incompatibility groups to show the relatively minor changes in the ctRNA consensus sequence which are considered to have led incompatibility testing to assign the plasmids to different groups. Therefore, these authors argue that closely related replicons had been assigned to different incompatibility groups, and this had unnecessarily complicated the classification system.

Another recognized difficulty in incompatibility group classification of plasmids is the multi-replicon composition of some plasmids. A multi-replicon plasmid could conceivably be maintained in a host cell line by a secondary compatible replicon when otherwise it would have been lost due to its incompatible primary replicon (Smalla *et al.*, 2000[b]). An example of a multi-replicon plasmid is the well studied F plasmid which has three independent replicons, designated FIA (the dominant replicon), FIB, and FIC, each specifying different incompatibility groups (Espinosa *et al.*, 2000). The IncQ-like plasmid pIE1107 provides a relevant example of a two-origin plasmid expressing two different incompatibilities. The pIE1107 plasmid is incompatible with the IncQ prototype plasmid RSF1010, but if the non-functional *oriV*-like region is deleted, the plasmid derivative, pIE1108, is compatible with RSF1010 (Tietze, 1998).

With the introduction of molecular techniques, new tools have been applied to plasmid classification in an attempt to better define, and provide a more reliable basis for, plasmid classification. However, these molecular tools are not without their own inherent problems.

Couturier *et al.* (1988) proposed the use of a range of replicon specific probes for the classification of plasmids. Replicon typing, as this method has become known,



involves the genotypic classification of plasmids based on colony, dot or Southern hybridization signals obtained with a collection of probes constructed from representatives from known plasmid groups. These probes were generally constructed using at least one of the incompatibility determinants associated with replication or partitioning functions. These authors allowed for a 15-20 % hybridization mismatch so that a family of replicons would include replicons with an approximate 80 % similarity to part or all of the DNA fragment used as a probe. This collection of *inc/rep* probes, as they are referred to, have been extended over the years to include new plasmid groups (Smalla *et al.*, 2000[b]).

However, application of replicon typing has exposed the limitations of this technique. By Southern hybridization using 14 *inc/rep* probes, Sobecky *et al.* (1997) attempted to replicon type 297 plasmids isolated from marine aerobic heterotrophic bacteria. None of the 14 *inc/rep* probes (4 broad host range and 10 narrow host range replicon probes) hybridized to any of the plasmids. Even lowering the stringency of the probes to allow for a <75 % similarity to the probe sequence did not reveal any significant homology to any of the *inc/rep* probes used (Sobecky *et al.*, 1997). Similarly, Kobayashi and Bailey (1994) reported that 68 % of plasmids isolated from bacteria in the sugar beet phyllosphere did not hybridize to the *inc/rep* probes. These two selected cases illustrate that *inc/rep* probes developed for plasmids from clinical isolates might fail to identify plasmids in environmental bacteria other than *Enterobacteriaceae*. Replicon typing by hybridization has also failed to identify relationships between plasmids when DNA sequencing has clearly shown them to share a common evolutionary history. This was found for the IncN-like plasmid pGSH500 which did not hybridize to the Couturier IncN probe (Da Silva-Tatley and Steyn, 1993; Smalla *et al.*, 2000[b]).

Used on its own as a plasmid classification method, restriction fragment length polymorphism (RFLP) analysis provides no information on DNA homology or plasmid incompatibility. This problem was exposed in a study by Lilley *et al.* (1996) in which RFLP analysis was solely used to classify plasmids isolated from the phyllosphere and rhizoplane of sugar beets. Yet, RFLP analysis has solved inconsistencies in incompatibility group assignments. Replicon typing, using the IncX probe derived from R6K, of plasmids classified as belonging to the IncX group



failed to give signals for a number of plasmids assigned to this group on the basis of incompatibility testing (Jones *et al.*, 1993). Analysis of these plasmids revealed a RFLP pattern common to all these plasmids, but dissimilar to that of R6K, and it was on this basis that it was suggested that the IncX group be subdivided into two classes (Jones *et al.*, 1993).

The limitations of incompatibility testing and replicon typing has meant that further plasmid characterization is required to confirm plasmid group assignment by either method. Confirmation is usually now sought in sequence data which provides information on plasmid genotypic relatedness. Cloning and sequencing of either the minimal replicon or the whole plasmid is the most straightforward approach to characterizing newly isolated replicons. Osborn *et al.* (2000) cautions that traditional phylogenetic analysis cannot be applied to many features of plasmids as a direct consequence of the overriding influence of horizontal transfer in and between plasmids. Sequencing data, like replicon-specific probes or plasmid-backbone specific PCR primers, also provides no information on the molecular interactions of incompatibility determinants which only the incompatibility phenotype of related plasmid backbones will show.

A feature that all plasmids have is a replicon, and in most plasmids there is an associated replication initiating protein. Consequently, it has been proposed that with the increasing collection of sequence data, a more practical approach to plasmid classification would be the comparison of the amino acid sequence of replication initiating proteins (del Solar *et al.*, 1998). Replication initiating proteins may have a common function, but not all plasmids, such as ColE1 (see Chapter 1) encode a replication initiating protein. Nevertheless, these appear to be the exception, and other features of their replicons may be compared.

What is evident is that the diversity of replicon structure and the modular composition of plasmids, where different modules may have a different ancestry, complicates plasmid classification methodology. It is equally evident that classification of plasmids should not be made on the basis of one method alone. A combined approach to plasmid characterization and classification, such as that used by Smalla *et al.* (2000[a]) in an investigation of IncQ-like plasmids in piggery manure, is preferable.



In this study, RFLP analysis, replicon typing, and incompatibility testing was used to determine the genotypic and phenotypic relationships of four IncQ-like plasmids isolated from the piggery manure. Genotypic analysis clearly characterized the isolated plasmids as IncQ-like, yet incompatibility testing demonstrated differing incompatibility phenotypes dependent on the plasmid pairs. This finding resulted in the first suggestion to subdivide the IncQ group of plasmids into two groups, designated IncQ $\alpha$  and IncQ $\beta$  (Smalla *et al.*, 2000[a]). The authors also used PCR primers for total community screening so as to provide information on the diversity and potential prevalence of IncQ-like plasmids in this environment. However, the primers used by Smalla *et al.* (2000[a]) would not have detected the IncQ-like plasmids that have been the main focus of this work.

The overall aim of the work presented in this chapter was the investigation of the relatedness of plasmid pTC-F14 to selected members of the IncQ family of plasmids. Incompatibility and complementation studies provided insights into the evolutionary history of the IncQ plasmid family, which consequently suggested a mechanism by which some IncQ-like plasmids have changed their incompatibility status to overcome group-associated incompatibility. RSF1010 determinants expressing incompatibility towards the pTC-F14 replicon were sought, and RSF1010 interference with *oriV*<sub>pTC-F14</sub> was demonstrated. The pTC-F14 RepB protein was also shown to have primase activity.



### **3.2) Materials and methods**

**3.2.1) Bacterial strains, bacteriophage, media and growth conditions.** The bacterial strains, plasmids and bacteriophage used in this study are shown in Appendix B. Plasmid maps for pGL10 and pIE1108 are given in Appendix D. *E. coli* DH5 $\alpha$  and *E. coli* XL1-Blue cells were grown at 37°C in Luria-Bertani medium (Appendix A), and ampicillin (100  $\mu\text{g}\cdot\text{ml}^{-1}$ ); chloramphenicol (20  $\mu\text{g}\cdot\text{ml}^{-1}$ ), kanamycin, (30  $\mu\text{g}\cdot\text{ml}^{-1}$ ); tetracycline (20  $\mu\text{g}\cdot\text{ml}^{-1}$ ) were added as required. For the primase assay, phage infected *E. coli* XL1-Blue was grown at 37°C on H-plates with Top agarose used as the overlay medium (Appendix A).

**3.2.2) DNA techniques, sequencing and analysis.** Plasmid preparation, restriction endonuclease digestions, gel electrophoresis, and cloning were carried out using standard methods (Sambrook *et al.*, 1989; Ausubel *et al.*, 1993). The polymerase chain reaction was performed in a PCR Sprint Temperature Cycling System (Hybaid) using the Expand High Fidelity PCR System DNA polymerase (Roche Molecular Biochemicals). After an initial denaturation of 60 s at 94°C, 25 cycles of 30 s at 94°C, 30 s at 55°C (for primers SEQORI and PORI) and 90 s at 72°C were performed. A final extension step of 120 s at 72°C before cooling to 4°C completed the reaction. Annealing temperatures were based on the average annealing temperature, and extension times were altered as required, for the primer pairs used to make constructs in this study (Appendix B and C). Sequencing was done by the dideoxy chain termination method, using an ABI PRISM<sup>TM</sup> 377 automated DNA sequencer.

**3.2.3) Incompatibility assay.** The ability of an incoming plasmid to displace a resident plasmid was used as the test for incompatibility. Transformation-competent, plasmid-containing *E. coli* DH5 $\alpha$  host cells were transformed with a second plasmid and plated on antibiotic-containing medium which selected only for the incoming plasmid. Twenty colonies were re-streaked so as to obtain colonies derived from single cells on solid medium containing an antibiotic which selected only for the newly-acquired plasmid. The twenty transformants were again re-streaked to single colonies on solid medium selecting only for the newly-acquired plasmid. The single



colonies were then plated onto two sets of solid medium containing a single antibiotic to separately test for the presence of the newly acquired plasmid or the plasmid which was resident at the start of the experiment. Controls to check for spontaneous loss of resident plasmids were carried out using the same procedure except that plasmid-containing *E. coli* DH5 $\alpha$  competent cells were taken through two cycles of growth on solid medium without antibiotic selection before testing for retention of the resident plasmid.

**3.2.4) Complementation assay.** Complementation in *trans* of an IncQ group plasmid *oriV* was detected by selection on solid medium for the replication of an *oriV*<sub>incQ</sub>-ColE1 construct if the complementing replicon was co-resident in *E. coli* GW125a (*polA* mutant). Transformation-competent *E. coli* GW125a containing either IncQ or IncQ-like plasmids were divided into three aliquots. One aliquot of the set was transformed with the *oriV*<sub>incQ</sub>-ColE1 construct, another with the ColE1 vector, and the last aliquot was transformed with a PolA-independent plasmid that was used as a transformation-competency control. The aliquots were plated on solid medium with selection for the incoming *oriV*<sub>incQ</sub>-ColE1 construct.

**3.2.5) Primase assay.** The replication deficient M13 bacteriophage M13 $\Delta lac110$  was used to construct M13 $\Delta F109E$  by cloning the *oriV*<sub>pTC-F14</sub> from pTC-F109E (Appendix B) into this phage. Intact phage were isolated and titered according to Ausubel *et al.* (1993). Growth curve experiments were performed as follows. An *E. coli* XL1-Blue transformant was grown to saturation in Luria-Bertani broth at 37°C. Approximately  $2 \times 10^4$  p.f.u M13 $\Delta F109E$  was then added to 200  $\mu$ l of the *E. coli* XL1-Blue transformant culture. After 10 minutes incubation at room temperature, this suspension was added to 20 ml Luria-Bertani broth containing the appropriate selection, and grown for 12 hours at 37°C with constant shaking. At the appropriate time interval, 1 ml of this culture was removed, the cells pelleted by centrifugation, and the phage titered from the supernatant as follows. 100  $\mu$ l from a serial dilution of the supernatant was added to 200  $\mu$ l *E. coli* XL1-Blue (pTC-F101) pre-grown to an optical density at 600nm of 0.8-1.0 in Luria-Bertani broth at 37°C with selection. After 10 minutes incubation at room temperature, this suspension was added to 5 ml molten Top agarose (Appendix A), and the overlay poured onto H-plates (Appendix

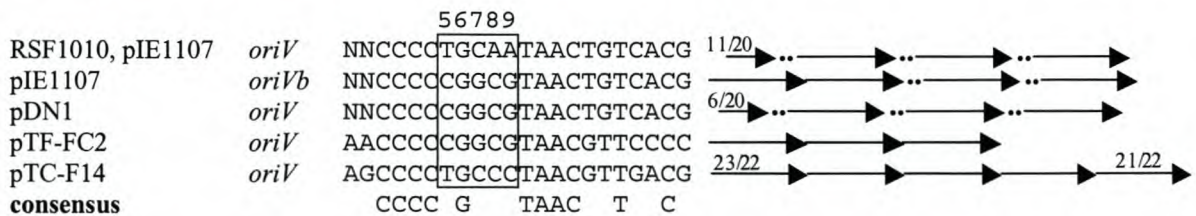


A). Once the Top agarose overlay had set, the plates were incubated overnight at 37°C, and the phage titer calculated.



### 3.3) Results

**3.3.1) IncQ and IncQ-like plasmids share iteron sequence similarity.** IncQ and IncQ-like plasmid iteron sequences were aligned and a consensus sequence determined for the IncQ family of plasmids (Fig 3.3.1). Since it has been shown that iterons of IncQ and IncQ-like plasmids are the primary determinants which exert incompatibility (Persson and Nordström, 1986; Lin *et al.*, 1987; Dorrington and Rawlings, 1989), we wished to determine which members of the IncQ family were incompatible with each other.



**Fig 3.3.1:** Sequence and arrangement of the iterons of IncQ-like plasmids tested in this study. The lengths and positions of partial iterons are indicated above the arrows. Iterons of RSF1010 and pDN1 *oriV*, pIE1107 *oriVa* and *oriVb* are 20bp long with 2bp spacers (denoted NN and being either AA or CC). In the case of the pIE1107 *oriVb*, the first 2bp spacer is missing. Iterons of pTF-FC2 and pTC-F14 are 22bp long (except where shown) and spacers are absent. Boxed nucleotides numbered 5-9 are thought to be critical for RepC recognition. Numbering in accordance with RSF1010 iteron length.

**3.3.2) Incompatibility among IncQ and IncQ-like plasmids.** Plasmids pIE1107 and RSF1010 were previously found to be incompatible due to the presence of non-essential IncQ-identical iterons present on pIE1107. If these iterons were deleted, the resulting plasmid, pIE1108, was fully compatible with RSF1010 (Tietze, 1998). We tested the functional replicon of pTC-F14 for incompatibility with the IncQ plasmid R300B replicon (pKE462), the IncQ-like plasmid pIE1108, and the pTF-FC2 replicon (pDER404). The replicons which, according to Rep protein sequence identity, were most closely related to each other (pIE1108 and R300B; pTC-F14 and pTF-FC2 [see section 2.3.2]), were fully compatible in *E. coli* (Table 3.3.1).

**Table 3.3.1:** IncQ-like plasmid displacement by other IncQ-like plasmids in *E. coli*

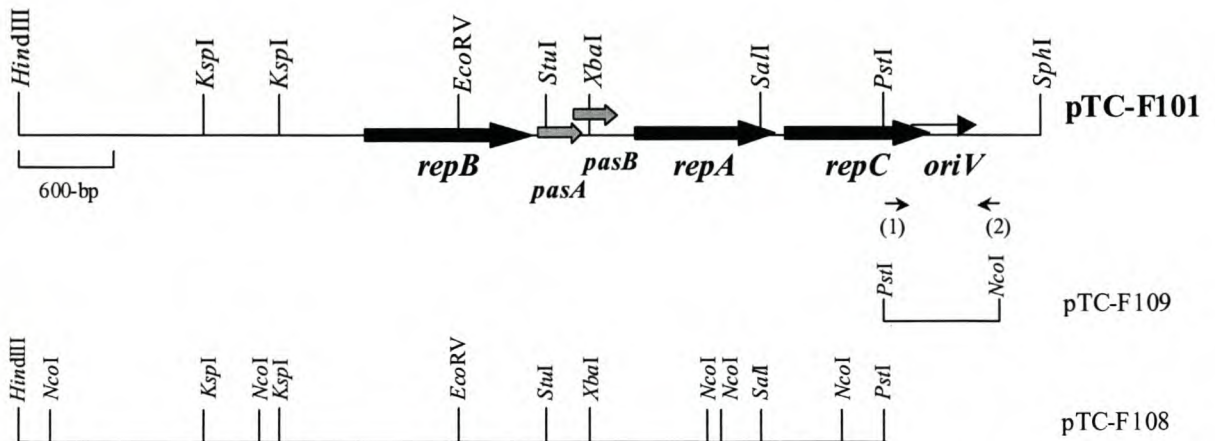
Incoming replicon (plasmid selected)	% Resident (unselected) plasmids remaining			
	R300B (pKE462)	pIE1108	pTF-FC2 (pDER404)	pTC-F14 (pTC-F101)
R300B (pKE462)	-	100	100	0
pIE1108	100	-	0	100
pTF-FC2 (pDER404)	100	100	-	100
pTC-F14 (pTC-F101)	NE <sup>(a)</sup>	100	100	-

<sup>(a)</sup>NE, plasmid not established as pTC-F14 could not be transformed into *E. coli* (pKE462).



Asymmetrical (unidirectional) incompatibility between plasmids pIE1108 and pTF-FC2, as reported by Tietze (1998), was demonstrated in the incompatibility testing of this study. If pIE1108 was the resident plasmid and either pTC-F14 or pTF-FC2 the incoming selected plasmid, neither plasmid displaced pIE1108. However, an incoming selected pIE1108 was able to displace pTF-FC2, but not a resident pTC-F14 replicon. A resident IncQ R300B replicon would not allow pTC-F14 to become established, while pTF-FC2 could not displace the R300B replicon. However, an incoming selected IncQ R300B replicon displaced a resident pTC-F14, but not a resident pTF-FC2. The four plasmids, therefore, were distinguishable from each other based on incompatibility testing. The plasmids which were originally isolated from heterotrophic bacteria appeared to have an incompatibility advantage over those isolated from chemolithoautotrophic bacteria when tested for displacement in *E. coli*.

As regions other than the *oriV* of plasmids have been isolated and shown to exert plasmid incompatibility against the parental plasmid (Novick, 1987; Dorrington and Rawlings, 1989; Tietze, 1998), the plasmids pTC-F108 and pTC-F109 (Fig 3.3.2) were used to search for intraplasmid (internal) incompatibility determinants on the pTC-F14 replicon.



**Fig 3.3.2:** Restriction enzyme map of the minimal replicon (pTC-F101) of pTC-F14, and subclones constructed in this study. The position of the primers [(1) SEQORI and (2) ORIR] used to construct pTC-F109 are shown by small labeled arrows.

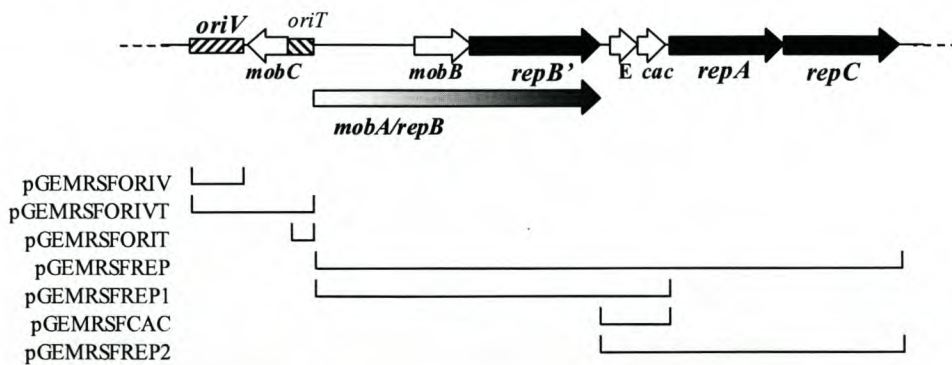
The plasmid pTC-F109 containing only the *oriV*<sub>pTC-F14</sub> demonstrated incompatibility with the resident pTC-F101, while pTC-F108 was compatible. Therefore, within the pTC-F14 replicon region, only *oriV*-associated plasmid incompatibility could be detected.



As R300B was able to displace the pTC-F14 replicon, the region expressing incompatibility was sought.

### 3.3.3) Identifying the IncQ plasmid locus expressing incompatibility to pTC-F14.

To determine the region of the IncQ plasmid expressing incompatibility towards the pTC-F14 replicon, a number of subclones were constructed for incompatibility testing (Fig 3.3.3). A derivative of plasmid RSF1010 (identical or nearly identical to R300B) was used as the source of DNA.



**Fig 3.3.3:** IncQ plasmid (RSF1010K) subclones used to identify the region on the IncQ plasmid that exerts incompatibility against the pTC-F14 replicon. The various regions indicated were PCR amplified and cloned into the pGEM-T<sup>®</sup> vector.

The construct pGEMRSFREP2 which expressed incompatibility towards pTC-F14 contains a 2.4-kb fragment incorporating the RSF1010 region spanning ORF E to *repC* (Fig 3.3.3; Table 3.3.2). As pGEMRSFCAC containing only ORF E and the *cac* gene, including the region coding for the 75-bp anti-sense ctRNA species, (Kim and Meyer, 1986) did not exert incompatibility, this suggested that the incompatibility determinant could be localized to the RSF1010 *repAC* operon (Fig 3.3.3; Table 3.3.2).

**Table 3.3.2:** pTC-F14 plasmid displacement by IncQ

Incoming replicon (plasmid selected)	% Resident (unselected) plasmids remaining
	pTC-F14 (pTC-F101)
pGEMRSFORIV	100
pGEMRSFORIVT	100
pGEMRSFORIT	100
pGEMRSFREP	0
pGEMRSFREP1	100
pGEMRSFCAC	100
pGEMRSFREP2	0



This result suggests that the RSF1010 *repAC* operon proteins interact with the pTC-F14 *oriV*. This interaction could be by recognition and binding of the RSF1010 proteins to *oriV*<sub>pTC-F14</sub>, and subsequent inhibition of replication as a consequence of this binding. Alternatively, expression of the RSF1010 *repAC* operon could be affecting the regulation of the pTC-F14 *repAC* operon.

**3.3.4) Complementation of pTC-F14 replication proteins by IncQ and IncQ-like plasmids.** Since the replication proteins of plasmids pTC-F14 and pTF-FC2 are fairly closely related (Fig 2.3.2), it was of interest to determine whether the pTC-F14 *oriV* region could be complemented in *trans* by pTF-FC2, and vice versa. To test this, *E. coli* GW125a (pDER412, pTF-FC2 replicon) and *E. coli* GW125a (pTC-F101, pTC-F14 replicon) were transformed with either pTC-F109 or pTV4164 (Appendix B) (Table 3.3.3 [rows 1 and 2]). Constructs pTC-F109 and pTV4164 contain the *oriV* region of pTC-F14 and pTF-FC2, respectively.

**Table 3.3.3:** Complementation of pTC-F14 replication proteins by IncQ and IncQ-like plasmids

<i>oriV</i> -ColE1 construct transformed <sup>(a)</sup>	Co-resident pTC-F14 deletion construct	Co-resident IncQ or IncQ-like plasmids				
		None	pTC-F14 (pTC-F101)	pTF-FC2 (pDER412)	pIE1108 <sup>(c)</sup> (pIE1108Cm)	R300B <sup>(d)</sup> (pKE462)
pTV4164	none	-	-	+	-	-
pTC-F109	none	-	+	-	-	-
pTC-F109	pGL10-BAC	+		+	+	± <sup>(d)</sup>
pTC-F109	pGL10-BA	-		-	-	-
pTC-F109	pGL10-B	-		-	-	-
pTC-F109	pGL10-AC	-		+	+	-
pTC-F109	pGL10-A	-		-	-	-
pTC-F109	pGL10-C	-		+	+	-
pTC-F109	pGL10 <sup>(b)</sup>	-		-	-	-

<sup>(a)</sup> The *oriV*-ColE1 construct was transformed into *E. coli* GW125a containing the co-resident plasmids indicated. Results are shown as growth [+] or no growth [-] on medium selecting only for the incoming *oriV*-ColE1 construct. Plasmid extractions to generate diagnostic banding patterns by restriction enzyme digestion, and re-streaking of transformants on selective medium was used to confirm results (not shown). <sup>(b)</sup> Compatible PolA-independent vector. <sup>(c)</sup> Transformants required 36 hours of growth at 37°C for visible colonies on solid medium, but re-streaking of transformants only required overnight growth at 37°C. <sup>(d)</sup> Few slow growing colonies appeared after 48 hours. Growth inhibition demonstrated to be due to incompatibility between pKE462 and pGL10-BAC when selection for both removed (not shown).

As the vectors in which the *oriV* regions were cloned (ColE1-based) do not replicate in strain GW125a (*polA*), and since the *oriV* regions cannot replicate unless the



replication proteins are provided in *trans*, transformants should be obtained only if replication protein complementation has occurred. Transformants were obtained when pTC-F109 was transformed into *E. coli* GW125a (pTC-F101), but not *E. coli* GW125a (pDER412) [Table 3.3.3]. This indicated that the replication proteins of pTC-F14, but not pTF-FC2, were able to complement the pTC-F14 *oriV*. In a reciprocal experiment, *E. coli* GW125a (pDER412) and *E. coli* GW125a (pTC-F101) were transformed with pTV4164 containing the pTF-FC2 *oriV*. Transformants were obtained for *E. coli* GW125a (pDER412), but not *E. coli* GW125a (pTC-F101) [Table 3.3.3]. Therefore, each *oriV* could be complemented in *trans* by its own replication proteins but not by those of the other plasmid. In similar tests, *E. coli* GW125a (pKE462, R300B replicon) and *E. coli* GW125a (pIE1108) recipients, when transformed with pTC-F109 or pTV4164 did not result in any transformants indicating that cross-complementation did not occur.

The complementation assay was then adapted to determine if all three replication proteins (RepB, RepA, and RepC) were plasmid-specific, or whether any of the pTC-F14 replication proteins could be exchanged for the equivalent protein from a different IncQ-family plasmid. To do this, the ability of pTC-F109 to replicate in *E. coli* GW125a in the presence of pIE1108, R300B, and pTF-FC2 was retested, except that, in addition, one or more of the replication proteins of pTC-F14 was supplied in *trans* (Table 3.3.3 [rows 3-9]). To supply the pTC-F14 replication proteins in *trans*, fragments of pTC-F14 containing various replication genes were amplified by PCR, and cloned in plasmid pGL10. The pGL10 plasmid vector contains an unrelated, compatible, RK2/RP4-based minimum replicon. When pTC-F109 was transformed into these heteroplasmid transformants, *E. coli* GW125a (pDER412; pTF-FC2 replicon) and *E. coli* GW125a (pIE1108) could complement the pTC-F14 replicon only when pTC-F14 RepC was supplied in *trans* (Table 3.3.3). This indicated that RepA and RepB of these two plasmids were able to substitute for the equivalent replication proteins of pTC-F14, but that pTC-F14 RepC was plasmid-specific.

The replication proteins of the IncQ (R300B) replicon, however, were unable to complement the replication proteins of pTC-F14. When all the pTC-F14 replicon proteins were supplied in *trans* (pGL10-BAC), in the presence of the co-resident R300B plasmid (pKE462) and pTC-F109, a few slow growing transformants were



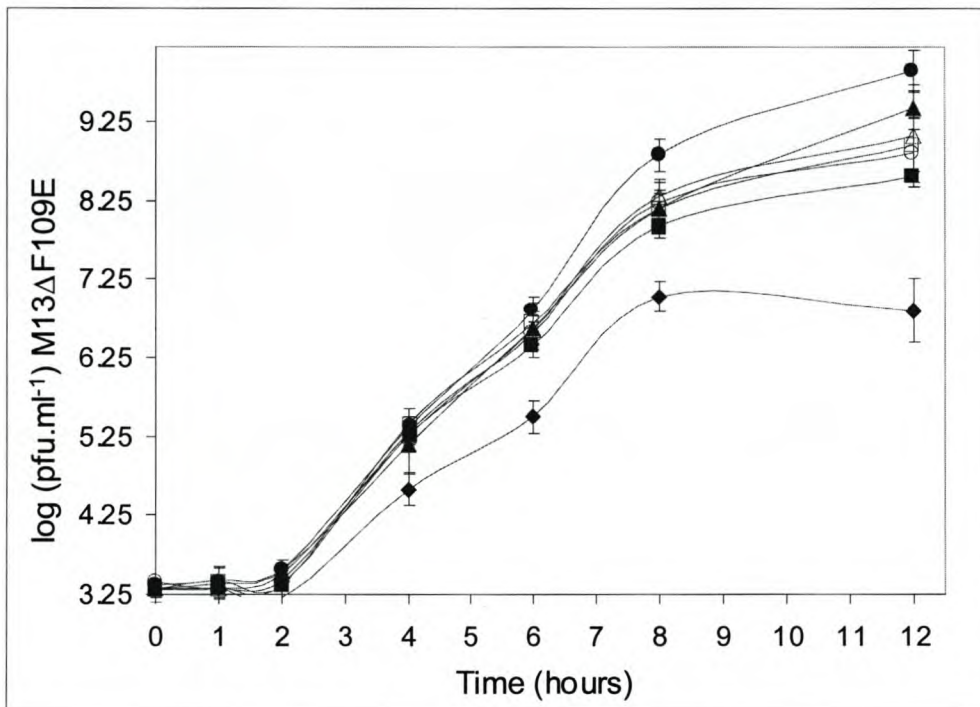
observed (Table 3.3.3). With pGL10-BAC present in *trans*, replication interference of the co-resident plasmids pTC-F109 and R300B (pKE462) was demonstrated when antibiotic selection for either pTC-F109 or pKE462 (R300B replicon) was removed (not shown). To demonstrate this interference, the *E. coli* GW125a (pKE462 + pGL10-BAC + pTC-F109) transformant was streaked onto three solid medium plates containing an antibiotic for one of the resident plasmids. Colonies were re-streaked twice, before being tested for the presence of the other two plasmids by streaking on solid media containing the appropriate antibiotic. When pKE462 (R300B replicon) had initially been selected, displacement of pTC-F109 (*oriV*<sub>pTC-F14</sub>) was demonstrated, but not pGL10-BAC (RK2/RP4 replicon vector). This suggested that pKE462 was interfering with pTC-F109 replication. The replication proteins of pTC-F14 were also found to interfere with pKE462 replication. This was demonstrated by displacement of pKE462 when pTC-F109 was initially selected in the displacement assay, and very weak growth was observed when testing for the presence of pKE462 after initial selection for pGL10-BAC (not shown). This suggested that the pTC-F14 replication proteins supplied in *trans* by pGL10-BAC were interfering with pKE462 replication.

There was concern as to whether the replication proteins were expressed from all constructs. Attempts to detect replication protein expression from all pTC-F14 constructs using an *in vitro* transcription-translation kit (*E. coli* S30 Extract system for circular DNA, Promega Corp, USA) were unsuccessful. This is possibly due to the use of the low copy number pGL10 vector (RK2/RP4-based), rather than the preferred ColE1-based vectors that the manufacturers recommend. Complementation when pGL10-C and pGL10-AC were present in *trans* suggests that expression of pTC-F14 *repC* was occurring, and complementation of pTC-F109 by pGL10-BAC suggests functional RepB and RepA expression. As pGL10-B and pGL10-BA were constructed using the same forward primer and cloning site as for pGL10-BAC, all constructs had the same promoter, and hence *repB* expression from these constructs was presumed to occur. However, to further substantiate expression of RepB from these constructs a primase assay was employed. Successful priming of the pTC-F14 *oriV* by these constructs would also verify the activity of pTC-F14 RepB. Unfortunately, no assay was available for functional RepA activity detection from pGL10-A. Such assays require purified RepA helicase, and a radiolabeled oligonucleotide annealed to a ssDNA oligonucleotide. Activity is detected by



separation of the radiolabeled oligonucleotide from the dsDNA molecule (see Scherzinger *et al.*, 1997, or Patel and Picha, 2000 for assay methodology).

**3.3.5) pTC-F14 RepB can prime  $oriV_{pTC-F14}$ .** The replication deficient M13 $\Delta lac110$  phage is defective in the region required for priming of DNA synthesis. Consequently, M13 $\Delta lac110$  produces plaques of small size, and has a reduced phage titer (Kim *et al.*, 1981). The efficiency of phage replication is increased if a priming site is cloned into M13 $\Delta lac110$ , and the corresponding host or plasmid-encoded primase is provided in *trans* (Honda *et al.*, 1989). To determine whether pTC-F14 RepB could prime  $oriV_{pTC-F14}$ , M13 $\Delta F109E$  was constructed by cloning  $oriV_{pTC-F14}$  into the mutant phage M13 $\Delta lac110$ .



**Fig 3.3.4:** Phage titers of M13 $\Delta F109E$  recombinant phage determined during growth in *E. coli* XL1-Blue (pGL10) [■], XL1-Blue (pTC-F101) [●], XL1-Blue (pGL10-B) [○], XL1-Blue (pGL10-BA) [△], XL1-Blue (pGL10-BAC) [□], XL1-Blue (pTC-KmMO) [▲] and XL1-Blue (RSF1010) [◆]. The vertical error bars represent the standard deviation between three different assays.

By providing the *repB* gene product of pTC-F14 in *trans* (pGL10-B, pGL10-BA, pGL10-BAC) the efficiency of M13 $\Delta F109E$  replication was increased and reached a phage titer approximately 2.5-fold higher than that of *E. coli* XL1-Blue (pGL10), the assay control, in a 12-hour period (Fig 3.3.4). Although the increase in phage titer was not as marked as when a similar experiment was carried out with the pTF-FC2



*repB* and its  $oriV_{p_{TF-FC2}}$  (Dorrington *et al.*, 1991), the ability of the pTC-F14 *repB* to complement its own *oriV* was highly repeatable. The slight reduction in the rate of increase in phage titer at the 6-hour time interval was routinely seen, and may correspond to a phage assembly phase, reflecting a synchronous phage replication cycle (Fig 3.3.4).

The plasmid construct pTC-KmMO was derived from pGL10-BAC by removal of the pGL10 vector and  $P_{tac}$  promoter sequence and their replacement with a kanamycin resistance gene. Plasmid pTC-KmMO contained the entire pTC-F14 replicon except that *mobA* was removed, and presumably the *repB*, *repA*, and *repC* genes were expressed from the kanamycin resistance gene promoter (Appendix B). This construct was found to replicate in *E. coli* GW125a. The phage titer after 12-hours when pTC-KmMO was resident in *E. coli* XL1-Blue was 7.4-fold higher than when pGL10 was resident. When pTC-F101 was resident, the phage titer was 21-fold higher than that of *E. coli* XL1-Blue (pGL10). This elevated phage titer is most likely a consequence of optimal expression of RepB and the MobA-repB fusion protein from the pTC-F14 replicon. The above set of results confirmed that  $oriV_{p_{TC-F14}}$  does have a priming site upon which the RepB product acts, and that RepB of pTC-F14 encodes a plasmid-template specific primase.

As the IncQ replication proteins were apparently unable to complement the pTC-F14 RepB (Table 3.3.3), it was considered worth attempting to confirm this result using the primase assay. Plasmid RSF1010K was used to provide the IncQ replication proteins in *trans* in this assay system. When compared to the *E. coli* XL1-Blue (pGL10) assay control that cannot act on the pTC-F14 *oriV*, the replication of M13ΔF109E was significantly inhibited as shown by reduced phage titer (Fig 3.3.4). In this assay system there were no replication proteins of pTC-F14 present, unlike the complementation studies (see section 3.3.4), and it has been shown that IncQ plasmid-pTC-F14 incompatibility appears to be IncQ *repAC* operon related (see section 3.3.3). A possible interpretation is that the replication proteins of the RFS1010 *repAC* operon were able to bind to the  $oriV_{p_{TC-F14}}$ , but were unable to initiate replication. This non-productive binding could have inhibited phage M13ΔF109E replication. DNA-binding studies would be able to address this interpretation, but this current study did not investigate this further.



### 3.4) Discussion

**3.4.1) Division of the IncQ plasmid group.** This study clearly illustrates the inadequacy of using plasmid incompatibility as a classification system. Incompatibility testing with the apparent IncQ-like plasmid pTC-F14 established that this plasmid was incompatible with the IncQ prototype plasmid (RSF1010/R1162/R300B) (Table 3.3.1). In an *E. coli* host, a resident pTC-F14 plasmid was displaced by an incoming IncQ prototype plasmid, but was unable to become established if the IncQ prototype plasmid was resident (Table 3.3.1). Therefore, strictly applying the definition of plasmid incompatibility for group classification, this result confirmed that pTC-F14 was an IncQ-like plasmid.

However, this incompatible phenotype was not found when pTC-F14 was tested against other members of the IncQ plasmid family. Stable segregational inheritance was observed when pTC-F14 was challenged by pTF-FC2 and pIE1108 (Table 3.3.1). However, the sequence relatedness of the pTC-F14 replication proteins, and its replicon structure, to the IncQ prototype plasmid clearly indicates that pTC-F14 is a member of the IncQ plasmid family (Fig 2.3.2). Compatible phenotypes between related IncQ-like plasmid pairs has previously been reported. Both pTF-FC2 and pIE1108 were found to be compatible with the IncQ prototype plasmid (RSF1010/R1162/R300B), while the plasmid pair pIE1108 and pTF-FC2 display asymmetric compatibility (Dorrington and Rawlings, 1989; Tietze, 1998). By definition, these plasmids should represent different incompatibility groups.

It has already been proposed that the IncQ group should be subdivided into two groups. Smalla *et al.* (2000[a]) had observed compatible phenotypes between IncQ-like plasmids isolated from piggery manure. Consequently, these authors proposed that the IncQ group be subdivided into IncQ $\alpha$  (RSF1010-type) and IncQ $\beta$  (pIE1107-type) to accommodate the related, yet compatible, IncQ-like plasmids that had been isolated from this animal-associated environment.

From the work presented in this study, it is now proposed that the two IncQ groups be subdivided further. The reason for this proposal is two-fold. Firstly, the two



chemolithoautotroph-originating IncQ-like plasmids pTC-F14 and pTF-FC2 are distinguishable from each other, and from the IncQ prototype plasmid, based on incompatibility phenotypes. It has been shown that the plasmids pTF-FC2 and pTC-F14 are compatible, and similarly the plasmid pair pIE1108 and RSF1010 are compatible (Tietze, 1998). However, pTC-F14 is incompatible with the IncQ prototype plasmid, while pTF-FC2 is compatible with the IncQ prototype plasmid (Table 3.3.1; Dorrington and Rawlings, 1989). Secondly, the relationship of the IncQ-like plasmid replication proteins, determined by amino acid identity alignments, clearly divides the IncQ plasmid family into two phylogenetic groups (Fig 2.3.2). The replication proteins of the chemolithoautotroph-originating plasmid isolates, pTC-F14 and pTF-FC2, are grouped apart from those of IncQ-like plasmids isolated from heterotrophic bacteria. Consequently, it is now proposed that based on the amino acid sequence relationship of the replication proteins of those IncQ-like plasmids which have been DNA sequenced, the IncQ group be divided into two groups, and that each initial division be further subdivided to represent each IncQ-like plasmid type in that division. In practice, IncQ group 1 plasmids are assigned to either IncQ1 $\alpha$  (RSF1010-type), or IncQ1 $\beta$  (pIE1107-type) subgroups. Similarly, the IncQ-like group 2 plasmids are assigned to either IncQ2 $\alpha$  (pTF-FC2-type) or IncQ2 $\beta$  (pTC-F14-type). Isolation of IncQ-like plasmids which cannot be assigned to either of these subgroups can result in new subgroups being created in accordance with this nomenclature. Division of the IncQ group is further supported by the two different mobilization modules identified on IncQ-like plasmids. The IncQ1 group have the IncQ-type mobilization module comprising three *mob* genes (*mobABC*), whereas the IncQ2 group have the IncP-type mobilization module comprising five *mob* genes (*mobABCDE*).

**3.4.2) Evolution of new incompatibility groups.** The clustering together of the pTC-F14 and pTF-FC2 replication proteins following amino acid sequence alignment (Fig 2.3.2) suggests that these two IncQ-like plasmids have a common ancestor. Furthermore, these two plasmids were isolated from bacteria that grow in very similar environments. As identical plasmids are incompatible, one would predict that soon after divergence these promiscuous broad host range plasmids would have encountered each other and competed for the same "replication space". However, it



has been shown in this study that pTC-F14 and pTF-FC2 are now compatible, suggesting that these two replicons have evolved to become accommodating of each other.

One may speculate on the steps by which these two previously incompatible plasmids have evolved to compatibility. For iteron-regulated plasmids the direct repeats (iterons) themselves exert the strongest incompatibility towards the parental plasmid (Persson and Nordström, 1986; Lin *et al.*, 1987; Dorrington and Rawlings, 1989). Therefore, for compatible plasmids to evolve one might expect the direct repeats to change first. However, this change would have to be to the point where the RepC protein of evolving sister plasmids no longer recognize the iterons of each other. Therefore, the iterons and RepC protein of the same plasmid would be expected to co-evolve. What is not predictable is whether the RepA and RepB proteins also co-evolve with RepC evolution in a way that discriminates between replicons.

Investigations into IncQ iteron recognition by RepC has identified two regions in the IncQ iteron sequence in which base substitutions alter the RepC binding affinity, and subsequently a possible iteron-discriminating region has also been identified. Miao *et al.* (1995[b]) demonstrated that nucleotide substitutions between nucleotides 2 and 11 of the RSF1010 iteron (refer to Fig 3.3.1 and nucleotide numbering) abolished RSF1010 RepC binding, but that substitutions between nucleotides 14 and 18 only lowered the RSF1010 RepC binding affinity. The differing incompatibility expressed towards the IncQ prototype plasmid (RSF1010/R1162/R300b) by pIE1107 *oriVa* (identical to RSF1010 iteron sequence) and *oriVb* (compatible with the IncQ prototype plasmid) was suggested to be a result of nucleotide differences in the region designated nucleotides 5-9 (boxed in Fig 3.3.1), as the flanking nucleotides of these two pIE1107 iteron sequences are otherwise identical (Tietze,1998). Consequently, Tietze (1998) proposed that these five nucleotides are essential for RepC iteron recognition. Interestingly, only one nucleotide is fully conserved between nucleotides 5-9, and this region is flanked by short sequences of absolute identity (Fig 3.3.1). The conservation of this guanine nucleotide suggests that this nucleotide is critical in this probable iteron discriminating region. Substitution of this guanine nucleotide in a two direct repeat *oriV*<sub>R1162</sub> sequence, cloned into pBR322, resulted in a mutant *oriV*<sub>R1162</sub> iteron sequence which was no longer incompatible with R1162 (Lin *et al.*, 1987). It



was also found that substitution of this guanine nucleotide in iterons ligated to a kanamycin gene resulted in the replication proteins of R1162, supplied in *trans*, being unable to act upon the iteron sequence to initiate replication of this satellite plasmid (Lin *et al.*, 1987).

From the above discussion, it appears that alteration of the iteron sequence in a specific region can completely abolish RepC binding. This suggests that if the iteron sequence changes, the plasmid requires co-evolution of the RepC initiating protein to maintain specificity for its iterons for functional replication. Using the region highlighted by Tietze (1998) as the probable iteron-discriminating region, this sequence (bp 5-9) differs by 3-bp between the compatible plasmids pTC-F14 and pTF-FC2 (Fig 3.3.1). The inability of either plasmid's RepC protein to substitute for the replication initiating protein of the other suggests that the iterons have evolved apart, and that the RepC proteins have co-evolved to the point where there is no longer cross-complementation. This argument seems to be substantiated by the finding that only when the RepC of pTC-F14 was supplied in *trans* could the RepA and RepB proteins of pTF-FC2 and pIE1108 substitute for those of pTC-F14 (Table 3.3.3). Therefore, RepC of pTC-F14 is specific for the parental plasmid, and this correlates with the compatibility phenotype observed when pTC-F14 and either of these two plasmids are co-resident in the same host cell (Table 3.3.1).

The finding that cross-complementation by RepA and RepB of pTF-FC2 and pIE1108 occurred suggests that the helicase and primase proteins are less plasmid-template specific than the RepC protein. From the above discussion on iteron-RepC co-evolution, there is a suggestion that of the three Rep proteins, the RepC protein is under the most pressure to evolve. Amino acid sequence alignment of the Rep proteins indicates a 62 % amino acid identity between the RepC orthologs of the IncQ plasmid family, while the RepA (44 %) and RepB (15 %) orthologs are not as conserved. This suggests that although the RepC protein is under the most pressure to evolve, there must be constraints on RepC evolution. Conversely, the RepB and RepA proteins seem to have less constraints on their evolution as each amino acid alignment of the RepB and RepA orthologs showed them to be less similar than the RepC orthologs, but they could still cross-complement. As it has been shown that



RepA and RepB do not physically interact with RepC (Scherzinger *et al.*, 1997), this could allow for the independent evolution of these proteins from that of RepC.

**3.4.3) Interference with pTC-F14 replication by Rep proteins of the IncQ prototype plasmid.** Incompatibility between RSF1010 and pTC-F14 was shown not to be due to the RSF1010 *oriV* region, but the RepAC encoding region (Table 3.3.2). This suggested that these replication proteins of the IncQ prototype plasmid (RSF1010/R1162/R300B) interfere with the replication of pTC-F14, and this may be the mechanism by which incompatibility between these two plasmids is expressed.

Replication inhibition of phage M13 $\Delta$ F109E (*oriV*<sub>pTC-F14</sub>) by a RSF1010 replicon (Fig 3.3.4) supports the suggestion that replication inhibition could be the consequence of RSF1010 replication protein interaction with *oriV*<sub>pTC-F14</sub>. Therefore, protein interference with replication at the origin is thought to be the mechanism of replication inhibition. A similar finding was reported by Tietze (1998) for the plasmid pairs p95L28 and pDER412. The plasmid p95L28 is a ColE1-based vector (pUCBM20) into which the *oriVb* and the *repAC* operon of pIE1107 was cloned. Cloned on its own into this vector, *oriVb*<sub>pIE1107</sub> was compatible with pDER412 (pTF-FC2 replicon), but when the pIE1107 *repAC* operon was present (p95L28), the construct was incompatible with the pTF-FC2 replicon (Tietze, 1998). In this current study, replication interference by the replication proteins of the IncQ R300B plasmid (pKE462) was demonstrated by displacement of pTC-F109 (*oriV*<sub>pTC-F14</sub>; see section 3.3.4), in the absence of antibiotic selection for pTC-F109, even when pGL10-BAC (*repBAC*<sub>pTC-F14</sub>) was present in *trans*. However, pGL10-BAC was also able to displace the unselected pKE462 in a reciprocal displacement assay. This implies pTC-F14 replication protein interference with pKE462 replication. The mechanism of replication inhibition has not been investigated further in this study.



## Chapter Four

### Regulation of the pTC-F14 replicon, and cross-regulation by IncQ and IncQ-like plasmids

#### Contents

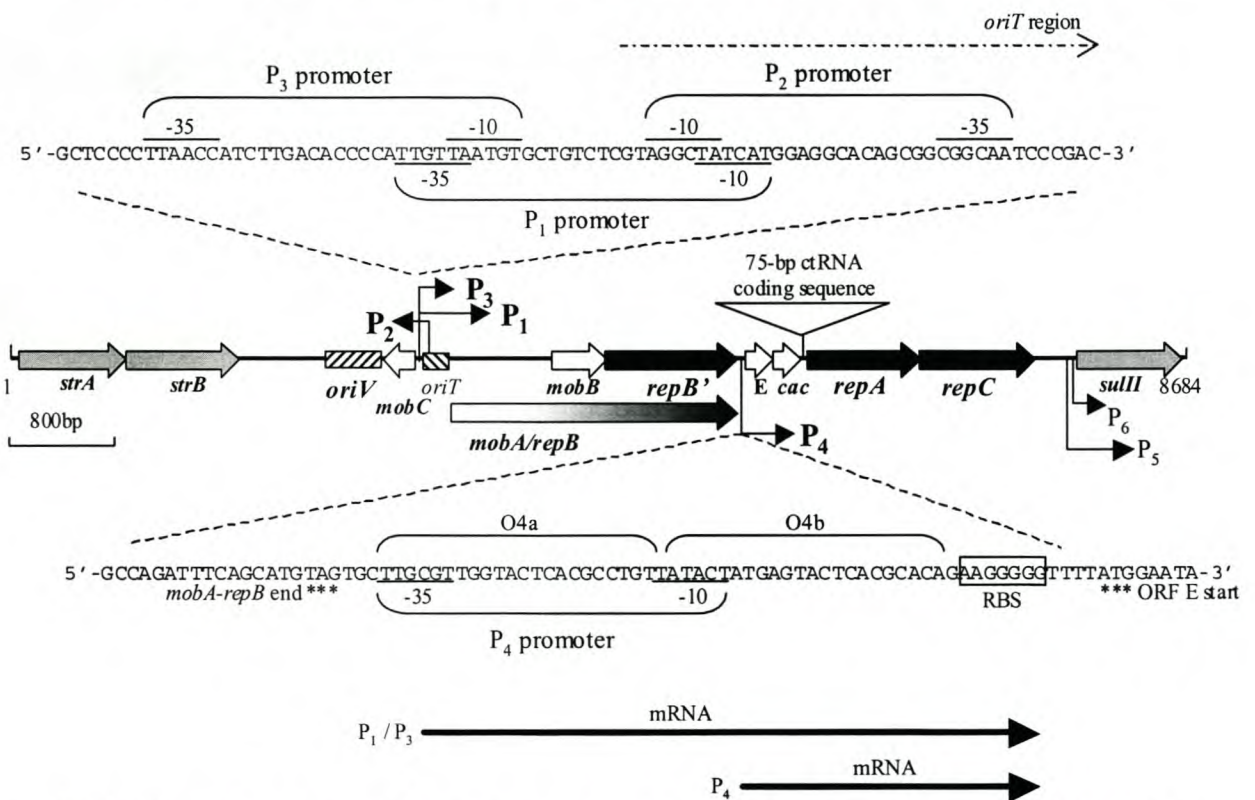
4.1)	Introduction	125
4.2)	Materials and Methods	130
	4.2.1) Bacterial strains, media and growth conditions	130
	4.2.2) DNA techniques, sequencing and analysis	130
	4.2.3) Northern blot	130
	4.2.4) Reverse-transcriptase polymerase chain reaction (RT-PCR)	131
	4.2.5) $\beta$ -galactosidase assays	132
4.3)	Results	134
	4.3.1) Number and size of pTC-F14 replicon transcripts	134
	4.3.2) RT-PCR analysis of the pTC-F14 replicon	136
	4.3.3) Regulation of putative promoter regions as determined by reporter-gene studies	139
4.4)	Discussion	143



#### 4.1) Introduction

The maintenance of a plasmid is dependent on the recognition of its promoter sequences by the Sigma ( $\sigma$ )-factors and RNAP of the bacterial species in which it is resident. The exceptional broad host range of the IncQ prototype plasmid implies that its promoter regions are recognized by  $\sigma$ -factors and RNAP of different bacterial species.

Analysis of the broad host range IncQ prototype plasmid (RSF1010/R1162/R300B) DNA sequence has identified six putative promoter regions from which transcription of the eleven recognized open reading frames of the IncQ plasmid is initiated (Fig 4.1). These promoter regions, designated P<sub>1</sub> to P<sub>6</sub>, have been shown to function in *E. coli*, and all the promoter regions, with the exception of P<sub>2</sub>, initiate transcription in the same direction (Derbyshire *et al.*, 1987; Scholz *et al.*, 1989)



**Fig 4.1:** Composite diagram of the six promoter regions, and negative regulatory elements of the IncQ plasmid. The promoters are labeled P<sub>1</sub> to P<sub>6</sub>, with the direction of transcription initiation indicated by the arrow-heads. The DNA sequence of the promoter regions is shown, and the regions (-35 and -10) contacted by the *E. coli* RNAP in each promoter is indicated. The region encoding the 75-bp ctRNA species is shown, and the two sites to which the *cac* gene product (F repressor) binds in the P<sub>4</sub> promoter region are labeled O4a and O4b. The direction of transcription and polycystronic nature of the two mRNA species transcribed from promoters P<sub>1</sub>/P<sub>3</sub> and P<sub>4</sub> are shown by the arrows. Adapted from Derbyshire *et al.* (1987) and Maeser *et al.* (1990).



Analysis of RSF1010/R1162/R300B mRNA transcripts has suggested that the IncQ prototype plasmid promoter regions P<sub>1</sub>/P<sub>3</sub> and P<sub>4</sub> are the most relevant to IncQ prototype plasmid replicon gene transcription (Fig 4.1). Using S1 nuclease protection assays, Derbyshire *et al.* (1987) identified three promoter regions upstream of the RSF1010 *oriT*. The promoter regions these authors designated as P<sub>1</sub> and P<sub>3</sub> were shown to initiate transcription of genes downstream of the *oriT* region, while the third promoter region, P<sub>2</sub>, initiated transcription towards the RSF1010 *oriV* region. No plausible reason for the functional significance of the tandem promoters P<sub>1</sub> and P<sub>3</sub> has been suggested, but initiation of transcription in the same direction has been demonstrated from both P<sub>1</sub> and P<sub>3</sub> (Derbyshire *et al.*, 1987). Bagdasarian *et al.* (1986) had previously shown that a large polycistronic transcript covering the mobilization and replication region from *mobA(repB)* to *repC* was initiated from the P<sub>1</sub>/P<sub>3</sub> promoter region. These authors also found that a second mRNA molecule was transcribed from promoter P<sub>4</sub>, and suggested that this transcript was most likely polycistronic. The second mRNA molecule initiated from promoter P<sub>4</sub> was identified when two S1-resistant RNA-DNA hybrid signals were obtained when a DNA fragment incorporating the *repAC* operon was hybridized to RNA. The larger RNA-DNA hybrid was explained by the transcript initiated from promoter P<sub>1</sub>/P<sub>3</sub>, while the second smaller hybrid was the same size as that predicted for a transcript initiated from P<sub>4</sub> (Bagdasarian *et al.*, 1986). This finding supported the suggestion by Haring *et al.* (1985) that the RepA and RepC proteins were translated from a polycistronic mRNA molecule.

This work led to the conclusion that only two mRNA transcripts are required to satisfy the expression of the IncQ prototype plasmid replication genes. No transcription termination sites have been identified, although downstream of *repC* a G+C-rich region which is followed by a T-rich sequence is believed to be a Rho ( $\rho$ )-independent termination signal (Scholz *et al.*, 1989).

Regulation of IncQ prototype plasmid (RSF1010/R1162/R300B) replicon gene expression is thought to occur by both transcriptional and post-transcriptional regulatory mechanisms. Transcriptional regulation controls the initiation of



transcription from the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster (Fig 4.1). It is the combined action of the MobC and the MobA(RepB) proteins that is essential for the maximal repression of the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster (Frey *et al.*, 1992). The P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster is arranged such that promoters P<sub>1</sub> and P<sub>3</sub> are tandem overlapping promoter regions, with the –10 region of promoter P<sub>2</sub> overlapping the –10 region of the P<sub>1</sub> promoter (Fig 4.1; Frey *et al.*, 1992). The MobC protein of the IncQ prototype plasmid mobilization module is divergently transcribed from promoter P<sub>2</sub>, relative to the initiation of transcription from P<sub>1</sub>/P<sub>3</sub>. The MobC protein is an accessory protein thought to unwind the DNA in the vicinity of the *oriT nic* site, and enhance the cleavage of the *nic* site by the MobA relaxase (Zhang and Meyer, 1997; Rawlings and Tietze, 2001). Deletion studies determined that MobC was an important component of the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster repressor, but that this protein alone was not sufficient for repression of the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster (Frey *et al.*, 1992). The repression of the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster is therefore believed to be the consequence of both the MobC and MobA-RepB fusion protein binding to the *oriT* region preventing transcription from the P<sub>1</sub> and P<sub>3</sub> tandem promoters. Consequently, the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster is subject to autorepression.

Transcriptional regulation by the product of the *cac* gene, and post-transcriptional regulation by a ctRNA molecule, both control the initiation of transcription from the P<sub>4</sub> promoter region. Bagdasarian *et al.* (1986) concluded that neither RepA nor RepC regulated the P<sub>4</sub> promoter since a transcriptional reporter gene fusion construct containing only the region between RSF1010 *repB* and *repA* indicated that P<sub>4</sub> was repressed even when the RSF1010 replicon was not present in *trans*. DNA sequencing of RSF1010 revealed two open reading frames between promoter P<sub>4</sub> and *repA* (Scholz *et al.*, 1989). These two open reading frames were designated ORF E and ORF F, and coded for proteins of 70 and 68 amino acids, respectively (Scholz *et al.*, 1989). An investigation into the functions of these two genes downstream of P<sub>4</sub> (Fig 4.1), provided no biochemical function for the product of ORF E, but the product of ORF F (now the *cac* gene [Control of A and C genes]) was found to be a repressor of the P<sub>4</sub> promoter (Scholz *et al.*, 1989; Maeser *et al.*, 1990). Promoter-reporter gene constructs of P<sub>4</sub> fused to *lacZ* indicated that if the product of ORF F was present in *trans*, or in *cis*, β-galactosidase activity was repressed (Maeser *et al.*, 1990). Further



analysis of ORF F protein found that the protein dimer bound to two sites in the P<sub>4</sub> promoter region, and these two adjacent operator sites overlapped both the -35 and -10 regions of the P<sub>4</sub> promoter suggesting that binding of ORF F protein blocks the ability of RNAP to bind to the promoter (Maeser *et al.*, 1990). These two adjacent operator sites have subsequently been designated O<sub>4a</sub> and O<sub>4b</sub> by Maeser *et al.* (1990) (Fig 4.1).

As no discernable transcription termination site was identified, the binding of the F repressor protein to the operators in the P<sub>4</sub> promoter is thought to autoregulate not only itself, but *repA* and *repC* expression, hence the designation *cac* (Scholz *et al.*, 1989; Maeser *et al.*, 1990).

The expression of the *repAC* operon of the IncQ prototype plasmid also appears to be regulated post-transcriptionally by a 75-bp ctRNA molecule encoded by a region found between the *cac* gene (previously ORF F) and *repA* of the IncQ prototype plasmid (Kim and Meyer, 1986) (Fig 4.1). This ctRNA molecule was found to protect a complementary R1162 DNA fragment from S1 nuclease activity. From the protected DNA sequence it was found that this region included the ribosomal binding site of *repA* and the first two codons of the *repA* gene, suggesting that the ctRNA molecule competes with ribosomes for the sequence encoding the RBS of *repA* on the P<sub>4</sub> mRNA transcript (Kim and Meyer, 1986). Translational inhibition by the ctRNA of R1162 was demonstrated using promoter-reporter gene constructs. Mutation of the ctRNA sequence in either of the loop regions of the proposed two stem-loop ctRNA secondary structure relieved repression of a *repA* promoter-*lacZ* translational fusion construct. Another mutation outside the ctRNA coding sequence led to a detectable increase in R1162 copy number, and this was attributed to the mutation resulting in a decrease in ctRNA synthesis (Kim and Meyer, 1986).

Therefore, the expression of the IncQ prototype plasmid replicon genes are controlled by the regulation of two transcription initiation units, the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster and the P<sub>4</sub> promoter. However, only the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster appears to be essential for expression of the replicon genes, as constructs in which the P<sub>4</sub> promoter region has been deleted are stably maintained (Frey *et al.*, 1992). This indicates that



the long polycistronic mRNA initiated at the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster is sufficient to satisfy the replication requirements of the IncQ prototype plasmid.

Although the regulation of the IncQ prototype plasmid has been studied, the regulation of the replication genes has not been investigated for any other plasmid of the IncQ group. Considering that the plasmid backbone of pTC-F14 differs from the IncQ prototype plasmid in that a *pas*-like operon is found at the same locus as the ORF E and *cac* gene, and the mobilization module of pTC-F14 is of the IncP-type, rather than an IncQ-type (DNA sequencing data, L. van Zyl, per comm.), the regulation mechanism of pTC-F14 is potentially different. The initial aim of the work in this chapter was the identification of the number and size of transcripts from the pTC-F14 replicon, from which it was hoped that potential promoter regions could be identified. The construction of promoter-reporter gene fusions would then allow for a study of pTC-F14 replication gene regulation. Secondly, these promoter-reporter gene fusion constructs would also allow for an investigation into the possible cross-regulation of the pTC-F14 replication genes by related IncQ plasmids.



## **4.2) Materials and methods**

**4.2.1) Bacterial strains, media and growth conditions.** The bacterial strains and plasmids used in this study are shown in Appendix B. The plasmid map for pMC1403 is given in Appendix D. *E. coli* DH5 $\alpha$  and *E. coli* CSH50I<sup>q</sup> cells were grown at 37°C in Luria-Bertani medium (Appendix A), and ampicillin (100  $\mu\text{g}\cdot\text{ml}^{-1}$ ); chloramphenicol (20  $\mu\text{g}\cdot\text{ml}^{-1}$ ), kanamycin, (30  $\mu\text{g}\cdot\text{ml}^{-1}$ ) were added as required.

**4.2.2) DNA techniques, sequencing and analysis.** Plasmid preparation, restriction endonuclease digestions, gel electrophoresis, and cloning were carried out using standard methods (Sambrook *et al.*, 1989; Ausubel *et al.*, 1993). Sequencing was done by the dideoxy chain termination method, using an ABI PRISM<sup>TM</sup> 377 automated DNA sequencer.

**4.2.3) Northern blot.** Total RNA was isolated by an adapted protocol of Laing and Pretorius (1992). An overnight *E. coli* DH5 $\alpha$  culture was diluted 1:100 into 30 ml pre-warmed Luria-Bertani broth with antibiotic selection as required (Appendix A), and grown to an optical density at 600 nm of 0.7-0.8 at 37°C with constant shaking. The cells were harvested and resuspended in 300  $\mu\text{l}$  cold STE Buffer (0.25 M Tris-Cl pH 7.2, 0.1 M NaCl, 0.01 M EDTA pH 8.0). The suspension was added to 300  $\mu\text{l}$  phenol (pH 6.7, Sigma) containing 0.3 g glass beads, and vortexed for 30 seconds before being placed on ice. 4  $\mu\text{l}$  10 % SDS was then added, and the solution was incubated on ice for 15 minutes. The solution was centrifuged at high speed for 10 minutes after 300  $\mu\text{l}$  chloroform: isoamyl alcohol (24:1, v/v) had been added. The aqueous phase was removed, and extracted with one volume phenol: chloroform: isoamyl alcohol (24:24:1, v/v) until the interface was clear. The extracted nucleic acids were precipitated overnight at -20°C in the presence of 0.1 volume 5 M NaCl and two volumes absolute ethanol. To remove contaminating DNA, the precipitate was pelleted by centrifugation at high speed, the supernatant removed, and the pellet dried. The pellet was resuspended in 200  $\mu\text{l}$  DNase buffer (20 mM MgCl<sub>2</sub>, 2 mM Dithioerythritol), and incubated at 37°C for 15 minutes in the presence of 5 Units DNase I (RNase-free; Roche Molecular Biochemicals). The DNase-treatment was stopped by the addition of one volume phenol: chloroform: isoamyl alcohol (24:24:1,



v/v), the aqueous phase then removed, and RNA precipitated as above. Extracted RNA (30 µg) was incubated at 55°C for 15 minutes in three volumes RNA loading buffer (Appendix A) before being separated in a 1 % (w/v) agarose/18 % (v/v) formaldehyde denaturing gel in 1X MOPS electrophoresis buffer (Appendix A). Prior to blotting, the gel was soaked in 20X SSC (3 M NaCl, 0.3 M sodium citrate pH 7.0) for 1 hour. The RNA was then capillary blotted onto Hybond™-N<sup>+</sup> (AEC-Amersham) nylon membrane using 20X SSC as the transfer buffer. The transferred RNA was cross-linked under UV-light. The pTC-F14 replicon probes were labeled with [ $\alpha$ -<sup>32</sup>P]dATP (AEC-Amersham) using the Random Primed DNA labeling kit (Roche Molecular Biochemicals), and the probes hybridized overnight to the RNA at 60°C in hybridization buffer (7 % SDS, 1% bovine serum albumin [BSA], 1 mM EDTA pH 8.0, 0.25 M Na<sub>2</sub>HPO<sub>4</sub>). After hybridization, the membrane was washed at 60°C in 1X SSC containing 0.1 % SDS, and again in 0.1X SSC containing 0.1 % SDS. After these stringency washes the probe was detected by autoradiography using MG-SR X-ray film (Konica).

**4.2.4) Reverse-transcriptase polymerase chain reaction (RT-PCR).** Total RNA was isolated using the SV Total RNA Isolation System (Promega Corp). The isolated total RNA was resuspended in 200 µl DNase buffer (18 mM Tris-Cl pH 7.3, 9 mM MnCl<sub>2</sub>, 0.9 mM NaCl) and incubated for 1 hour 30 minutes at 28°C in the presence of 100 Units DNase I (RNase-free; Roche Molecular Biochemicals). The DNase treatment was stopped by the addition of one volume phenol: chloroform: isoamyl alcohol (24:24:1, v/v), and the RNA precipitated from the aqueous phase overnight at -20°C by the addition of 0.1 volume 5 M NaCl and two volumes absolute ethanol. A two-step RT-PCR protocol was used for cDNA synthesis and cDNA product detection. The protocol of the manufacturer of the 1<sup>st</sup> Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche Molecular Biochemicals) was used for the reverse-transcriptase reaction. 125 ng RNA was used for each reverse-transcriptase reaction. Prior to the addition of the kit components and the appropriate reverse primer, the RNA was heated to 70°C for 10 minutes, and then placed on ice. The reverse transcriptase reaction commenced with a 10 minute incubation at 25°C, followed by a 1 hour incubation at 42°C, and ended with a 5 minute incubation at 94°C, and a 5 minute incubation at 4°C. The polymerase chain reaction was performed in a PCR



Sprint Temperature Cycling System (Hybaid) using *Taq* DNA polymerase (Promega Corp). For standard PCR reactions, a protocol of an initial denaturation of 60 s at 94°C, 25 cycles of 30 s at 94°C, 30 s at 54°C and 90 s at 72°C was used. A final extension step of 120 s at 72°C before cooling to 4°C completed the reaction. Extension times were altered as required for the primer pairs (Appendix C), and 2 µl of the 20 µl (total volume) reverse-transcriptase reaction was used in each PCR reaction.

**4.2.4)  $\beta$ -galactosidase assays.** The putative promoter regions of the pTC-F14 replicon region were PCR amplified using primer pairs LACF14MOBBF/LACF14MOBBR (*mobB* promoter), LACF14MOBAF/LACF14MOBAR (*mobA* promoter), LACF14REPBPF/LACF14REPBPR (*repB* promoter), primer5/primer6 (*pas* operon promoter; S. Deane, per comm.), LACF14REPAF/LACF14REPAR (*repA* promoter), and LACF14REPCF/LACF14REPCR (*repC* promoter) (Appendix C). A PCR Sprint Temperature Cycling System (Hybaid) and Expand High Fidelity PCR System DNA polymerase (Roche Molecular Biochemicals) was used to amplify the putative promoter regions. After an initial denaturation of 60 s at 94°C, 25 cycles of 30 s at 94°C, 30 s at 61°C and 90 s at 72°C were performed. A final extension step of 120 s at 72°C before cooling to 4°C completed the reaction. The PCR products were digested with *EcoRI* and *BamHI* and cloned into pMC1403 (Appendix B and D) to construct a promoter-reporter gene (*lacZ*) fusion recombinant plasmid. As these recombinant plasmids are translational fusion constructs requiring in-frame ligation of the start codon of the promoter-associated gene to the reporter gene (*lacZ*), all constructs were sequenced with primer LACZPRI (Appendix C) to ensure the promoter fusions were correct. These constructs were transformed into *E. coli* CSH50I<sup>q</sup> and the  $\beta$ -galactosidase activity measured using the method of Miller (1972). Overnight cultures were diluted 1:100 into fresh pre-warmed Luria-Bertani medium containing the appropriate antibiotic selection and grown at 37°C for 4 hours. After this time period the optical density (600 nm) was recorded, and the culture diluted 1:5 into Z-buffer (Appendix A). The exception to this protocol was *E. coli* CSH50I<sup>q</sup> (pMCF14PAS) which was initially diluted 1:500 into fresh medium with appropriate selection, grown for 3 hours at 37°C, and diluted 1:20 into Z-buffer. Following dilution in Z-buffer, the culture suspensions were briefly vortexed in the



presence of toluene (1 %, v/v), and then placed at 37°C for 45 minutes to evaporate the toluene, before being placed at 28°C for the remainder of the assay. The assay was started with the addition of *o*-nitrophenyl- $\beta$ -D-galactoside (ONPG) (Appendix A) to a final concentration of 0.67 mg.ml<sup>-1</sup>. After 30 minutes incubation [6 minutes for *E. coli* CSH50I<sup>q</sup> (pMCF14PAS)] at 28°C, the reaction was stopped with the addition of 0.42 volume of 1 M Na<sub>2</sub>CO<sub>3</sub>. The reaction was then centrifuged at high speed for 5 minutes to remove the cells, and the supernatant optical density was measured at both 420nm and 550nm. The  $\beta$ -galactosidase activity was calculated according to the equation by Miller (1972), and the activity expressed as Miller Units (Eqn 3).

$$\text{Miller Units} = 1000 \times \frac{\text{OD}_{420} - 1.75 \times \text{OD}_{550}}{t \times v \times \text{OD}_{600}} \quad (3)$$

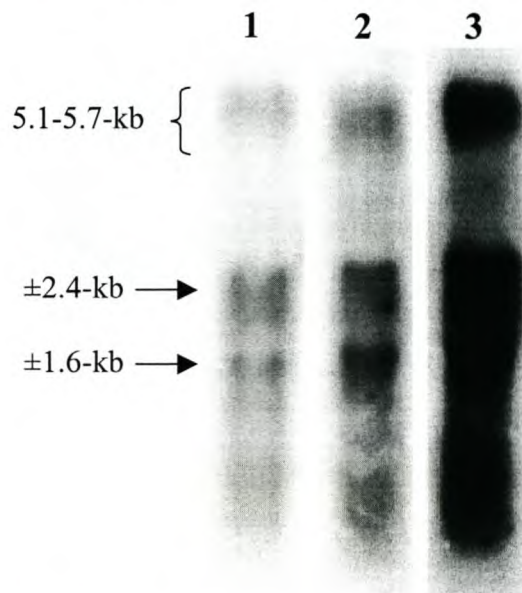
$t$  = time of the reaction in minutes

$v$  = volume of the culture used in the assay, in ml



### 4.3) Results

**4.3.1) Number and size of pTC-F14 replicon transcripts.** The number and size of pTC-F14 replicon transcripts was investigated using the Northern blot technique. Total RNA from *E. coli* DH5 $\alpha$  (pTC-F14Cm) was separated on a 1 % denaturing gel, capillary blotted onto a positively-charged membrane, and probed with DNA fragments of the pTC-F14 replicon. The DNA probes representing pTC-F14 *repC*, *repA*, *repB*, *repAC*, and *repBA* were excised using restriction endonucleases from PCR products previously cloned into the pGL10 vector (Appendix B; see section 3.3.4). The *mobA-repB* and *mobB* probes were excised from constructs kindly supplied by L. van Zyl (per comm.).



**Fig 4.3.1:** Example of three Northern blots to detect transcripts from the pTC-F14 replicon. The three blots shown were probed with *repC* DNA (lane 1), *repA* DNA (lane 2), and *repAC* DNA (lane 3).

Several signals were detected when *E. coli* DH5 $\alpha$  (pTC-F14Cm) RNA was probed with the above DNA probes. A signal for a high molecular weight mRNA species of 5.1-5.7-kb, two smaller signals of approximately 2.4-kb and 1.6-kb, as well as a smear in the region below 1-kb were detected (Fig 4.3.1). The results using the *repA*, *repC* and *repAC* DNA probes are shown in Fig 4.3.1, and similar results were obtained with a *repB* DNA probe, a DNA probe incorporating the region encoding *repB* to *repA*, and a DNA probe of the *mobA-repB* region (not shown). Signal intensities suggested low transcript levels, but increasing the amount of RNA loaded from 30 to 50  $\mu$ g yielded a saturated signal that was difficult to interpret (Fig 4.3.1; lane 3). No signal



was obtained when total RNA extracted from *E. coli* DH5 $\alpha$  was probed with any of the DNA probes (not shown), indicating that the hybridization signals obtained were pTC-F14 plasmid specific.

To determine if the 5.1-5.7-kb signal was the consequence of DNA contamination, purified pTC-F14Cm plasmid DNA was included as a control in the denaturing gel of the Northern blots. No signal from this experimental control was detected, indicating that plasmid DNA was not responsible for the signals obtained. Consequently, the 5.1 to 5.7-kb signal was interpreted as arising from probe hybridization to a mRNA species. As the 5.1 to 5.7-kb hybridization signal was detected with all the above DNA probes, this suggested that the transcript was polycistronic. Should a large polycistronic mRNA species be initiated upstream of *mobB* (in the *oriT*<sub>pTC-F14</sub> region) and terminated downstream of *repC* (Appendix D, Fig C), the size for such a transcript is predicted to be  $\pm 5.7$ -kb. The estimated size for such a transcript was within the size range detected, and it was consequently deduced that the 5.1-5.7-kb signal obtained corresponded to a polycistronic mRNA transcript initiated upstream of *mobB* that encodes all the genes of the pTC-F14 replicon.

Signals in the approximate size range of 2.4-kb, 1.6-kb, and a smear below 1-kb were routinely detected when *E. coli* DH5 $\alpha$  (pTC-F14Cm) RNA was probed with any of the above DNA probes (Fig 4.3.1). However, the size estimates of these two signals coincides with the size range in which the *E. coli* 23S, 16S and 5S rRNA species are found (Fig 4.3.1). Therefore, the  $\pm 2.4$ -kb and  $\pm 1.6$ -kb mRNA transcripts detected could have been due to plasmid mRNA entrapped by the two *E. coli* rRNA species. Interpretation of the Northern blots required consideration of this possibility. The  $\pm 2.4$ -kb signal is the approximate size of a transcript that could encode the *pasAB-repAC* operon (2.5-kb) (Fig 4.3.1). An alternative explanation is that just the *repAC* operon is encoded on a bicistronic mRNA. The predicted size of such a bicistronic transcript is 1.8-kb, and conceivably this is within the range of the  $\pm 1.6$ -kb signal (Fig 4.3.1). Therefore, the *repAC* operon may be transcribed on two RNA species ( $\pm 2.5$ -kb and  $\pm 1.8$ -kb). Should pTC-F14 *repB* be transcribed on a mRNA molecule separately from the mRNA transcript encoding *repAC*, then the estimated signal size would be either a 1.1-kb transcript if *repB* is transcribed on a monocistronic mRNA,

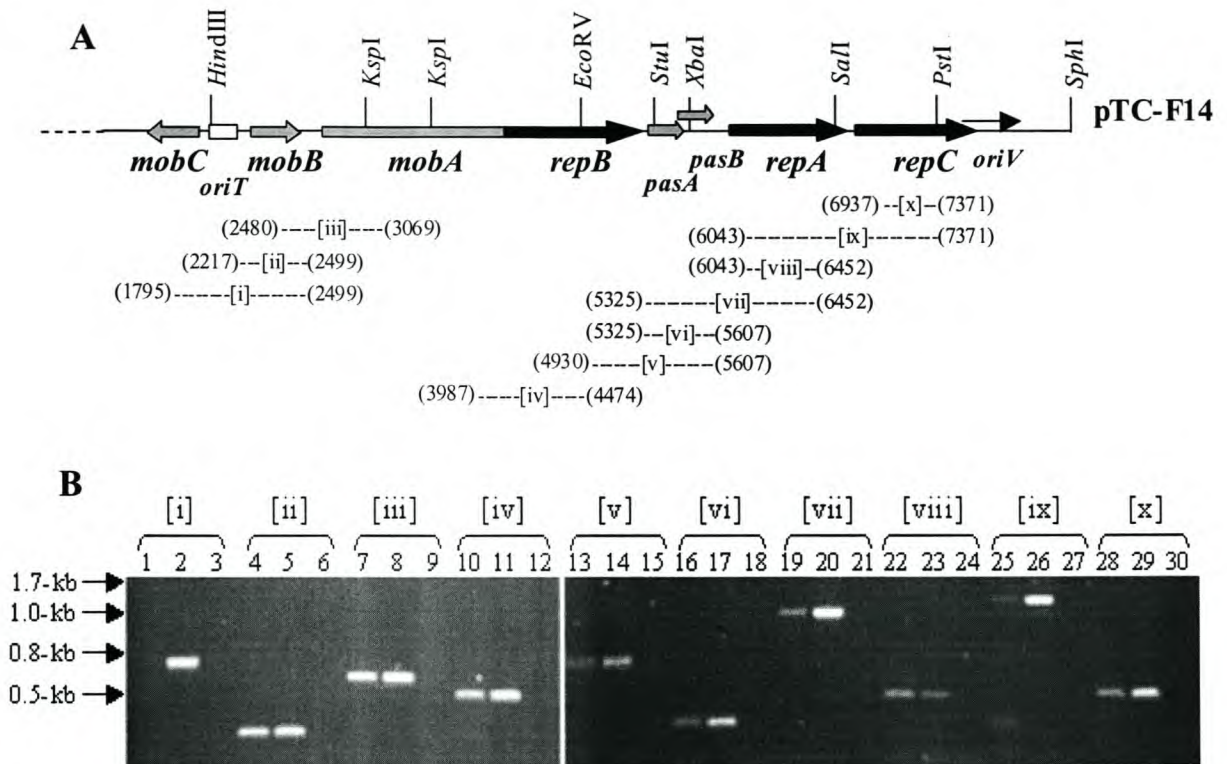


or 2.5-kb if transcribed as part of the *mobA-repB* gene fusion. Given that the sizes of the transcript signals correspond to the location of the *E. coli* rRNA species, it was not possible to unequivocally determine whether smaller transcripts were synthesized from pTC-F14.

**4.3.2) RT-PCR analysis of the pTC-F14 replicon.** RT-PCR was used to confirm the 5.1-5.7-kb transcript detected by Northern blot analysis. To overcome problems associated with reverse-transcriptase (RT) dissociation from large RNA templates, the experimental design was such that combinations of overlapping forward and reverse primers would substantiate detection of any polycistronic mRNA species (Fig 4.3.2). If a large polycistronic mRNA encoding the entire region from pTC-F14 *mobB* to *repC* was indeed transcribed, smaller mRNA transcripts would not be detected by this method.

The RT-PCR reactions (Fig 4.3.2) were each performed three times to confirm the result obtained. To detect DNA contamination in the RNA extracts used for the RT-PCR reactions, PCR reactions were performed with each primer pair using a three-fold higher concentration of the *E. coli* DH5 $\alpha$  (pTC-F14Cm) total RNA extract than was used in each RT-PCR reaction. No amplification products were detected. During RT-PCR analysis of pTC-F14 transcripts, DNA contamination was accounted for by control reactions in which the AMV reverse-transcriptase was not added to the RT reaction. PCR products were not obtained for any of the above control reactions. Occasionally, a second smaller PCR product was observed following PCR of the cDNA. A second PCR product can be seen in lane 25 of Fig 4.3.2, but as it does not correspond to the product size expected, and was not bigger than the PCR reaction control (middle lane of each set; Fig 4.3.2), it was deemed to be the result of mis-priming.





**Fig 4.3.2:** RT-PCR analysis of the pTC-F14 replicon. (A) Bracketed numbers represent the primers and the location of the primers (base pair position) relative to the pTC-F14 DNA sequence (Appendix E) used in each RT-PCR reaction. The reverse primers used in the RT reaction are as follows: (2499) RTF14MOBBR, (3069) RTF14MOBAR, (4474) RTF14REPBR, (5607) RTF14PASBR, (6452) RTF14REPAR, and (7371) RTF14REPCR. Latin numerals refer to the primer pairs used in the PCR reactions following the RT reaction (B). The forward primers used in the PCR reaction are as follows: (1795) RTF14MOBCJMOBB, (2217) RTF14MOBBF, (2480) RTF14MOBBJMObAF, (3987) RTF14MOBAJREPBF, (4930) RTF14REPBJPASF, (5325) RTF14PASAF, (6043) RTF14REPAF, and (6937) RTF14REPCF. The RT-PCR reaction products shown in (B) are grouped in sets of three, with the same primer pair used in each reaction of the set. Lanes 1, 4, 7, 10, 13, 16, 19, 22, 25, and 28 are RT-PCR reactions performed on extracted *E. coli* (pTC-F14Cm) total RNA. Lanes 2, 5, 8, 11, 14, 17, 20, 23, 26, and 29 are PCR reactions performed on purified pTC-F14Cm DNA as a PCR control. Lanes 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30 are exactly the same RT-PCR reactions as those performed in the first lane of each set except that AMV reverse-transcriptase was not added to the reaction mix.

It appears from the RT-PCR analysis (Fig 4.3.2) that there is indeed a large transcript that is initiated upstream of the pTC-F14 *mobB* gene and extends beyond the *repC* gene. The RT-PCR experimental design was such that by using the primer combinations shown in Fig 4.3.2[A], PCR products obtained would link gene junctions allowing for the identification of polycistronic transcripts. Since the PCR step of the two-step RT-PCR reaction did not yield a reaction product when using the RTF14MOBCJMOBB (1795) forward primer and the RTF14MOBBR (2499) reverse primer (Fig 4.3.2[B]; lane 1 [i]), this was interpreted as showing that no transcript was initiated from *mobC*, or upstream of the *mobC* gene. However, reaction products were obtained for all the other primer pairs (Fig 4.3.2), which was interpreted as

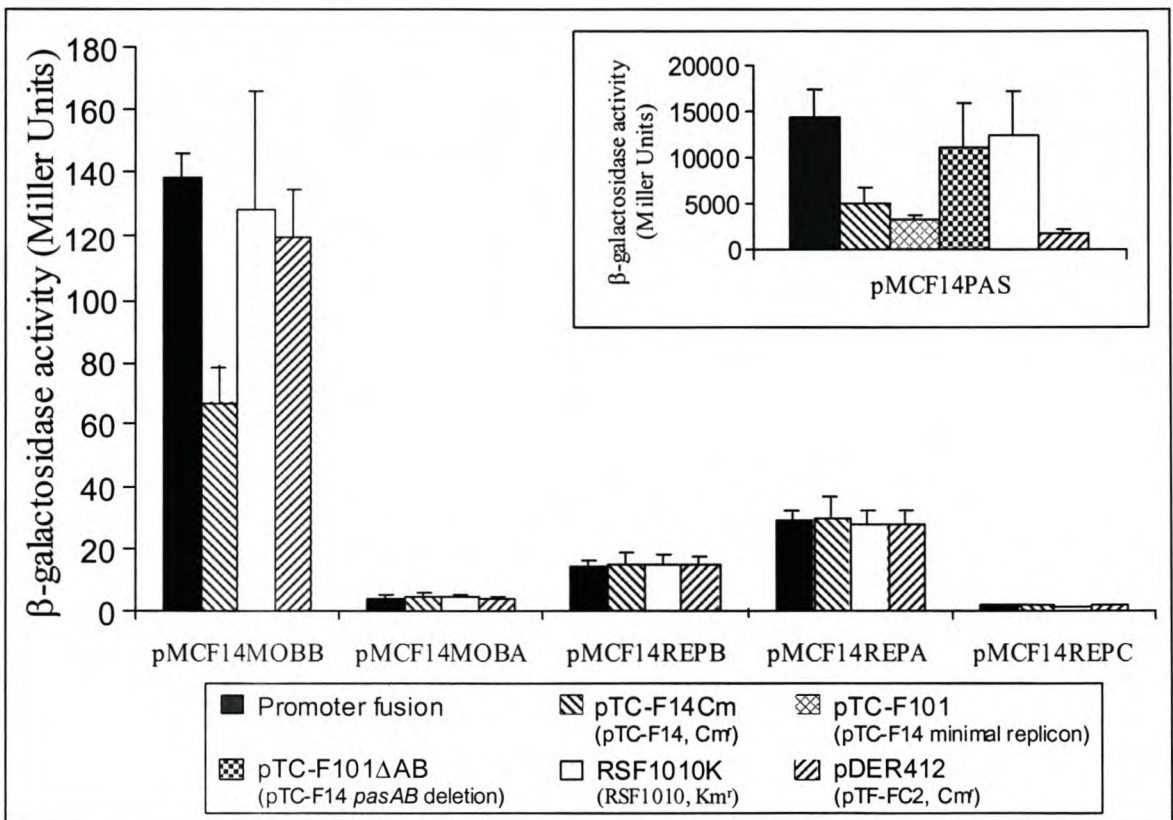


demonstrating that a large mRNA transcript is initiated upstream of *mobB* for expression of all the pTC-F14 replicon genes. A putative promoter has been identified upstream of the *oriT nic* site (Appendix E) which differs by one nucleotide from the *E. coli*  $\sigma^{70}$  consensus sequence in both its  $-35$  region (TTGACTI) and  $-10$  region (TACAAT), and has a spacer of N<sub>16</sub>. This spacer region is within the most common range of N<sub>16-18</sub> for a promoter (Harley and Reynolds, 1987). Should a transcript be initiated from this promoter that terminates downstream of *repC*, a transcript size of approximately 5.7-kb is predicted. This corresponds to the estimated size of the high molecular weight RNA species detected by Northern blot analysis (Fig 4.3.1).

As a consequence of this large polycistronic mRNA species described above, detection of smaller transcripts initiated from other promoters on the pTC-F14 replicon would be difficult to detect using the RT-PCR technique. To try and address this problem, different forward and reverse primer combinations were used in an attempt to determine the end-point of cDNA synthesis in the RT reactions. The largest product that could be detected was a 2.4-kb product with the primer pair RTF14REPBJPASF (4930) and RTF14REPCR (7371) (not shown). A putative promoter region has been identified upstream of *pasA* (Appendix E; see section 4.3.3) that has one nucleotide difference in its  $-35$  region (TTCACA), and a two nucleotide difference in its  $-10$  region (TATATC) when compared to the *E. coli*  $\sigma^{70}$  consensus sequence, as well as a spacer region of N<sub>17</sub>. It is thus likely that a transcript is initiated from this *pas*-promoter region (see section 4.3.3). However, the product from the RTF14REPBJPASF and RTF14REPCR primer pair shows that cDNA products from smaller mRNA transcripts initiated from the pTC-F14 replicon (such as from the *pas* operon) cannot be distinguished from the cDNA product of the large transcript (5.1-5.7-kb) in this RT-PCR experiment. This is because the cDNA product originating from the large polycistronic mRNA transcript (5.1-5.7-kb) will always provide a template for whatever combination of primer pair is used for PCR reactions to detect smaller cDNA products. Real-time PCR would have allowed one to distinguish between the large (5.1-5.7-kb) transcript and smaller mRNA transcripts, but this equipment was not available at the time.



**4.3.3) Regulation of putative promoter regions as determined by reporter-gene studies.** To further investigate the number and strength of promoters of the pTC-F14 replicon, it was decided that the region upstream of each pTC-F14 gene from *mobB* to *repC* would be PCR amplified and cloned in-frame into the pMC1403 promoter-reporter gene (*lacZ*) fusion vector. To allow sufficient DNA fragment length to accommodate putative promoter regions, the following PCR product size for each upstream gene region was cloned into the pMC1403 vector: 364-bp (*mobB*), 395-bp (*mobA*), 419-bp (*repB*), 307-bp (*pasA*), 397-bp (*repA*), and 400-bp (*repC*). Furthermore, to ensure that no host cell background  $\beta$ -galactosidase activity interfered with the assays, and that any  $P_{tac}$  controlled genes added in *trans* were repressed in the absence of IPTG, *E. coli* CSH50I<sup>q</sup> was used as the host cell as this strain contains *lacI*<sup>q</sup> on the F' plasmid (Appendix B).



**Fig 4.3.3:** Cross-regulation of putative pTC-F14 promoter regions by related IncQ plasmids present in *trans*, as determined by  $\beta$ -galactosidase activity. The plasmid present in *trans* to the promoter – reporter gene construct (given below each set of assays; Appendix B) is indicated by the boxed legend. The  $\beta$ -galactosidase activity for each construct was calculated from the mean of three different assays, and within each assay the  $\beta$ -galactosidase activity was measured from three samples. The vertical error bars represent the standard deviation between three different assays.



The  $\beta$ -galactosidase activity for each putative promoter fusion and how this activity was affected by the parental pTC-F14 (pTC-F14Cm) plasmid and the related IncQ plasmids RSF1010 (RSF1010K) and pTF-FC2 (pDER412) when placed in *trans* is shown in Fig 4.3.3.

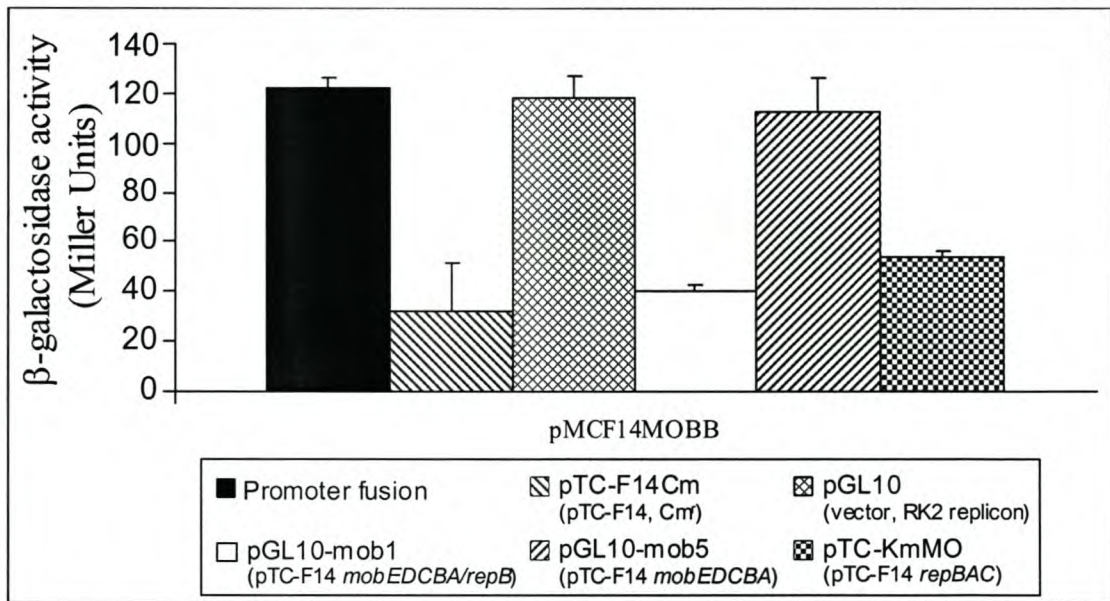
The strongest  $\beta$ -galactosidase activity measured for the pTC-F14 replicon was for the *pas* gene promoter translational fusion construct pMCF14PAS (Fig 4.3.3). The  $\beta$ -galactosidase activity of  $14426 \pm 2943$  units obtained for *E. coli* CSH50I<sup>q</sup> (pMCF14PAS) was in stark contrast to the substantially lower  $\beta$ -galactosidase activities measured for the other pTC-F14 promoter-*lacZ* fusion constructs. The *mobB* promoter region (pMCF14MOBB) gave the second highest activity of only  $138 \pm 7.95$  units (Fig 4.3.3). The negligible  $\beta$ -galactosidase activities (less than 5 units) measured for the putative promoter regions of pTC-F14 *mobA* and *repC* discounted these regions as containing promoter sequences active in *E. coli*, while *repB* and *repA* promoter-*lacZ* constructs had  $\beta$ -galactosidase activities of  $13.94 \pm 2.1$  units and  $28.83 \pm 2.0$  units, respectively (Fig 4.3.3). When pTC-F14 and related IncQ family plasmids were placed in *trans* with the *repB* and *repA* promoter-*lacZ* fusion constructs the  $\beta$ -galactosidase activities were unaffected (Fig 4.3.3). This suggested that the low level of  $\beta$ -galactosidase activity obtained was spurious and did not represent genuine plasmid promoter activity.

Repression of the pTC-F14 *mobB* and *pas* operon putative promoters by pTC-F14Cm was evident from the reduction in  $\beta$ -galactosidase activity when this plasmid was present in *trans* (Fig 4.3.3). The  $\beta$ -galactosidase activities of the pTC-F14 *mobB* (pMCF14MOBB) and *pas* operon (pMCF14PAS) promoter-*lacZ* constructs were reduced by 52 % and 65 %, respectively, when pTC-F14Cm was present in *trans* (Fig 4.3.3). The return to 77 % of *E. coli* CSH50I<sup>q</sup> (pMCF14PAS)  $\beta$ -galactosidase activity when pTC-F101 $\Delta$ AB (*pasAB* deletion) is present in *trans* suggests that the *pas* operon putative promoter is subject to regulation by products of the *pas* operon (Fig 4.3.3). The putative promoter region of the *pas* operon also appears to be the only promoter region which responds to cross-regulation by a related IncQ plasmid. The  $\beta$ -galactosidase activity of the pTC-F14 *pas* operon promoter-*lacZ* fusion (pMCF14PAS) was reduced by 88 % when the closely related plasmid pTF-FC2



(pDER412) was present in *trans* (Fig 4.3.3). Further investigation into the native repressor of the pTC-F14 *pas* operon promoter, and repression of the *pas* operon promoter by pTF-FC2, is being undertaken in our laboratory by S. Deane.

The repressor of the pTC-F14 *mobB* putative promoter was localized to the region of the pTC-F14 replicon upstream of the *pas* operon. This was determined by  $\beta$ -galactosidase assays of *E. coli* CSH50I<sup>d</sup> (pMCF14MOBB) into which constructs containing various regions of the pTC-F14 replicon had been transformed. Placement in *trans* of a pTC-F14 replicon containing a deletion of the *pasAB* genes (pTC-F101 $\Delta$ AB) did not relieve the repression of the *mobB* promoter ( $41 \pm 0.7$  units; not shown), while pGL10-AC (Appendix B) did not repress  $\beta$ -galactosidase activity ( $139 \pm 13.06$  units; not shown). This indicated that the repressor of this promoter was located upstream of the pTC-F14 *pas* operon. The regulation of the *mobB* promoter-*lacZ* fusion was plasmid specific as neither RSF1010 nor pDER412 reduced expression of the fusion construct (Fig 4.3.3).



**Fig 4.3.4:** Regulation of the pTC-F14 *mobB* promoter region as determined by  $\beta$ -galactosidase assays. The  $\beta$ -galactosidase activity for each construct was calculated from the mean of three different assays, and within each assay the  $\beta$ -galactosidase activity was measured from three samples. The vertical error bars represent the standard deviation between three different assays.

The instability of the *mobB* promoter-fusion construct when pTC-F14Cm was present in *trans* is thought to account for the 73 % reduction of  $\beta$ -galactosidase activity for this assay (Fig 4.3.4). This is in contrast to the 52 % reduction in  $\beta$ -galactosidase



activity for the same assay reported in Fig 4.3.3. No other plasmid in *trans* to the *mobB* promoter-fusion construct caused promoter-reporter gene construct instability.

The pGL10-mob1 construct, which contains the pTC-F14 mobilization region from *mobE* to *repB*, encodes a fully functional MobA-RepB fusion protein. When pGL10-mob1 was transformed into *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB)  $\beta$ -galactosidase activity was reduced to 33 % of the activity measured for *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB) (Fig 4.3.4). To investigate which of the pGL10-mob1 proteins was the repressor, the pGL10-mob5 construct, which encodes a truncated MobA-RepB protein (primase domain removed), was transformed into *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB). With pGL10-mob5 in *trans* the  $\beta$ -galactosidase activity of the *mobB* promoter-*lacZ* construct was 92 % of the activity measured for *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB), suggesting that the RepB protein was the repressor of the *mobB* promoter (Fig 4.3.4).

To confirm the RepB primase of pTC-F14 could repress the *mobB* promoter, the construct pTC-KmMO was transformed into *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB). The relaxase domain of the MobA-RepB fusion protein had been removed in the construction of pTC-KmMO by ligation of PCR amplified pTC-F14 *repBAC+oriV* to a kanamycin cassette (see section 3.3.5; Appendix B). With pTC-KmMO present in *trans*, the  $\beta$ -galactosidase activity was reduced by 57 % compared to *E. coli* CSH50I<sup>q</sup> (pMCF14MOBB) (Fig 4.3.4). This allowed for the identification of RepB (the primase domain of the MobA-RepB fusion protein) as the repressor of the *mobB* promoter.



#### 4.4) Discussion

Northern blot and RT-PCR analysis indicated that an approximately 5.7-kb transcript is initiated upstream of the pTC-F14 *mobB* gene, and this transcript size suggests that the transcript encodes all the replication proteins of the replicon. However, neither technique clearly showed how many smaller transcripts were synthesized. Expression of the replication genes and the *mobB* and *mobA* genes of the mobilization module on a large polycistronic transcript is similar to the strategy used by the RSF1010/R1162/R300B IncQ prototype plasmid (Bagdasarian *et al.*, 1986). Analysis of the DNA sequence around the pTC-F14 *oriT-mobB* region identified a putative promoter sequence upstream of the *oriT nic* site. This promoter sequence shows good DNA sequence similarity (TTGACT-N<sub>16</sub>-TACAAT) to the *E. coli*  $\sigma^{70}$ -RNAP type consensus sequence (Harley and Reynolds, 1987). A smaller DNA fragment (171-bp) upstream of *mobB* than was used for the construction of pMCF14MOBB (364-bp of upstream DNA), and which did not include the putative promoter region (Appendix E), was amplified using the same reverse primer as for pMCF14MOBB, and fused to the *lacZ* gene of pMC1403. This construct, pMCF14MOBB2, gave no detectable  $\beta$ -galactosidase activity supporting the view that the putative promoter above was the likely promoter.

A smaller transcript is thought to be produced specifically for the expression of the IncQ replicon ORF E-*cac-repAC* operon (Bagdasarian *et al.*, 1986; Maeser *et al.*, 1990). The transcript analysis techniques used in this current study could not, however, determine if a transcript encoding only the *pas-repAC* operon was similarly produced. The much higher  $\beta$ -galactosidase activity determined for the putative *pas* promoter region than for any of the other pTC-F14 replicon promoter regions tested, strongly suggests that a transcript is initiated from this promoter. However, it must be taken into consideration that for translational reporter-gene fusions, increased reporter enzyme synthesis is not only a reflection of the amount of transcript, but also the strength of the ribosomal binding site and the context of the fusion.

A transcript size of about 2.5-kb can be expected for such a polycistronic transcript, and when DNA of either *repC* or *repA* was used as a probe for a Northern blot, a



corresponding transcript size was detected (Fig 4.3.1). Since the highly synthesized 23S rRNA transcript is also of this size ( $\pm 2.5$ -kb), the abundant rRNA species may have entrapped the pTC-F14 mRNA transcripts. Therefore, it cannot be unequivocally said that the 2.5-kb signal corresponds to a transcript encoding the *pas-repAC* operon. Analysis of DNA sequence upstream of the *pas* operon has also revealed a likely promoter (TTCACA-N<sub>17</sub>-TATATC) which shows high DNA sequence similarity to the *E. coli*  $\sigma^{70}$ -RNAP type consensus sequence.

Reporter gene (*lacZ*) fusions were made to investigate the regulation of each of the pTC-F14 *mobB*, *mobA*, *repB*, *pasA*, *repA*, and *repC* genes. Only the reporter-gene fusions to the *mobB*, *repB*, *pasA*, and *repA* putative promoter regions gave  $\beta$ -galactosidase activities above background levels. Upstream regions of both pTC-F14 *repB* and *repA* cloned in frame to the *lacZ* reporter gene gave low, but detectable,  $\beta$ -galactosidase activities, and both promoter regions appeared not to be regulated by products of the pTC-F14 replicon (Fig 4.3.3). Therefore, the low level of expression was presumably due to spurious background promoter activity. The upstream promoter regions of *mobB* and the *pas* operon demonstrated the highest levels of  $\beta$ -galactosidase expression in *E. coli* (Fig 4.3.3). The *pas* promoter was shown to be regulated by products of the *pas* operon, suggesting that this operon is subject to autorepression. Should the *pasAB-repAC* operon be transcribed together on a polycistronic mRNA molecule, this could indicate that auto-repression of the *pas* promoter regulates, at least in part, the expression of the *repA* and *repC* genes. However, deletion of the *pas* operon had no effect on pTC-F14 copy number (S. Deane, per comm.). RT-PCR showed that the *pas* operon is encoded on the large polycistronic mRNA transcript initiated upstream of *mobB* (Fig 4.3.2). This suggests that the large 5.1-5.7-kb polycistronic transcript that also encodes the pTC-F14 replication genes provides sufficient expression of pTC-F14 replication proteins to maintain plasmid copy number, irrespective of the presumed, more highly expressed, smaller transcript of the *pasAB-repAC* operon.

The *mobB* promoter was shown to be repressed by the pTC-F14 RepB protein (Fig 4.3.4). Removal of the primase domain (RepB) of the MobA-RepB fusion protein (pGL10-mob5) relieved the repression of this promoter, while a replication proficient



construct (pTC-KmMO) which has the relaxase domain (MobA) of the fusion protein removed, repressed the promoter (Fig 4.3.4). No region of the pTC-F14 replicon downstream of the MobA-RepB fusion protein was found to repress the promoter. Furthermore, the pTC-F14 *mobCDE* products did not appear to influence the *mobB* promoter strength (Fig 4.3.4), as repression of the *mobB* promoter by pTC-F14 (pTC-F14Cm) was found to be no greater than for RepB alone (pGL10-mob1 and pTC-KmMO). This suggested that RepB is sufficient for repression of the *mobB* promoter (Fig 4.3.4).

The mechanism of RepB repression of the pTC-F14 *mobB* promoter requires further investigation. It is worth noting that the P<sub>1</sub>/P<sub>2</sub>/P<sub>3</sub> promoter cluster of the IncQ prototype plasmid required the concerted action of both the MobC and MobA(RepB) proteins for the full repression of the promoter cluster (Frey *et al.*, 1992). The influence of the *mobCDE* gene products (Appendix D; Fig C) on repression of the pTC-F14 *mobB* promoter was not specifically investigated in this study, and it must be taken into consideration that the mobilization module of pTC-F14 is of the IncP-type, rather than the IncQ-type. The pTC-F14 mobilization module encodes five Mob proteins and thus is unrelated to the mobilization module of the IncQ prototype plasmid which only encodes three Mob proteins (Rawlings and Tietze, 2001). The *mobCDE* gene products of the related IncP-type mobilization module of pTF-FC2 (Rohrer and Rawlings, 1993) did not appear to influence transcription initiation towards the pTC-F14 *mobB* gene. Repression of the pTF-FC2 P2 promoter (transcription towards the *mobB* gene) by the pTF-FC2 MobA protein was demonstrated in a study by Rohrer and Rawlings (1993). However, the pTF-FC2 plasmid did not require the pTF-FC2 RepB (primase) protein to be fused to its MobA protein for pTF-FC2 mobilization, and so the possibility that the RepB may function as a co-repressor of the pTF-FC2 P2 promoter was not investigated by Rohrer and Rawlings (1993). In this current study there was no evidence for repression of the pTC-F14 *mobB* promoter by pTC-F14 MobA.

Regulation of the pTC-F14 putative promoter regions by related IncQ plasmids was also investigated. It was found that there was no cross-regulation of the pTC-F14 large transcript (5.1-5.7-kb) by pTF-FC2 (pDER412) or RSF1010 (RSF1010K). The closely related IncQ-like plasmid pTF-FC2 was, however, able to repress the *pas*



promoter of the pTC-F14 replicon. When pTF-FC2 plasmid was resident in *E. coli* CSH50I<sup>q</sup> (pMCF14PAS), the  $\beta$ -galactosidase activity of the pTC-F14 *pas* promoter-*lacZ* fusion was reduced by 88 % (Fig 4.3.3). The repressor has been localized to the *pas* operon of pTF-FC2, and is currently under investigation (S. Deane, per comm.).



## Chapter Five

### General Discussion

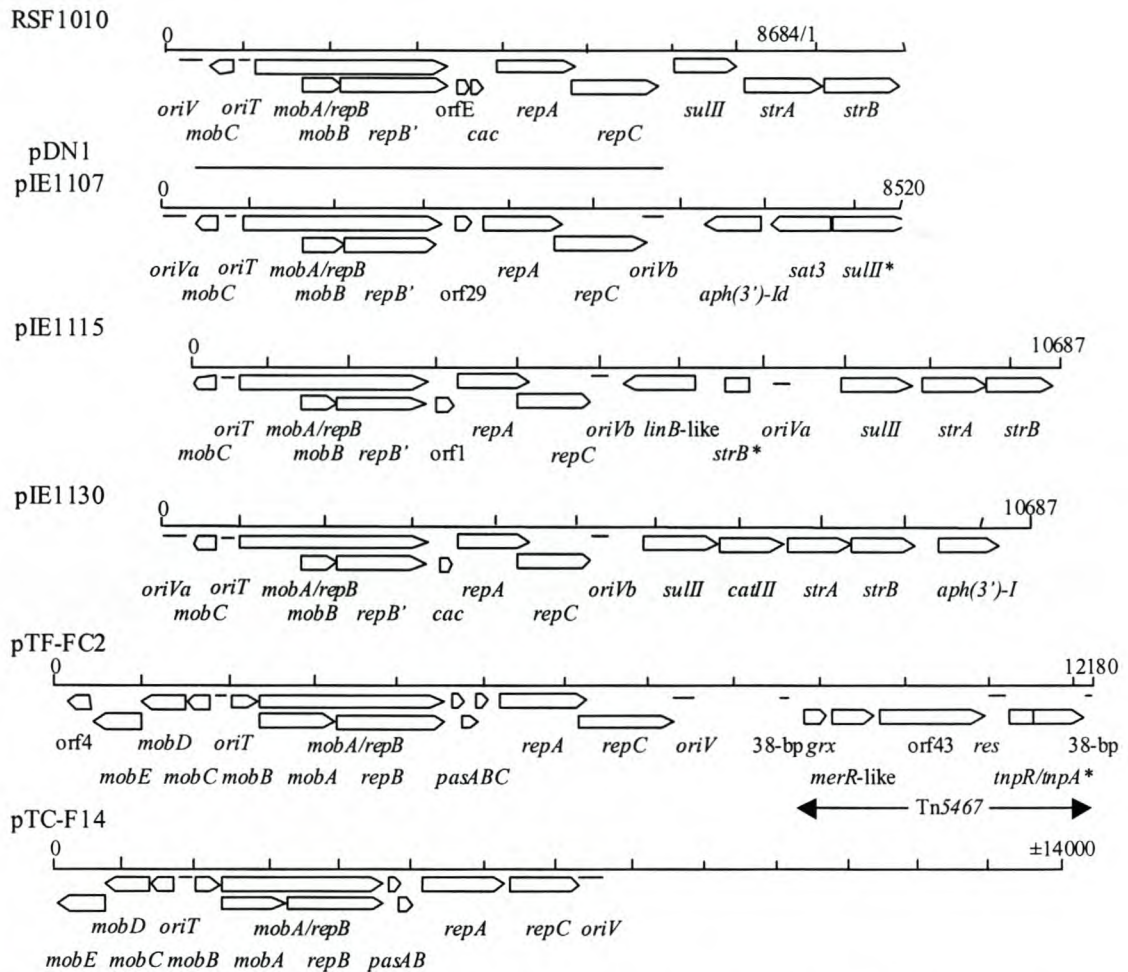
Prior to commencing this study, the only IncQ-like plasmid not isolated from human- or animal-associated environments was plasmid pTF-FC2 (Dorrington and Rawlings, 1990). The isolation of pTC-F14 from the bacterium *At caldus* strain “f” represents the second IncQ-like plasmid isolated from a biomining environment, and a variation of IncQ plasmid structure that has previously not been reported. A limited range of ecological niches have been screened for IncQ plasmids, but it has been the investigation of plasmids found in heterotrophic bacteria that has provided the largest proportion of IncQ group members. Isolation of IncQ plasmids from such diverse environments as piggery manure and the acidic, inorganic biomining environment demonstrates the promiscuity of the IncQ plasmids, and suggests that these plasmids are present in many other ecological niches. Expansion of the search for IncQ plasmids into other environments is likely to lead to the discovery of new IncQ plasmid replicons.

#### Variation in the IncQ replicon structure

Variation in the IncQ replicon structure is seen when comparing the IncQ-like plasmids referred to in this study. In light of this, brief mention needs to be made of the accessory genes carried by the IncQ-like plasmids to be discussed (Fig 5.1). The IncQ and IncQ-like plasmids encode a wide variety of antibiotic resistance genes which have been acquired either through transposon, insertion sequence, or integron activity. The different G+C ratio, codon usage, and higher prevalence of 6-bp restriction endonuclease sites than the IncQ plasmid backbone indicates that the accessory DNA has been acquired relatively recently (Rawlings and Tietze, 2001). Although antibiotic resistance genes have been commonly found associated with the IncQ plasmids, not all IncQ plasmids carry resistance genes. For instance, the accessory genes identified on pTF-FC2 include a glutaredoxin-like gene (*grx*), a *merR*-like regulator, and there is also clear evidence of transposon Tn21 activity (Fig 5.1; Clennel *et al.*, 1995). In another case, an IncQ plasmid has been isolated which



contains no accessory DNA. Plasmid pDN1, isolated from the sheep foot rot-causing pathogen *D. nodosus*, is currently the smallest IncQ-like plasmid, consisting solely of the IncQ plasmid backbone, and consequently is only 5112-bp in size (Whittle *et al.*, 2000).



**Fig 5.1:** Genetic maps of IncQ plasmids referred to in this study. Plasmid pDN1 is represented by a line since its genetic map is identical to the plasmid backbone of pIE1107. The *repB*, *repA* and *repC* replication module genes, and *mobA*, *mobB*, *mobC*, *mobD* and *mobE* genes of the mobilization module together form part of the plasmid backbone. Other genes found in the plasmid backbone include: *cac*, control of *repA* and *repC* regulator; *orfE*, unknown function; *orf29*, unknown function; *orf1*, unknown function; *pasA*, *pasB* and *pasC*, plasmid addiction system genes. The location of the *oriV* (origin of vegetative replication), and *oriT* (origin of transfer) regions on each plasmid is indicated. Accessory genes identified on IncQ plasmids include: *aph(3')-I*, kanamycin and neomycin aminoglycoside phosphotransferase; *catIII*, chloramphenicol acetyltransferase; *grx*, glutaredoxin-like gene; *linB*-like, lincosamide nucleotyltransferase; *merR*-like, *merR*-like regulator; *strA* and *strB*, streptomycin aminoglycoside phosphotransferase; *sullI*, sulfonamide-resistant dihydropteroate synthase; *tnpA*, transposase; *tnpR*, resolvase; *res*, site of cointegrate resolution by transposon resolvase; *orf4*, unknown function, *orf43*; unknown function. Genes which have been inactivated by deletions are indicated by asterisks. Reproduced from Rawlings and Tietze (2001).

In comparing the IncQ plasmid backbones, the IncQ prototype plasmid RSF1010/R1162/R300B, as well as pDN1, pIE1107, pIE1115, and pIE1130 display



similarity in the structural organization of the replication and mobilization modules (Fig 5.1). General similarities of the above plasmid backbones include an overlapping *repA* and *repC* gene organization, and synthesis of the MobB protein from within the MobA-RepB coding sequence. These plasmids also have an IncQ-type mobilization module that consists of three *mob* genes (Fig 5.1). In comparing the pTF-FC2 plasmid structure with the above IncQ-like plasmids, the *repA* and *repC* genes also overlap, but the *mobB* coding sequence is independently located upstream of the *mobA/repB* coding sequence. The pTF-FC2 plasmid also has a mobilization module of the IncP-type which consists of five *mob* genes that are unrelated to the *mob* genes of the IncQ prototype plasmid (Rawlings and Tietze, 2001). A further variation on the pTF-FC2 plasmid structure is pTC-F14 which does not have overlapping *repA* and *repC* genes, but also has an IncP-type mobilization module.

Another feature of the IncQ plasmid backbone is the identification of between one and three open reading frames situated between the *repB* and *repA* genes (Fig 5.1). The IncQ prototype plasmid RSF1010/R1162/R300B has two open reading frames at this locus. The function of the IncQ prototype plasmid's ORF E gene is unknown, but the *cac* gene is thought to be responsible for the transcriptional regulation of RSF1010/R1162/R300B *repAC* expression (Maeser *et al.*, 1990). Only plasmid pIE1130, isolated from an unidentified bacterium in piggery manure, contains an open reading frame with sequence similarity to the *cac* gene. Plasmids pIE1107 and pIE1115 were also isolated from an animal-associated environment and, like pIE1130, were found to have a single open reading frame at this locus (Tietze, 1998; Smalla *et al.*, 2000[a]). The coding sequence at this locus on pIE1107 and pIE1115 is unrelated to the *cac* gene and consequently has no known function. Similarly, a single open reading frame of unknown function, and designated ORF78, is found at the same locus on the pDN1 replicon (Whittle *et al.*, 2000; Fig 5.1). The two biomining environment IncQ plasmid isolates, pTF-FC2 and pTC-F14, however, contain a proteic plasmid addiction system (*pas*) at this locus on the IncQ plasmid backbone. This raises the question of whether the *pas* genes are situated here merely by chance, or whether they have a role related to replicon function. Evidence suggests that the *pas* genes may not necessarily have a replicon-related role. Firstly, the *pas* genes of both pTF-FC2 (Smith and Rawlings, 1997) and pTC-F14 (S. Deane, per comm.) could be deleted without a noticeable change in copy number. Secondly, the *pasAB*



operon has recently been identified on a plasmid isolated from *P. fluorescens*, but the plasmid, designated pAM10.6, does not have an IncQ-like replicon (Peters *et al.*, 2001). The identification of a *pasAB* operon on an unrelated replicon (pAM10.6) suggests that this *pasAB* operon is a module that may be acquired by different replicons, and thus not uniquely associated with IncQ plasmids.

The products of the pAM10.6 *pasAB* operon show significant amino acid identity with the PasA of pTF-FC2 (79.7 %) and pTC-F14 (77 %), and with the PasB of pTF-FC2 (77.5 %) and pTC-F14 (74.2 %). However, unlike pTF-FC2 which has a third *pas* gene (*pasC*), the *pasAB* operons of pTC-F14 and pAM10.6 do not have an associated *pasC* gene. The product of the *pasC* gene is believed to enhance the ability of PasA to neutralize the PasB toxin (Smith and Rawlings, 1997), so the lack of a *pasC* gene in the pTC-F14 and pAM10.6 *pasAB* operon is intriguing. This raises the question as to whether the *pasAB* operon is less evolved than the *pasABC* operon of pTF-FC2, or whether pTC-FC2 *pasC* is a fortuitous development after the insertion of this plasmid stability module in the pTF-FC2 replicon. Alternatively, the *pasAB* operon could be a more evolved and refined form of the original *pasABC* stability system, resulting from the loss of *pasC* from this plasmid addiction module. Current work suggests that the *pasABC* configuration is more effective at plasmid stabilization than *pasAB* (S. Deane, per comm.).

When comparing other features of the IncQ replicon, one notes variation in the location and number of *oriV* regions on each IncQ plasmid, as well as the number and sequence of direct repeats in the iteron containing region of each origin (see section 3.3.1; Fig 5.1). The IncQ prototype plasmid RSF1010/R1162/R300B, pDN1, pTF-FC2 and pTC-F14 all have one *oriV*, but only RSF1010/R1162/R300B does not have the *oriV* located proximally to the *repC* gene (Fig 5.1). The dislocation of the IncQ prototype *oriV* from its *repC* gene indicates that there is no spatial requirement for the location of the *oriV* relative to the *repC* gene, although Henderson and Meyer (1999) reported that spatial variation of the *oriV* relative to the *oriT* affected conjugal transfer frequencies. In contrast to the above plasmids, plasmids pIE1107, pIE1115 and pIE1130 have two *oriV*-like regions. Osborn *et al.* (2000) used the model of plasmid speciation through cointegrate formation (Sýkora, 1992; see section 1.5) to try to address how a plasmid may acquire more than one copy of *oriV* DNA. To apply this



model to pIE1107, pIE1115 and pIE1130, one needs to assume cointegrate formation between two different IncQ-like plasmids. This is then followed by the partial deletion of one of the cointegrate replicons such that the *oriV* of the deleted cointegrate replicon is retained by the new plasmid. The potential consequence of acquiring a second origin is clearly demonstrated by pIE1107. Plasmid pIE1107 has two *oriV*-like regions, one of which was responsible for the displacement of RSF1010 from a host. It was the non-functional pIE1107 *oriVa* region which was found to be the incompatibility determinant, while the replication-active *oriVb* region expressed no incompatibility towards RSF1010 (Tietze, 1998). Therefore, the acquisition of a second *oriV* region may contribute to a plasmid's competitive advantage in a plasmid-rich environment. That is, a multi-origin plasmid potentially carries more than one incompatibility determinant that can act against plasmids attempting to establish themselves in the same host as the multi-origin plasmid. Smalla *et al.* (2000[a]) suggested that there was a high prevalence of IncQ plasmids in the manure environment from which pIE1107, pIE1115 and pIE1130 were isolated. Therefore, a high prevalence of related replicons implies strong competition for the “replication space” in this environment.

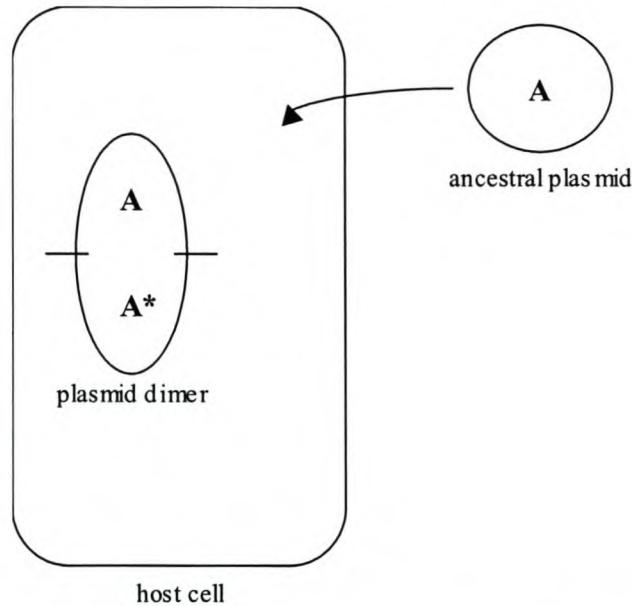
### **Evolution of a new incompatibility group, and implications for IncQ plasmid classification.**

The consequence of origin pairing (‘handcuffing’) is the inhibition of a new round of replication (see section 1.3.2.5). Origin pairing of incompatible iteron-regulated plasmids through cross-recognition of the related plasmid's iteron sequence by RepC orthologs interferes with copy number control, and can result in plasmid loss from the host population – incompatibility. Therefore, compatible plasmids suggest the ability of the RepC protein to distinguish between the parental plasmid's iteron sequence, and the iteron sequence of a related plasmid replicon.

The plasmid speciation model of Sýkora (1992) could explain how the two compatible replicons of pTF-FC2 and pTC-F14 arose. To do this, one must assume that a dimer of an ancestral IncQ plasmid formed through either recombination of plasmid progeny monomers, or defective replication termination (Snyder and Champness, 2003). Pressure for one of the dimer's replicons to begin to acquire



mutations leading to a sufficiently different replicon to the original replicon, could be provided by the dimer being challenged by an ancestral, and identical, plasmid attempting to displace the plasmid dimer from the host (Fig 5.2).



**Fig 5.2:** An ancestral and identical (incompatible) plasmid (A) attempting to occupy the same host as a dimer of the ancestral plasmid, places the plasmid dimer under pressure to acquire mutations in a replicon (A\*) so as to lead to a compatible and functional replicon (A\*) to prevent displacement of the dimer.

The mutations acquired in the second replicon (A\*) of the dimer would have to allow for both an unrelated and functional replicon to evolve, so that the second replicon of the dimer is compatible with the ancestral plasmid, and ensures maintenance of the dimer in the presence of the ancestral plasmid. Resolution of the plasmid dimer after mutation of the dimer's second replicon to an unrelated, but functional, replicon would lead to two compatible sister plasmids. Such a mechanism for plasmid speciation could have given rise to the compatible plasmids pTF-FC2 and pTC-F14.

Irrespective of whether compatible plasmids can arise through the mechanism described above, or if it is the consequence of independent replicon evolution, this study proposes that the driving force for the acquisition of mutations that led to IncQ replicon compatibility is the same. It is believed that concomitant evolution of an IncQ plasmid's iteron sequence and its RepC protein leads to compatible replicons through the ability of the RepC proteins to now distinguish between the parental plasmid and a related replicon (see section 3.4.2).



In addition, the iteron-mediated ‘handcuffing’ copy number control model predicts that iteron-bound RepC proteins interact through a protein-interaction domain to affect ‘handcuffing’ (Chattoraj, 2000). In accordance with this model, the protein binding domain of the RepC protein must, therefore, alter sufficiently so that the RepC orthologs no longer recognize each other.

The incompatibility-based plasmid classification system requires that only incompatible plasmids belong to the same group. As found by Smalla *et al.* (2000[a]), and as demonstrated in this study, this system for the classification of IncQ plasmids is not workable. To account for the incompatibility phenotypes expressed (Table 3.3.1) and the phylogenetic relationships of the IncQ replication proteins (Fig 2.2.3), it is now proposed that the IncQ group be divided such that plasmids can be classified as IncQ1 $\alpha$  (RSF1010-type), IncQ1 $\beta$  (pIE1107-type), IncQ1 $\gamma$  (pIE1130-type), IncQ2 $\alpha$  (pTF-FC2-type), and IncQ2 $\beta$  (pTC-F14-type) (see Gardner, M.N., S.M. Deane and D.E. Rawlings (2001) *J Bacteriol* 183: 3303-3309).

The major division of the IncQ group into IncQ1 and IncQ2 is supported by the phylogenetic relationship of the replication proteins. Amino acid sequence alignment of the replication proteins clearly divides the IncQ plasmid family into two groups (Fig 2.2.3). An additional criterion by which plasmids of the IncQ2 group (pTF-FC2 and pTC-F14) can be separated from the IncQ1 plasmids is the type of mobilization module present on the plasmid. The IncQ2 group plasmids have an IncP-type mobilization module (*mobABCDE*), while the mobilization module of IncQ1 group plasmids have an IncQ-type (*mobABC*).

One can speculate as to what would be required for the establishment of an IncQ3 group. Using only the phylogenetic relationships of RepC orthologs, an IncQ3 group could be established if the RepC protein of a new IncQ-like replicon is clearly unrelated to either of the two groups determined (Fig 2.3.2). Due to the conservation of the RepC protein sequence (62 %), relative to the RepA (44 %) and RepB (15 %) proteins, it is proposed that RepC be used for group classification, with the phylogenetic relationships between the RepA and RepB orthologs supporting the



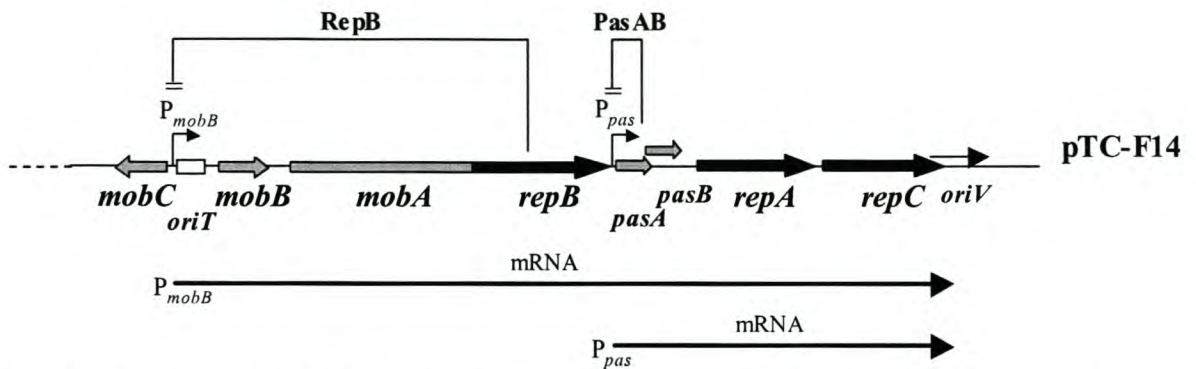
assignment. If the sequence of a newly isolated IncQ-like plasmid indicated a deep division in the phylogenetic relationship of the RepB, RepA, and RepC orthologs, the question that would arise is, how deep should that division be to warrant the recognition of an IncQ3 group as opposed to a new subdivision of IncQ1 or IncQ2? The suggestion is that all IncQ family plasmids that have an IncQ-type mobilization module are placed in the IncQ1 group, and all those with an IncP-type mobilization module be placed into the IncQ2 group; irrespective of the deepness of the division of their RepB, RepA, or RepC proteins. Therefore, the depth of phylogenetic subdivisions when comparing replication protein amino acid sequences are secondary considerations. Establishment of an IncQ3 group would most likely require that the plasmid isolate carries a new mobilization module that has not yet been described for IncQ-like plasmids. Alternatively, there may be a new theme in the arrangement of a replicon that has retained many of the characteristics of IncQ-like plasmids.

The occurrence of the *pas* genes has little phylogenetic value. For example, *Aeromonas salmonicida* was recently found to harbour an IncQ-like plasmid, pRAS3, which has a replicon that is closely related to pTF-FC2, has an IncP-type mobilization module (as does pTF-FC2), but does not have a *pas* system (L'Abée-Lund and Sørum, 2002). Based on the phylogeny of its replication proteins and its mobilization module, plasmid pRAS3 is an IncQ2 group plasmid.

### **Regulation of the pTC-F14 replicon.**

Analysis of the transcripts produced from the pTC-F14 replicon suggested that a large 5.1-5.7-kb transcript is initiated upstream of the pTC-F14 *mobB* gene and terminates downstream of the *repC* gene (Fig 4.3.1 and Fig 4.3.2). DNA sequence analysis of the pTC-F14 *oriT* region found upstream of *mobB* identified a promoter sequence (TTGACT-N<sub>16</sub>-TACAAT) with good DNA sequence similarity to the *E. coli*  $\sigma^{70}$  RNAP consensus sequence. Promoter-reporter gene (*lacZ*) studies suggested that this promoter was regulated by the product of the *repB* gene (Fig 4.3.4; Fig 5.3).





**Fig 5.3:** Regulation of the pTC-F14 replicon. The two polycistronic mRNA transcripts believed to be initiated from the promoters ( $P_{mobB}$  and  $P_{pas}$ ) are shown. The RepB protein was identified as a regulator of the  $P_{mobB}$  promoter, while the  $P_{pas}$  promoter is autoregulated by a product(s) of the *pas* operon.

Rohrer and Rawlings (1993) demonstrated that the P2 promoter of pTF-FC2 (transcript initiation towards *mobB*) was repressed by the MobA protein. These authors determined that the pTF-FC2 plasmid did not require the pTF-FC2 RepB (primase) protein to be fused to its MobA protein for pTF-FC2 mobilization. However, the IncQ prototype plasmid primase (RepB) was found to be active in conjugal transfer, as it is required for the initiation of complementary strand synthesis following transfer of the ssDNA molecule through the conjugal pore (Henderson and Meyer, 1996; Henderson and Meyer, 1999). Henderson and Meyer (1996) had noted from earlier investigations into IncQ mobilization that these studies had not accounted for the possible priming of the ssDNA molecule by components of the cloning vector into which the IncQ mobilization module had been inserted. Similarly, the *mobABCDE* genes of the pTF-FC2 mobilization module were cloned into the pACYC184 cloning vector (Rohrer and Rawlings, 1993), and by doing so, this may have obscured the role pTF-FC2 RepB plays in mobilization of this plasmid. Consequently, the effect of pTF-FC2 RepB on regulation of the P2 (*mobB*) promoter was never tested.

The RepB protein of pTC-F14 could potentially participate in plasmid mobilization. The binding of RepB to the  $P_{mobB}$  promoter, which is within the *oriT* region of pTC-F14 (Fig 5.3), may position the RepB protein for priming of the ssDNA molecule during conjugal transfer. By binding to this region, RepB probably blocks access for RNAP to the promoter, thereby repressing expression of the replication genes.



With RepB possibly providing the link between pTC-F14 mobilization and replication, it still remains to be determined how pTC-F14 regulates mobilization independently of replication. A plasmid correcting its copy number in a host does not require activation of mobilization, yet for pTC-F14 the *mobB* and *mobA* genes are believed to be transcribed on the same mRNA as the replication genes (Fig 5.3; see Chapter 4). Therefore, it is speculated that in the case of pTC-F14, replication is regulated by RepB, but that the *mobCDE* gene products regulate mobilization. It was found for pTF-FC2 that the promoter for the *mobCDE* gene cluster was 100 times stronger than the *mobB/mobA/repB* promoter (Rohrer and Rawlings, 1993). This pTF-FC2 *mob*-specific *mobCDE* promoter was autoregulated by MobC, which suggests that regulation of *mobCDE* gene expression sets the mobilization frequency.

Initiation of transcription from the pTC-F14  $P_{pas}$  promoter (Fig 5.2) was found to be autoregulated by products of the *pas* operon (Fig 4.3.3). We were unable to determine unequivocally if the *pasAB-repAC* operon was transcribed on a single polycistronic mRNA by Northern blot or RT-PCR analysis, but deletion of the pTC-F14 *pas* operon did not affect plasmid copy number (S.Deane, per comm.). The pTC-F14  $P_{pas}$  promoter was, however, found to be cross-regulated by products of the pTF-FC2 replicon (Fig 4.3.3), and this response is currently under further investigation in our laboratory.

Apart from products of the pTF-FC2 replicon repressing the pTC-F14  $P_{pas}$  promoter, cross-regulation was not found for any of the other pTC-F14 putative promoter regions (Fig 4.3.3). Specifically, the RepB of pTF-FC2 was not able to regulate the large *mobB-repC* transcript of pTC-F14. However, the pTF-FC2 RepB was able to substitute for the pTC-F14 RepB in complementation assays (Table 3.3.3). This may suggest that the RepB orthologs have two domains: (i) a primase domain, and (ii) a regulatory domain. The regulatory functions of the RepB orthologs appear to be plasmid specific, but evidently, the primase functions have not evolved sufficiently to be plasmid specific. To support this suggestion, one needs to determine the reverse of what was tested in this study. One needs to test if the RepB of pTC-F14 can regulate pTF-FC2 *mobB* gene expression. It appears, therefore, that, like the RepC orthologs, the RepB orthologs have also evolved apart, as closely related RepB proteins would have resulted in plasmid incompatibility.



## Future studies

Although there are many avenues for research into the replication of IncQ family plasmids, the regulation of replication gene expression evidently remains one aspect that requires further investigation. The possible regulatory role of the open reading frame found on some IncQ-like plasmids at the same locus as the RSF1010/R1162/R300B ORF E and *cac* genes needs to be determined. As discussed above, some of these open reading frames have no homology to the ORF E or *cac* gene of RSF1010/R1162/R300B, and consequently no apparent function. The observation that the plasmid addiction system (*pas* genes) module, located in similar positions on pTF-FC2 and pTC-F14, can be deleted without affecting copy number, suggests that the *pas* module does not play a replicon regulatory role. The possibility exists that the *pas* system may play a more subtle regulatory role. However, no evidence for this has been found.

There is also the possibility of an additional *pasAB-repAC* transcript being produced by pTC-F14 (see Chapter 4). Real-time PCR now provides the means to overcome the problems experienced in this study, and would allow for the determination of the number of pTC-F14 replicon transcripts, and their relative size. This quantitative PCR could provide information on the expression levels of the different transcripts, and whether this differs between hosts.

Further regulation studies would also assist in resolving whether pTC-F14 mobilization is regulated independently of plasmid replication. By determining the dependency of the IncP-type mobilization system on replication, a general regulatory mechanism for IncQ2 group plasmids may become apparent.

On a personal note, I am intrigued by the process by which compatible IncQ plasmids arise. Site directed mutagenesis of the suggested iteron-discriminating region, together with iteron binding studies using purified RepC proteins from incompatible IncQ replicons, could confirm that the affinity of RepC orthologs for related iterons decreases with a change in iteron sequence away from the consensus. This would clearly demonstrate how RepC specificity for the parental iteron sequence can allow



the RepC proteins to distinguish between related replicons, and provide evidence for iteron-RepC alteration leading to plasmid compatibility.



**Appendix A: Growth media, additives, buffers and solutions**

All media, solutions and buffers were sterilized by autoclaving at 121°C for 20 minutes. Heat labile substances were filter sterilized using 0.22 µm membrane filters (Micron separations Inc.)

**A) Luria-Bertani medium**

Bactotryptone	10 g.l <sup>-1</sup>
Yeast extract	5 g.l <sup>-1</sup>
NaCl	5 g.l <sup>-1</sup>

Solid media contained 1.5 % (w/v) agar.

**B) H-Plate medium**

Bactotryptone	10 g.l <sup>-1</sup>
NaCl	8 g.l <sup>-1</sup>
Agar	1.5 % (w/v)

**C) Top agarose (overlay medium for H-plates)**

Bactotryptone	10 g.l <sup>-1</sup>
NaCl	8 g.l <sup>-1</sup>
Agarose	0.6 % (w/v)

**D) Media additives**

Medium additive stocks were made as follows:

Ampicillin (Ap)	100 mg.ml <sup>-1</sup>
Chloramphenicol (Cm)	20 mg.ml <sup>-1</sup>
Kanamycin (Km)	30 mg.ml <sup>-1</sup>
Tetracycline (Tc)	20 mg.ml <sup>-1</sup>

These stocks were stored at -20°C.

X-gal (5-bromo-4-chloro-3-indolyl-β-galactoside)

X-gal (2 % w/v)	0.2 g
Dimethylformamide	10 ml

The solution was stored in aliquots at -70°C



**E) Buffers and solutions**

## 10X MOPS electrophoresis buffer (RNase-free)

MOPS	41.85 g
NaOAc.3H <sub>2</sub> O	6.8 g

Make to 800 ml with DEPC-treated water. Add 20 ml 0.5 M EDTA pH 8.0 and adjust to pH 7.0 with 10 N NaOH. Make to 1 liter, autoclave and store in the dark at 4°C.

## RNA loading buffer (v/v)

formamide	56 %
12.3 M formaldehyde	19 %
10X MOPS buffer	11 %
1 % Bromophenol Blue	1.0 %
ethidium bromide	2.0 %

## Z-buffer

Na <sub>2</sub> HPO <sub>4</sub>	4.26 g
NaH <sub>2</sub> PO <sub>4</sub> .2H <sub>2</sub> O	3.12 g

Dissolve in 450 ml water, adjust to pH 7.0, and make to 480 ml. Autoclave

Dissolve 0.375 g KCl in 10 ml water, and autoclave.

Dissolve 0.123 g MgSO<sub>4</sub>.7H<sub>2</sub>O in 10 ml water, and autoclave.

Combine the above solutions, and add 1.35 ml β-mercaptoethanol.

ONPG-(*o*-nitrophenyl-β-D-galactoside) stock (4 mg.ml<sup>-1</sup>)

ONPG	0.4 g
------	-------

Dissolve in 100 ml 0.1 M Phosphate buffer pH 7.0. Store at -20°C in the dark.

0.1 M Phosphate buffer is prepared by combining 61 ml of 0.2 M Na<sub>2</sub>HPO<sub>4</sub> with 39 ml 0.2 M NaH<sub>2</sub>PO<sub>4</sub>.2H<sub>2</sub>O, and then making to 200 ml with water.



**Appendix B: Strains, plasmids, and bacteriophage used in this study**

Strain, plasmid or phage	Description <sup>a</sup>	Reference or source
<b>Strains</b>		
<i>Acidithiobacillus caldus</i> strain “F”	Isolated from BHP Billiton Minerals Technology Bionic plant (Randburg, South Africa)	This study
<i>Agrobacterium tumefaciens</i> LBA4404	Rif <sup>r</sup> , Nononcogenic, T-DNA deletion, <i>vir</i> intact	Hoekema <i>et al.</i> (1983)
<i>Escherichia coli</i> CSH501 <sup>q</sup>	<i>rspL</i> Δ( <i>lac-pro</i> ) (F’ <i>traD36 proAB lacI<sup>f</sup></i> ΔM15)	Smith and Rawlings (1998)
<i>E. coli</i> DH5α	φ80 <i>dlacZ</i> ΔM15 <i>endA1 recA1 gyrA96 thi-1 hsdR17</i> (r <sub>k</sub> <sup>-</sup> , m <sub>k</sub> <sup>+</sup> ) <i>relA1 supE44 deoR</i> Δ( <i>lacZYA-argF</i> ) U169	Promega Corp, Madison
<i>E. coli</i> GW125a	<i>recA</i> , <i>polA</i> mutant of AB1157	Dorrington and Rawlings (1989)
<i>E. coli</i> XL1-Blue	F’::Tn10 <i>proA<sup>+</sup>B<sup>+</sup></i> Δ( <i>lacZ</i> )M15/ <i>recA1 endA1 gyrA96</i> (Nal <sup>r</sup> Tet <sup>r</sup> ) <i>thi hsdR17</i> (r <sub>k</sub> <sup>-</sup> , m <sub>k</sub> <sup>+</sup> ) <i>relA1 supE44</i>	Stratagene
<i>Pseudomonas putida</i>	Prototrophic	Franklin <i>et al.</i> (1981)
<b>Plasmids</b>		
pTC-F14	Natural 14.4-kb plasmid from <i>Acidithiobacillus caldus</i> strain “F”	This study (Appendix D)
pTC-F14Cm	Cm <sup>r</sup> (Cm <sup>r</sup> gene cloned into pTC-F14)	This study
pIb	Ap <sup>r</sup> (unique <i>XbaI</i> site of pTC-F14 interrupted by pBluescript)	This study
pK13	Ap <sup>r</sup> (unique <i>BamHI</i> site of pTC-F14 interrupted by pBluescript)	This study
pTF-FC2	Natural 12.2-kb plasmid from <i>Acidithiobacillus ferrooxidans</i> FC2.	Rawlings and Woods (1985)
pKE462	Tc <sup>r</sup> , As <sup>r</sup> , R300B replicon	Dorrington and Rawlings (1989)
RSF1010K	Km <sup>r</sup> , 1-1704-bp of RSF1010 replaced by Tn903	G. Ziegelin
pIE1108	St <sup>r</sup> , Km <sup>r</sup> , pIE1107 replicon with a set of non-essential IncQ iterons deleted	Tietze (1998); Appendix D
pIE1108Cm	Cm <sup>r</sup> , Region (St <sup>r</sup> , Km <sup>r</sup> ) replaced with Cm <sup>r</sup> gene	This study
pDER404	Cm <sup>r</sup> , pTF-FC2 <i>ClaI-PstI</i> fragment complete replicon	Rawlings and Woods (1985)
pDER412	Cm <sup>r</sup> , pTF-FC2 plasmid with pBR325 Cm <sup>r</sup> gene	Rawlings <i>et al.</i> (1984)
pTV4164	Ap <sup>r</sup> , pUC19 with <i>oriV</i> <sub>pTF-FC2</sub> fragment	Dorrington <i>et al.</i> (1991)
pTC-F101	Km <sup>r</sup> , <i>HindIII</i> to <i>SphI</i> pTC-F14 replicon fragment	This study
pTC-F108	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with pTC-F101 <i>HindIII</i> to <i>PstI</i> fragment	This study
pTC-F109	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified <i>oriV</i> <sub>pTC-F14</sub> (bp 7363-8100)	This study
pTC-F109E	Ap <sup>r</sup> , pGEM-T <sup>®</sup> Easy vector with PCR amplified <i>oriV</i> <sub>pTC-F14</sub> (bp 7363-8100)	This study
pACYC184	Tc <sup>r</sup> , Cm <sup>r</sup> , p15A replicon, cloning vector	Chang and Cohen (1978)
pBluescript(KS)	Ap <sup>r</sup> , <i>LacZ'</i> , ColE1 replicon, vector	Stratagene



Strain, plasmid or phage	Description <sup>a</sup>	Reference or source
pGEM-T <sup>®</sup>	Ap <sup>r</sup> , T-tailed PCR product cloning vector	Promega Corp
pGEM-T <sup>®</sup> Easy	Ap <sup>r</sup> , T-tailed PCR product cloning vector with <i>EcoRI</i> site in multiple cloning site	Promega Corp
pKK223-3	Ap <sup>r</sup> , P <sub>lac</sub> , ColE1 replicon, expression vector	Pharmacia Biotech
pGL10	Km <sup>r</sup> , RK2/RP4 replicon, cloning vector	A. Toukdarian (Appendix D)
pMC1403	Ap <sup>r</sup> , promoterless <i>lacZYA</i> operon, ColE1 replicon	Casadaban <i>et al.</i> (1983)
pUC19	Ap <sup>r</sup> , <i>LacZ'</i> , ColE1 replicon, cloning vector	Yanisch-Perron <i>et al.</i> (1985)
pUCBM21	Ap <sup>r</sup> , <i>LacZ'</i> , ColE1 replicon, cloning vector	Roche Molecular Biochemicals
pGEMRSFORIV	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified <i>oriV</i> <sub>RSF1010</sub> fragment (bp 2209-2772)	This study
pGEMRSFORIVT	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified RSF1010 <i>oriV</i> to <i>oriT</i> fragment (bp 2209-3364)	This study
pGEMRSFORIT	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified <i>oriT</i> <sub>RSF1010</sub> fragment (bp 2755-3364)	This study
pGEMRSFREP	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified RSF1010 <i>mobA</i> to <i>repC</i> fragment (bp 3161-7725)	This study
pGEMRSFREP1	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified RSF1010 <i>mobA</i> to <i>cac</i> fragment (bp 3161-5926)	This study
pGEMRSFCAC	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified RSF1010 ORF E and <i>cac</i> fragment (bp 5271-5926)	This study
pGEMRSFREP2	Ap <sup>r</sup> , pGEM-T <sup>®</sup> vector with PCR amplified RSF1010 ORF E to <i>repC</i> fragment (bp 5271-7725)	This study
pGL10-C	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repC</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 6743-7756)	This study
pGL10-AC	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repAC</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 5786-7756)	This study
pGL10-A	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repA</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 5786-6727)	This study
pGL10-BA	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repBA</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 4057-6727)	This study
pGL10-B	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repB</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 4057-5199)	This study
pGL10-BAC	Km <sup>r</sup> , pGL10 containing PCR amplified pTC-F14 <i>repBAC</i> cloned behind the P <sub>lac</sub> promoter; acquired from pKK223-3 (bp 4057-7756)	This study
pTC-KmMO	Km <sup>r</sup> , PCR amplified pTC-F14 <i>repBAC</i> ligated to Km <sup>r</sup> gene (bp 4057-8100)	This study
pMCF14MOBB	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>mobB</i> promoter region cloned into pMC1403 (bp 1859-2223)	This study



Strain, plasmid or phage	Description <sup>a</sup>	Reference or source
pMCF14MOBB2	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>mobB</i> promoter region cloned into pMC1403 (bp 2052-2223)	This study
pMCF14MOBA	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>mobA</i> promoter region cloned into pMC1403 (bp 2294-2689)	This study
pMCF14REPB	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>repB</i> promoter region cloned into pMC1403 (bp 3716-4135)	This study
pMCF14PAS	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>pas</i> promoter region cloned into pMC1403 (bp 4956-5263)	S. Deane
pMCF14REPA	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>repA</i> promoter region cloned into pMC1403 (bp 5481-5878)	This study
pMCF14REPC	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>repC</i> promoter region cloned into pMC1403 (bp 6423-6823)	This study
pTC-F101ΔAB	Km <sup>r</sup> , <i>StuI-XbaI</i> deletion of pTC-F101 <i>pas</i> operon	S. Deane
pGL10-mob1	Km <sup>r</sup> , PCR amplified pTC-F14 <i>mobE</i> to <i>mobA(repB)</i> cloned into pGL10 (bp 70-5240)	L. van Zyl
pGEM-mob5	Ap <sup>r</sup> , PCR amplified pTC-F14 <i>mobE</i> to <i>mobA</i> (truncated <i>repB</i> ) cloned into pGEM-T <sup>®</sup> (bp 70-4375)	L. van Zyl
pGL10-mob5	Km <sup>r</sup> , <i>EcoRI-PstI</i> cloning of pTC-F14 <i>mobE</i> to <i>mobA</i> from pGEM-mob5	This study
<b>Bacteriophage</b>		
M13Δ <i>lac</i> 110	Partial 105-bp <i>ssi</i> deletion of phage M13	Honda <i>et al.</i> (1989)
M13ΔF109E	M13Δ <i>lac</i> 110 into which <i>oriV<sub>pTC-F14</sub></i> was cloned from pTC-F109E	This study

<sup>a</sup> Ap, ampicillin; Cm, chloramphenicol; Km, kanamycin; Rif, rifampicin; St, streptothricin and Tc, tetracycline resistance.



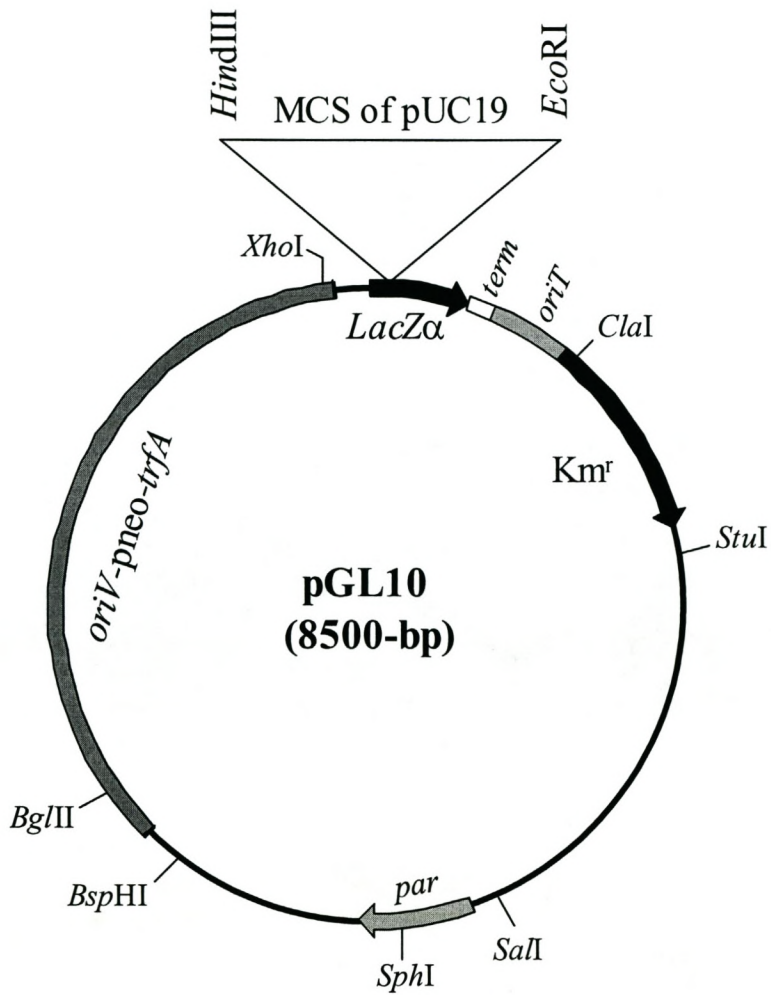
**Appendix C: Primers used in this study**

Primer <sup>b</sup>	Sequence	Source
TACREPB ( <i>EcoRI</i> )	5' -TCAGGAATTC <u>CCCCGGAGCTTCAG</u> -3'	This study
TACREPBE ( <i>PstI</i> )	5' -TTCTCTGCAGTCATGCCGCTGTG-3'	This study
TACREPA ( <i>EcoRI</i> )	5' -TATTGAATTC <u>CCCCGGCAGCGCC</u> -3'	This study
TACREPAE ( <i>PstI</i> )	5' -TATTCTGCAGAGGGGGTGCATAGC-3'	This study
TACREPC ( <i>EcoRI</i> )	5' -TCGCGAATTCGTGTGGCTATACCCAG-3'	This study
TACREPCE ( <i>HindIII</i> )	5' -TCTGAAGCTTGCTGGCTTAGCGTG-3'	This study
SEQORI	5' -TATCGAGATGGCAGAGGTGCGAG-3'	This study
ORIR ( <i>HindIII</i> )	5' -TGTCAAGCTTGGCACTCTCCTG-3'	This study
INCQVFOR ( <i>EcoRI</i> )	5' -TATGAATTCGGCATGTTCCGCGTCC-3'	This study
INCQVREV ( <i>PstI</i> )	5' -TCACTGCAGGCTGAATGATCGACC-3'	This study
INCQTFOR ( <i>EcoRI</i> )	5' -AGCGAATTCGGTGCATCATTTCAGCC-3'	This study
INCQTREV ( <i>SalI</i> )	5' -TCTAGTCGACCTTCATCCATGTCGC-3'	This study
INCQREPF ( <i>XbaI</i> )	5' -TGACTCTAGAACTGGCCTAACGGC-3'	This study
INCQREPR ( <i>SalI</i> )	5' -TATAGTCGACCTATGGAGCTGTGCG-3'	This study
INCQCACF ( <i>XbaI</i> )	5' -ACTGTCTAGAAGGTCATGGGTCTGC-3'	This study
INCQCACR ( <i>SalI</i> )	5' -TGCTGTGACGCCTCCAGAATATTG-3'	This study
LACF14MOBBF ( <i>EcoRI</i> )	5' -TGT <u>CGAATTC</u> CAGTGCAGTCCCTGC-3'	This study
LACF14MOBBF2 ( <i>EcoRI</i> )	5' -ACAC <u>CGAATTC</u> TACGCATGTTCCCGC-3'	This study
LACF14MOBBR ( <i>BamHI</i> )	5' -AGGAGGATCCACGGTAAATGGCATC-3'	This study
LACF14MOBAF ( <i>EcoRI</i> )	5' -TAT <u>CGAATTC</u> CAGGATGCGGACCTCG-3'	This study
LACF14MOBAR ( <i>BamHI</i> )	5' -TCTAGGATCCAGGGCTACCATTTCC-3'	This study
LACF14REPBF ( <i>EcoRI</i> )	5' -TACAGAATTCAGGCGACTAGGTGG-3'	This study
LACF14REPBR ( <i>BamHI</i> )	5' -TCAGGGATCCTGCTCCGATTTTCATC-3'	This study
PRIMER5 ( <i>EcoRI</i> )	5' -TACTGAATTC <u>TACCAGTGTGCCCATCG</u> -3'	S. Deane
PRIMER6 ( <i>BamHI</i> )	5' -GTAGGGATCCACTTCGGTGGTAATCGG-3'	S. Deane
LACF14REPAF ( <i>EcoRI</i> )	5' -TATCGAATTCGACTTGGCTGCGCTG-3'	This study
LACF14REPAR ( <i>BamHI</i> )	5' -TTAAGGATCCAGTTCTGGCGGTTTCG-3'	This study
LACF14REPCF ( <i>EcoRI</i> )	5' -TAGCGAATTCATCAGCCCTGATCG-3'	This study
LACF14REPCR ( <i>BamHI</i> )	5' -TATAGGATCCGGATCGTGTCTTGCG-3'	This study
LACZPRI	5' -CGCCAGCTGGCGAAAGGGGG-3'	A. Smith
RTF14REPCR	5' -ATCTCGATACGGGTGTGTGG-3'	This study
RTF14REPCF	5' -AGAGCCGCTTGGTGTGATG-3'	This study
RTF14REPAR	5' -ACCAACGGGCATTGTGATC-3'	This study
RTF14REPAF	5' -CTATCTGGCCGAGAAGATC-3'	This study
RTF14PASBR	5' -AGAAGTCTCCAGCTTGGAG-3'	This study
RTF14PASAF	5' -AAGCGATCCTTGAGCACCTG-3'	This study
RTF14REPBJPASF	5' -TGACCAGTCGGCTATTGAGG-3'	This study
RTF14REPBR	5' -GATGACCGCCTGATAGTTGC-3'	This study
RTF14MOBAJREPBF	5' -TGACCGGAACGACGAATAC-3'	This study
RTF14MOBAR	5' -TGGCGTCGCTTGTGTTGGTTC-3'	This study
RTF14MOBBJMOBAF	5' -CTTTGAAAGCCGCATCGAG-3'	This study
RTF14MOBBR	5' -CTCGATGGCGGCTTTCAAAG-3'	This study
RTF14MOBBF	5' -TACCGTGCAAGGACTGGAAC-3'	This study
RTF14MOBCJMOBB	5' -TGATGACATCGGCTTTCCCG-3'	This study

<sup>b</sup> Restriction enzyme sites incorporated into primers are indicated in brackets, and underlined in the primer sequence.

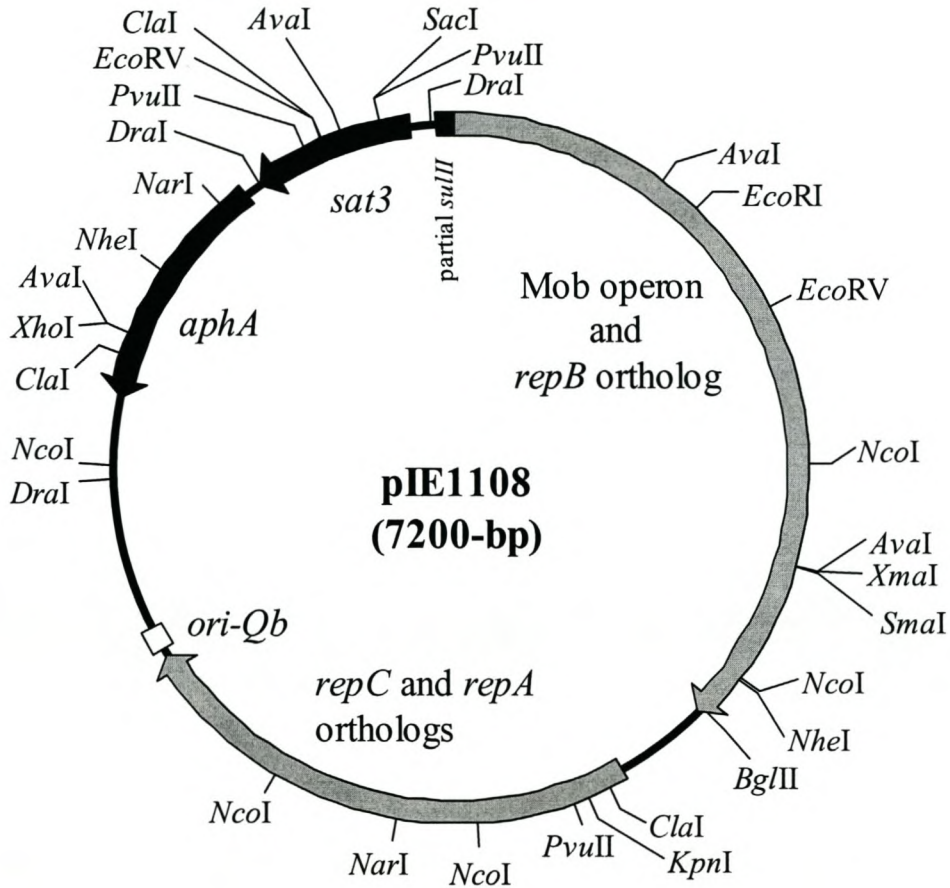


## Appendix D: Plasmid maps



**Fig A:** Circular map of the RK2-based broad host range cloning vector pGL10. Map drawn from information supplied by A. Toukdarian.

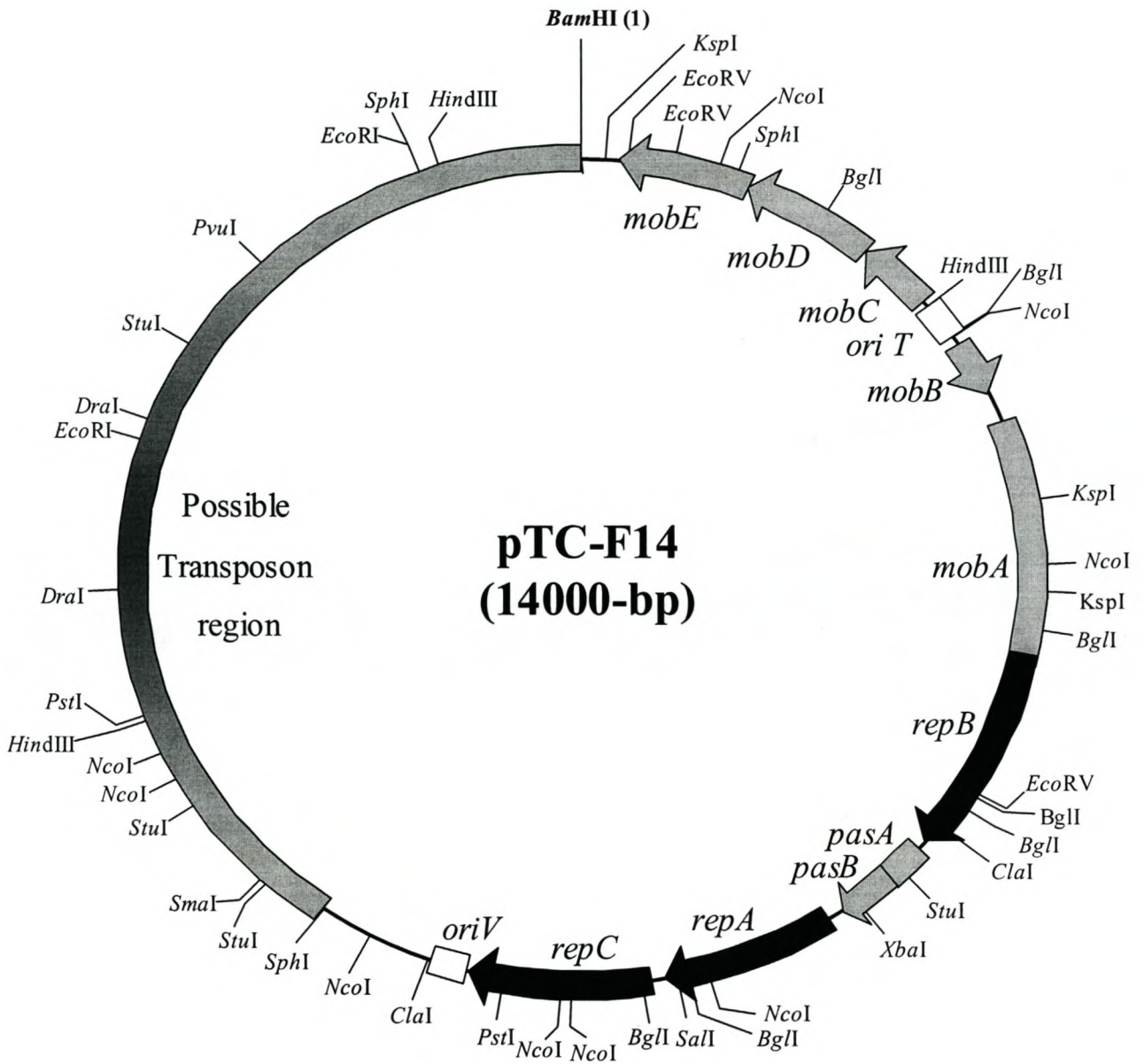
**Appendix D:** (cont.)



**Fig B:** Circular map of the pIE1107 plasmid derivative pIE1108 used in this study. Constructed from Entrez accession number Z74787 and from Tietze (1998). This plasmid has a novel kanamycin and neomycin resistance gene (*aphA*), a streptothricin-acetyltransferase-3 gene (*sat3*) providing resistance to streptothricin, and a truncated sulfonamide resistance gene (*suIII*).

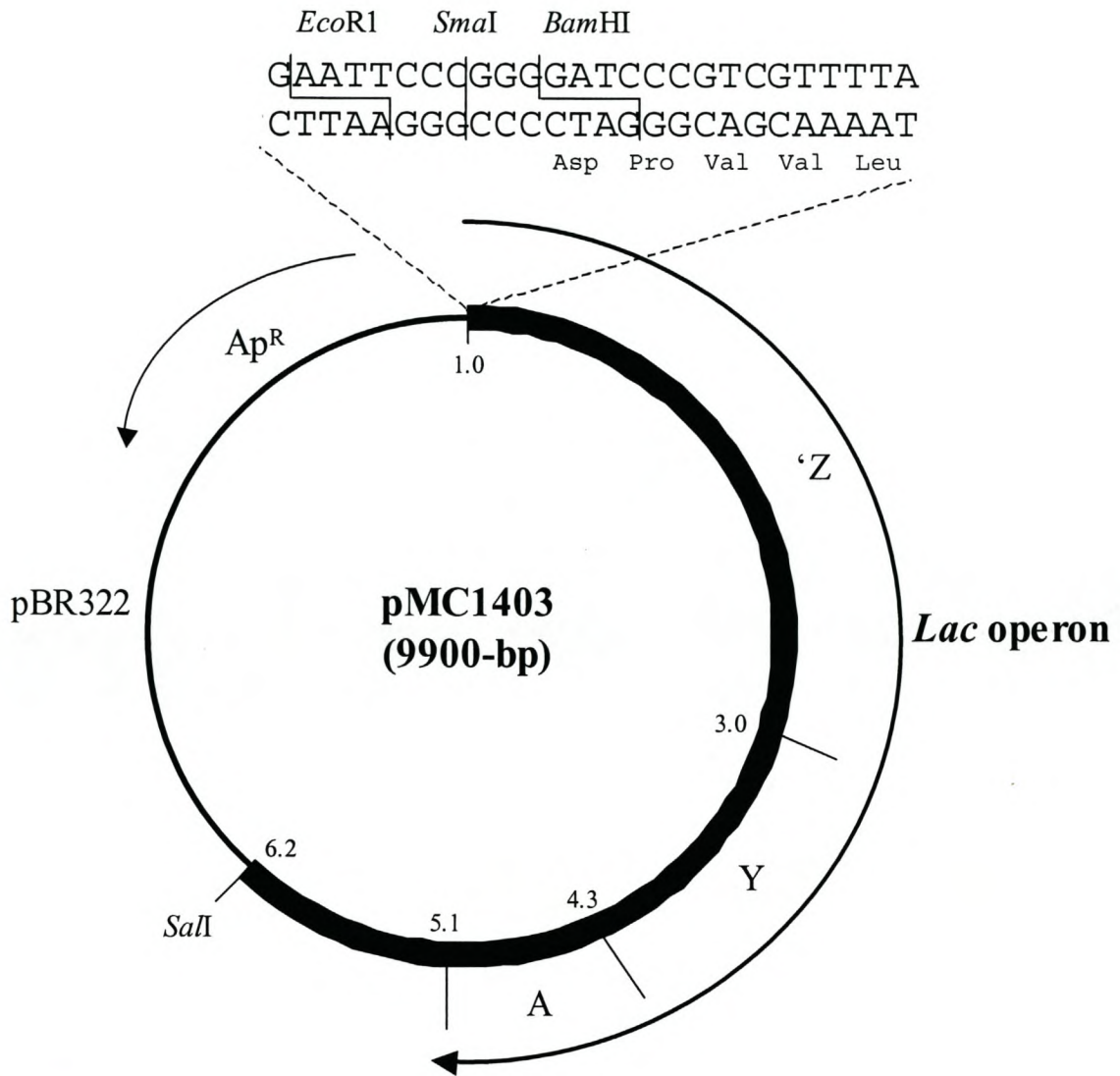


**Appendix D: (cont.)**



**Fig C:** Circular map of pTC-F14 compiled by M. Gardner, G. Goldschmidt, and L. van Zyl. With reference to Appendix E, note the *Bam*HI site is used as bp = 1.

**Appendix D:** (cont.)



**Fig D:**  $\beta$ -galactosidase hybrid protein gene fusion vector pMC1403. The three unique cloning sites for construction of promoter-reporter gene (*lacZ*) fusions are indicated. The vector contains the *lac* operon with the *lacZ* gene ('Z') for  $\beta$ -galactosidase, the *lacY* gene (Y) for lactose permease in *E. coli*, and the non-essential *lacA* gene (A) for transacetylase. The vector is missing the *lac* promoter, operator, and translation initiation site, as well as the first 7 $\frac{1}{3}$  non-essential amino acid codons. Reproduced from Casadaban *et al.* (1983).



**Appendix E: Annotation of pTC-F14 (1921-8389 using *Bam*HI as bp=1)**

Universal code

```

1921      CCATCCTGTCGTTCTCCTCTGTGCCGGTTTTCTTAAGCTTTGCATAGATTCTACTTCAATTG
                                     HindIII                                     -35 region
1981      ACTTTCGTAAAGAAAGCACATACAATACGTATAAGCTATATTTTGAACAGGATGGACTTG
          Possible promoter region  -10 region  oriT nic site
          6 bp → ← 6 bp
2041      TCATACACATGCTACGCATGTTCCCGCCAAGCCACGCCCAACCATTTCGGCGCTACGCTTC
                                               NcoI
2101      TCACGGTCCGGCTCATCCTGCAAAGGGCAAGCCCCTTGACCCCGGCCAGCGCGGCCAT
          RBS  mobB
2161      GGCTTCTTGCTGACCCAGACACACCCCGCCGCCGGAAGGAGACACGATGCCATTTACC
                                               M P F T

2221      GTGCAAGGACTGGAACCTCTGGATGCGGTTCGTCAATGTCCGGCTAACCGCATCGGAGAAA
          V Q G L E P L D A V V N V R L T A S E K

2281      GCGCGTCTACGGGAGGATGCGGACCTCGCCGGATTGAGTGTTCGAGCTGGTGCGCCGC
          A R L R E D A D L A G L S V S E L V R R

2341      CGCTACTTTGGCAGGCCGATTGTGGCTCATGCGGATGCGGTCTTGCTCAAAGAGCTACGC
          R Y F G R P I V A H A D A V L L K E L R

2401      CGCATCGGGGGATTGCTCAAGCATGTACACAATGAAAGCGGTGGAGCGTACAGCCAGCAA
          R I G G L L K H V H N E S G G A Y S Q Q

2461      ACTGCCCGTTCTGGTCACTTTGAAAGCCGCATCGAGGGGCTGAGTCATGATCGTTAA
          T A A V L V T L K A A I E G L S H D R *

2521      AAAGGTCAAGAGCAATCGAACAAAGGGCAAAGCGGCGAGTATACGCGATCTGACTAACTA

2581      CATCCGGGAGCCGCAGAACCAGGAATCCAAATGAGAAGGTACTTTACCGGAACGGACGGGG

2641      TTTTCATCAGCGACACTCATGCCGCCAGCGGGAGGAATGGTAGCCCTGGCAGCGGAAGC
          RBS  mobA
          M V A L A A E A

2701      GGTACGCAGCCGCAATCCGGTCAATCACTACATCCTGAGCTGGCGGGAAGGGGAGCAGCC
          V R S R N P V N H Y I L S W R E G E Q P

2761      TAGCCCGGAGCAGGTGGAAGAAGCGGTAAGCATCTTTCTGGATGAACTTGGCTTGCAGGA
          S P E Q V E E A V S I F L D E L G L Q E

2821      ACACCAGGTTATCTACGCCCTGCACAAGGACACGGATAACCTGCATCTGCATATCGCCGT
          H Q V I Y A L H K D T D N L H L H I A V

2881      CAATCGCGTACACCCAGAAACGCTCAAGTGTGTGGAGATCAACAAGGGCTTTGACCTTGA
          N R V H P E T L K C V E I N K G F D L E

2941      ATCCGCGCATCGAGCTATTGCCCGGATTGAACATGCACAGGGATGGCAACGGGAGCAGAA
          S A H R A I A R I E H A Q G W Q R E Q N

3001      CGGGCGCTATGAAGTGTTAGAGAATGGCGAGCTAGGACGAGAACACCTGGAACCAAACAA
          G R Y E V L E N G E L G R E H L E P N K
                                               SacII
3061      GCGACGCCAACCGGAGCAGCGCAAGCGGGACAAGGAGAACCGCACCGGGGAGAAATCCGC
          R R Q P E Q R K R D K E N R T G E K S A

3121      GGAACGTATCGCCATTGAGATCGGTGCGCCGATCATCAAGCAAGCGCAAAGCTGGGAACA
          E R I A I E I G A P I I K Q A Q S W E Q
    
```



3181 GTTGACCCGGAACTGGCAGCACAAAGGGATGCGCTACGAGCAGAAGGGCAGTGGCGCATT  
L H R E L A A Q G M R Y E Q K G S G A L

3241 GCTTTGGGTTCGGTGTAGGTAGCGGTCAAAGCCAGCAGTGCCGACCGGGAAGCGAGCCTCGG  
L W V G E V A V K A S S A D R E A S L G

3301 CAAACTGCAAAGCGGCTGGGCGCCTATGAACCCGCGCAAGCACCTTCGCCTGTGGCGCA  
K L Q K R L G A Y E P A Q A P S P V A Q

3361 ACGGAAGCCGGAACCGCTTCAACCCGACAGGCCGGAGTGGGAGGACTTCATGGCTGGACG  
R K P E P L Q P D R P E W E D F M A G R  
*NcoI*

3421 CAAAATGCACTACGCAGAAAAGAACGCGGCCAAGCTCTCCATGGACCAGCGGCAGGAACA  
K M H Y A E K N A A K L S M D Q R Q E Q

3481 GGAACGCAAGGCATTGCAAGCGCGGCAGCAGGAACAACGCAAGGTACTCCTGGGCGGACG  
E R K A L Q A R Q Q E Q R K V L L G G R

3541 ATGGAATGGCAAGGGAGAAGCGCTGAACGCGCTGCGCGGCGTGCTTGCCGAGAACAGGC  
W N G K G E A L N A L R G V L A A E Q A  
*SacII*

3601 CGCGGAGAAAGCCGCCCTGAAAGAACGTCACCAGCAGGAACGGCAACAGTGGCGGCAACA  
A E K A A L K E R H Q Q E R Q Q W R Q Q

3661 GTACCGCCCGTATCCCGACTTCGAGCAGTGGTTACGACGCGAGCACGGGGCAGAACAGGC  
Y R P Y P D F E Q W L R R E H G A E Q A

3721 GACTAGGTGGCGCTACCGGGAGGCGGAACCGCAACGCATCGAGGGCGACTCGACCGAAGC  
T R W R Y R E A E P Q R I E G D S T E A

3781 GCCGAAACCTCGGGACATTTCGCGCCTACCGGGCGGAGATCGTGGGACAGGAGGTCCGTTA  
P K P R D I R A Y R A E I V G Q E V R Y

3841 CACCCCAAGAGTGGCGCGGGCGGGCCGGGGGTGTGTCTTTTGTGGACAAGGGCAG  
T P K S G A G G G P G G V S F V D K G R

3901 GATCATCGAAATCCACGATTGGCGGAACCAGGACACCACTCTTGCGGGCGCTCCAGCTCTC  
I I E I H D W R N Q D T T L A A L Q L S

3961 GGCGCAGAAATGGGGCAAGTTCACCGTGACCGGGAACGACGAATACAAAGCGCTGTGCGT  
A Q K W G K F T V T G N D E Y K A L C V

4021 GAAGCTGGCGGTGGAGCACGGTTTCCAGATCACGAACCCGGAGCTTCAGGAGGTCATTTCG  
K L A V E H G F Q I T N P E L Q E V I R

4081 GCAGGAACGGCAACGGATGCGGCAGGAGAGGGGCACAGGCGATGAAATCGGAGCAGATCAA  
Q E R Q R M R Q E R A Q A **M** K S E Q I K  
RBS *repB*

4141 GCCGTTTCGAGCGATAACCGGAAGCGGTTGGCGCCGAGCGCTACCGGGTGACAAGCATCAA  
P F E R Y A E A V G A E R Y R V T S I K

4201 AATGCGGCCAGATGGCAGCAAGCAAACCTTCATTCTCGATAAGCGGGACGGTATCACGGC  
M R P D G S K Q T F I L D K R D G I T R

4261 CGGCTTCACGCCGGAAGAAATCGCCCAAAGGACGCCCCGAAATGCAGCGTTTACAGCGCCG  
G F T P E E I A Q R T P E M Q R L Q R R

4321 GGGCGAAAACCTGTATTACAGCGCCTCTCCGAAGGGAAACACCATATCCTGATCGACGA  
G E N L Y Y T P L S E G K H H I L I D D



4381 CATGGACCGTGAGAACTGGACCGGCTGATTTCGTGATGGCTATCAGCCCGCGTTCGTGCT  
M D R E K L D R L I R D G Y Q P A V V L

4441 GGAATCCAGCCCCGCAACTATCAGGCGGTATCACCATTCCGAAGCTGGGGACCCCTT  
E S S P G N Y Q A V I T I P K L G T P F

4501 CGACAAGGACGTGGGGAACCGCCTGAGCGATGCGCTCAACCGGAATACGGCGACCCAA  
D K D V G N R L S D A L N R E Y G D P K

4561 GCTGTCTGGTGCCATCCATCCGCACCGCGCTCCCGGCTACGAGAACCGCAAGCCCAAGCA  
L S G A I H P H R A P G Y E N R K P K H

4621 CCGGCGGGAAGATGGCAGCTATCCCGAAGTGCCTTTGCTCAAGGCCGAGCGGCGGAGTG  
R R E D G S Y P E V R L L K A E R R E C  
*EcoRV*

4681 CGGCAAGACGCTGGCGCTTTCCCGGAGATCGACGCCGGATATCAGCGGCAGGCCCGCCA  
G K T L A L S R E I D A G Y Q R Q A A E

4741 GAAGGCCCTGAAAACGCCCGTAGAGCAATTTTCGGAGCCAAACAGTACCCAGACCACCCC  
K A L K T P V E Q F S E P N S T Q T T P

4801 TGTCTCTGAAAAACGCCACAGAGGCTTATTGGCGGCATTACCGCGATGTTTCGCAAGCG  
V S E K T A T E A Y W R H Y R D V R K R

4861 CCAACGCGGAATGCTGGACCTGTCCCGCGTGGACGCCATGATTGCCGTGCGGATGCGCGT  
Q R G M L D L S R V D A M I A V R M R V  
*Clai*

4921 TACCGGCTTTGACCAGTCGGCTATTGAGGGCGCCATCTACCAGTGTGCCCCATCGATCCG  
T G F D Q S A I E G A I Y Q C A P S I R

4981 GGAGCAGCAGGAAAGCCGGGACTGGACCGACTACGCCCGCCGACGGCCCGCTATGCCTA  
E Q Q E S R D W T D Y A R R T A R Y A Y

5041 CAGCGCCGAGGTGACCGGCAGGCCCGGACCTGGGTAAGTACCGCCAGCAGTGGGAGAA  
S A A G D R Q A A D L G K Y R Q Q W E K

5101 GCTGGAGGGGCGGAGCGGCAGCAGGAGCAGGCCAAGGCGCGGGAGATTGAGCGCAACGG  
L E G R E R Q Q E Q A K A R E I E R N G

5161 GCCGAGTATGAGTCGTTAGTATTCCTTCACAGCGGCATGATATACTTGTATATCATTTTGA  
P S M S R \*      <sup>-35 region</sup> CTTCACAGCGGCATGATATACTTGTATATCATTTTGA <sup>-10 region</sup> ●  
Possible promoter region (●= possible transcription start)

5221 GAGGAGC<sup>RBS</sup>TA AACATGCTTGCCATCCGATTACCCACCGAAGTGGAAAATCGCCTTGAGGC  
M L A I R L P T E V E N R L E A      *pasA*      *StuI*

5281 CTTGGCGCAGGCCACAGGACGCACGAAAACCTTTTACGCCCGGAAGCGATCCTTGAGCA  
L A Q A T G R T K T F Y A R E A I L E H

5341 CCTGGACGACCTCGAAGATTTGTACCTTGCAGAGCAACGCCTGATCGACATTCGCGCAGG  
L D D L E D L Y L A E Q R L I D I R A G

5401 CCGAAGCCGTACCTACACGCTAGAGGAAGTGGAGCGCGATCTTGGCTTGGCGGATTGAGT  
R S R T Y T L E E V E R D L G L A D \*      <sup>RBS</sup> *pasB*  
L A W R I E

5461 TTGATGACAAGGCCAAGAAAGACTTGGCTGCGCTGGATAAGAGCGTTGCCAAACGCATCA  
F D D K A K K D L A A L D K S V A K R I  
*XbaI*

5521 CGGCTTTTCTGCGAGAGCGCGTCCGCCATCTAGACGACCCGCGCAGCATTGGCGAAGCCC  
T A F L R E R V A H L D D P R S I G E A



5581 TCAAAGGCTCCAAGCTGGGAGACTTCTGGAAGTACCGCGTAGGGGATTGGCGAATCATTG  
L K G S K L G D F W K Y R V G D W R I I

5641 CCAGCATTGAGGATGAAGCCTTGCATATTCTTGTCTGCGTATTGGTAACCGCCGAGAGG  
A S I E D E A L R I L V V R I G N R R E

5701 TATACCGAAAGTAAACAACCCCGGCATGAAGGGCCGGGTTTCCCGCGCAACCGGATGAA  
V Y R K \*

5761 AGGACGGCGCGAGCCGGAGTGTCTAGTCCCCGGCAGCGCCTAACCACGCCAGTTCTGGAAA  
RBS **repA**

5821 GGAGAAACAGCATGGCTTTGGACATTATGGCGGCATTCACCAACGAACCGCCAGAACTGGA  
M A L D I M A A F T N E P P E L D

5881 TTTTCATCTGGCCGGGATTCTGGCCGGGACCGTGGGCGCACTGGTTGCACCTGGCGCGAC  
F I W P G F L A G T V G A L V A P G A T

5941 CGGCAAGAGCTTCTGGGCCTTGAAGCGGCGATGGCCGTGCGGTGCCGGGAGCCTGGCCG  
G K S F W A L E A A M A V A C R E P G G

6001 AGACCTGTGGGGCTTGGCGCGAAGTACACCGGGCGCGTTGTCTATCTGGCCGGAGAAGA  
D L L G L A P K Y T G R V V Y L A G E D

6061 TCCAGAACCAGCCTTAATCCGGCGCGTTCATGCCATAGGCCAGCATCTCAACCAGTCGGC  
P E P A L I R R V H A I G Q H L N Q S A

6121 CCGCGAAGCCATTGCTGAGAATCTGACCCTTGAGCCGATCATGGGCAAGCGGCTGAACAT  
R E A I A E N L T L E P I M G K R L N I

6181 CATGGACGAGGCCACCTGCATCGCATCATCGAGGCCAGCTCCGGGGCACGGCTGATCGT  
M D E A H L H R I I E A S S G A R L I V  
NcoI

6241 GCTGGACACCCTGAGCCGTATTACAATCTGGACGAGAACAGCAACGGTGCATGGCGCA  
L D T L S R I H N L D E N S N G A M A H

6301 TCTGGTGGCAACCCTGGAGCATATTGCCGCCAGTACCGGCGCATCGGTGCTGTACCTGCA  
L V A T L E H I A A S T G A S V L Y L H  
NcoI

6361 TCATGTGAGCAAAGGCAGCGCCATGGCCGGGCGAGACGGATCAGCAGCAAGCGGCGCGGG  
H V S K G S A M A G Q T D Q Q Q A A R G

6421 GGCATCAGCCCTGATCGACAATGCCCGTTGGTGC GGCTACGTTGCCAAAATGGCAGAGCA  
A S A L I D N A R W C G Y V A K M A E H

6481 TGAGGCCGCGCGACTGAGCGACCGTAGCTTTGACCGTGCGCCATTGGCGACCGGCGCGG  
E A A R L S D R S F D R A P I G D R R G

6541 CTTCTTCGTGCGCTTCGGTGTGAGCAAACAGAACTACGACGCAACATCGCTCGAAAGCTG  
F F V R F G V S K Q N Y D A T S L E S W  
Sali

6601 GTATCAGCGAGTGGATGGCGGGGTGTTGATGCCCGTCGACCTGGTGGAAAGCAACGAGCAA  
Y Q R V D G G V L M P V D L V E A T S N

6661 TGGGAAGGTAAAAAACGATGGCAAAGGAAAGCGCGATGAGATCTGACCCTGTGCTATCGC  
G K V K N D G K G K R D E I \*

6721 ACCCCCTGCCGATCCCCTTTTTGTGTGGCTATACCCAGGGAGGCAAGCCATGAAGAAGGG  
RBS **repC**  
M K K G

6781 AAGCGCGATGACCTTCGACCTCACCCACGCAAGACACGATCCGGCTCACTGTCTGGCTCC  
S A M T F D L T H A R H D P A H C L A P



6841 TGGTCTTTTTTCGCAGCCTCAAGCGTGGCGAGCGAAAAAACTCAAGTTGGACGTGACCTA  
 G L F R S L K R G E R K K L K L D V T Y

6901 CACTTACGGTAAGGATCGTGTGAGTTTAGCGGTCCAGAGCCGCTTGGTGTGATGATAT  
 T Y G K D R V E F S G P E P L G V D D M

6961 GCGTGTCTTGCAAGGTCTCGTTGCCGTCGCCGCCACCTGTGGGTCGAAAGTCGGGAAT  
 R V L Q G L V A V A A T C G S E S R G I

7021 CATGCTGCGCCCTGAGCCAAAAAGTGAAGCGGGACATCAGTTGAGGCTGTTGCTTGATTT  
 M L R P E P K S E A G H Q L R L L L D L  
*NcoI*

7081 AGAATGGGACGCCAAGGAAAAAGATGCCATGGTCGCTAAAGGCAGCTTCCGGCAACTTGC  
 E W D A K E K D A M V A K G S F R Q L A

7141 AAGCGAGCTTGGATATGCTACAGACAGTGGCGGCGGTTTCGTGTCAATCGAGAAAGTAT  
 S E L G Y A T D S G G A F R V I R E S I

7201 CGAACGACTATGGAAGGTCTCCGTGATCGTTCAGCGCGAGGGGAAACGCAGGGATTCCG  
 E R L W K V S V I V Q R E G K R Q G F R

7261 CATCCTCTCCGAATATGCCAGCGACGAAGACACCGGGAAGCTATTTGTTGCCTTGAACCC  
 I L S E Y A S D E D T G K L F V A L N P


7321 CAGACTGGCTGAGGCAATCATCGGAGCACGTCCACACACCCGTATCGAGATGGCAGAGGT  
 R L A E A I I G A R P H T R I E M A E V  
*PstI*


7381 GCGAGCACTGCAGACTGACCCTGCCCGTCTCATTTCATCAGCGGTATGCGGTTGGATCGA  
 R A L Q T D P A R L I H Q R L C G W I D

7441 CCCC GGCAAGTCCGGGCGCGTTCGAGCTGGACACGCTTTGCGGTTATATCTGGCCCGATCA  
 P G K S G R V E L D T L C G Y I W P D Q

7501 GGCCAGCGGGTCAACCATGCGTATGCGCCGACAGACTGCTCGTAATGCGCTGCCCGAGCT  
 A S G S T M R M R R Q T A R N A L P E L

7561 TACCACTGTAGGCTGGACGGTGAACGAATACGCCAAGGGAAAAATGGGAAATCAGCAGACC  
 T T V G W T V N E Y A K G K W E I S R P

7621   
 GAAGGCCCTTGCCCTAACGTTGACGAGCCCCCTGCCCTAACGTTGACGAGCCCCCTGCCCTA  
 K A L A L T L T S P C P N V D E P L P \*

7681   
 ACGTTGACGAGCCCCCTGCCCTAACGTTGACGAGCCCCCTGCCCTAACGTTGACACCACAAA  
*Clal* 8 bp

7741 ACACGCTAAGCCAGCATTGACGGGGCTTTGCGGAAGATTTGAAAAATCGATCCAAGATAT  
8 bp

7801 TCTCCAAGAATATCCACTATGCGCGGCCTTCAGCCGCCGCTTGATGCGTTCGTAGAAGCC  
12 bp

7861 TTGCACTTCGGCCTTCGGCCTGCGCGGGCTACGCCCTTGACAGACACACCCCGCCGCCCGC  
12 bp 12 bp 12 bp

7921 CCGCCCGCCGCCCGCTTGTACGTTCTCCAAGGTGGATAACGTGACTTACGTTATCCACCA

7981 ACCATGCCGCCGATAGTGAGCACCTACTGGGTAGACGAGCGGCTGATGGTGTAAATAACC  
*NcoI*

8041 TTGACAAATAGATGTAAATTTACTATATTACATTTACAATGTAATTACAGGAGAGTGCCA

8101 TGGAAGCAGGACCGATCACACAACAAGACGTTGCCGCCGCCGCTGACCAGCTTGCGCGCC

```
8161      CCGCCAACACCAATGCCGCGAAAATCCGGGAGATTCTGGGCCGGGGAGTCTCGCCACCA
8221      TTCAACGCCATTTGCAGGCCCTTAGAGACGCTCAGAGGGCTCCAGAGCTACCAGAGGCCG
8281      TCCAGACGATTCCAACGCCTCCTGAGGCCGTTTCTGAGGCTTCCAGAGGCATTTGGGCGG
                                                SphI
8341      CGGCCTGGGCAATGGCGGAGCAACGCCACGCGGAAAGTCTGGCGCATGC
```

**Fig E:** Annotated sequence of pTC-F14 basic replicon. Nucleotide sequence deposited in the GenBank database under accession number AF325537.



## Literature cited

- Abeles, A. L., K. M. Snyder, and D. K. Chattoraj. (1984).** P1 plasmid replication: replicon structure. *J Mol Biol* **173**:307-324.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-3402.
- Atlung, T. and F. G. Hansen. (2002).** Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in *Escherichia coli*. *J Bacteriol* **184**:1843-1850.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. (1993).** *Current protocols in molecular biology*. Wiley Interscience, New York.
- Bagdasarian, M. M., P. Scholz, J. Frey, and M. Bagdasarian. (1986).** Regulation of the *rep* operon expression in the broad host range plasmid RSF1010, p. 209-223. In R. P. Novick and S. Levy (eds.), *Evolution and environmental spread of antibiotic resistance genes*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Bahloul, A., F. Boubrik, and J. Rouviere-Yaniv. (2001).** Roles of *Escherichia coli* histone-like protein HU in DNA replication: HU-beta suppresses the thermosensitivity of *dnaA46ts*. *Biochimie* **83**:219-229.
- Bairoch, A., P. Bucher, and K. Hofmann. (1997).** The PROSITE database, its status in 1997. *Nucleic Acids Res* **25**:217-221.
- Barth, P. T. and N. J. Grinter. (1974).** Comparison of the deoxyribonucleic acid molecular weights and homologies of plasmids conferring linked resistance to streptomycin and sulfonamides. *J Bacteriol* **120**:618-630.
- Beese, L. S. and T. A. Steitz. (1991).** Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J* **10**:25-33.
- Benkovic, S. J., A. M. Valentine, and F. Salinas. (2001).** Replisome-mediated DNA replication. *Annu Rev Biochem* **70**:181-208.
- Biswas, E. E. and S. B. Biswas. (1999[a]).** Mechanism of DnaB helicase of *Escherichia coli*: structural domains involved in ATP hydrolysis, DNA binding, and oligomerization. *Biochemistry* **38**:10919-10928.
- Biswas, E. E. and S. B. Biswas. (1999[b]).** Mechanism of DNA binding by the DnaB helicase of *Escherichia coli*: analysis of the roles of domain  $\gamma$  in DNA binding. *Biochemistry* **38**:10929-10939.
- Biswas, E. E., P.-H. Chen, and S. B. Biswas. (2002).** Modulation of enzymatic activities of *Escherichia coli* DnaB helicase by single-stranded DNA-binding proteins. *Nucleic Acids Res* **30**:2809-2816.



- Blomberg, P., K. Nördstrom, and E. G. Wagner. (1992).** Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J* **11**:2675-2683.
- Brasch, M. A. and R. J. Meyer. (1988).** Integrative suppression of *dnaA46* by broad host-range plasmid R1162. *Mol Gen Genet* **215**:139-145.
- Breed, A. W., C. J. N. Dempers, G. E. Searby, M. N. Gardner, D. E. Rawlings, and G. S. Hansford. (1999).** The effect of temperature on the ferrous iron oxidation kinetics of *Leptospirillum ferrooxidans* in continuous culture. *Biotech Bioeng* **65**:44-53.
- Brennan, R. G. and B. W. Matthews. (1989).** The helix-turn-helix DNA binding motif. *J Biol Chem* **264**:1903-1906.
- Brierley, C. L. (1982).** Microbiological mining. *Sci Am* **247**:42-51.
- Casadaban, M. J., A. Martinez-Arias, S. K. Shapira, and J. Chou. (1983).**  $\beta$ -galactosidase gene fusions for analyzing gene expression in *Escherichia coli* and yeast. *Methods Enzymol* **100**:293-308.
- Caspi, R., D. R. Helinski, M. Pacek, and I. Konieczny. (2000).** Interactions of DnaA proteins from distantly related bacteria with the replication origin of the broad host range plasmid RK2. *J Biol Chem* **275**:18454-18461.
- Cereghino, J. L. and D. R. Helinski. (1993).** Essentiality of the three carboxyl-terminal amino acids of the plasmid RK2 replication initiation protein TrfA for DNA binding and replication activity in gram-negative bacteria. *J Biol Chem* **268**:24926-24932.
- Chang, A. C. and S. N. Cohen. (1978).** Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**:1141-1156.
- Chang, T.-L., M. G. Kramer, R. A. Ansari, and S. A. Khan. (2000).** Role of individual monomers of a dimeric initiator protein in the initiation and termination of plasmid rolling circle replication. *J Biol Chem* **275**:13529-13534.
- Chang, T.-L., A. Naqvi, S. P. Anand, M. G. Kramer, R. Munshi, and S. A. Khan. (2002).** Biochemical characterization of the *Staphylococcus aureus* PcrA helicase and its role in plasmid rolling circle replication. *J Biol Chem* **277**:45880-45886.
- Chattoraj, D. K. (2000).** Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol Microbiol* **37**:467-476.
- Clennel, A.-M., B. Johnson, and D.E. Rawlings. (1995).** Structure and function of Tn5467, a Tn21-like transposon located on *Thiobacillus ferrooxidans* plasmid pTF-FC2. *Appl Environ Microbiol* **61**:4223-4229.
- Combet, C., C. Blanchet, C. Geourjon, and G. Deleage. (2000).** NPS@: network protein sequence analysis. *Trends Biochem Sci* **25**:147-150.
- Couturier, M., F. Bex, P. L. Bergquist, and W. K. Maas. (1988).** Identification and classification of bacterial plasmids. *Microbiol Rev* **52**:375-395.



- Cozzarelli, N. R. (1980).** DNA gyrase and the supercoiling of DNA. *Science* **207**:953-960.
- Da Silva-Tatley, F. M. and L. M. Steyn. (1993).** Characterization of a replicon of the moderately promiscuous plasmid, pGSH5000, with features of both the mini-replicon of pCU1 and the ori-2 of F. *Mol Microbiol* **7**:805-823.
- Datta, N. and R. W. Hedges. (1971).** Compatibility groups among fi - R factors. *Nature* **234**:222-223.
- Datta, H. J., G. S. Khatri, and D. Bastia. (1999).** Mechanism of recruitment of DnaB helicase to the replication origin of the plasmid pSC101. *Proc Natl Acad Sci USA* **96**:73-78.
- De Graaff, J., J. H. Crosa, F. Heffron, and S. Falkow. (1978).** Replication of the nonconjugative plasmid RSF1010 in *Escherichia coli* K-12. *J Bacteriol* **134**:1117-1122.
- del Solar, G., J. C. Alonso, M. Espinosa, and R. Díaz-Orejas. (1996).** Broad-host-range plasmid replication: an open question. *Mol Microbiol* **21**:661-666.
- del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas. (1998).** Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**:434-464.
- del Solar, G. and M. Espinosa. (2000).** Plasmid copy number control: an ever-growing story. *Mol Microbiol* **37**:492-500.
- del Solar, G., A. M. Hernández-Arriaga, F. X. Gomis-Rüth, M. Coll, and M. Espinosa. (2002).** A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. *J Bacteriol* **184**:4943-4951.
- Dellis, S., J. Feng, and M. Filutowicz. (1996).** Replication of plasmid R6K  $\gamma$  origin *in vivo* and *in vitro*: dependence on IHF binding to the *ihf1* site. *J Mol Biol* **257**:550-560.
- Derbyshire, K. M. and N. S. Willetts. (1987).** Mobilization of the non-conjugative plasmid RSF1010: a genetic analysis of its origin of transfer. *Mol Gen Genet* **206**:154-160.
- Derbyshire, K. M., G. Hatfull, and N. Willetts. (1987).** Mobilization of the non-conjugative plasmid RSF1010: a genetic and DNA sequence analysis of the mobilization region. *Mol Gen Genet* **206**:161-168.
- Dew, D. W. and Miller, D. M. (1997).** The BioNic process, bioleaching of mineral sulphide concentrates for the recovery of nickel, p. M7.1.1-M7.1.9. In *International Biohydrometallurgy Symposium, IBS97*. Australian Mineral Foundation.
- Dew, D. W., E. N. Lawson, and J. L. Broadhurst. (1997).** The Biox® process for biooxidation of gold-bearing ores or concentrates, p. 45-80. In D. E. Rawlings (ed.), *Biomining: Theory, Microbes and Industrial processes*. Springer-Verlag, New York.
- Diaz, R. and W. L. Staudenbauer. (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.



- Dodd, I. B. and J. B. Egan. (1990).** Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* **18**:5019-5026.
- Doran, K. S., I. Konieczny, and D. R. Helinski. (1998).** Replication origin of the broad host range plasmid RK2. Positioning of various motifs is critical for initiation of replication. *J Biol Chem* **273**:8447-8453.
- Doran, K. S., D. R. Helinski, and I. Konieczny. (1999[a]).** A critical DnaA box directs the cooperative binding of the *Escherichia coli* DnaA protein to the plasmid RK2 replication origin. *J Biol Chem* **274**:17918-17923.
- Doran, K. S., D. R. Helinski, and I. Konieczny. (1999[b]).** Host-dependent requirement for specific DnaA boxes for plasmid RK2 replication. *Mol Microbiol* **33**:490-498.
- Dorrington, R. A. and D. E. Rawlings. (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 D. E. Rawlings. (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**:5697-5705.
- Dorrington, R. A., S. Bardien, and D. E. Rawlings. (1991).** The broad-host-range plasmid pTF-FC2 requires a primase-like protein for autonomous replication in *Escherichia coli*. *Gene* **108**:7-14.
- Durland, R. H. and D. R. Helinski. (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.
- Ellison, V. and B. Stillman. (2001).** Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell* **106**:655-660.
- Espinosa, M., G. del Solar, F. Rojo, and J. C. Alonso. (1995).** Plasmid rolling circle replication and its control. *FEMS Microbiol Lett* **130**:111-120.
- Espinosa, M., S. Cohen, M. Couturier, G. del Solar, R. Díaz-Orejas, R. Giraldo, L. Janniere, C. Miller, M. Osborn, and C. M. Thomas. (2000).** Plasmid replication and copy number control, p. 1-47. In C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood academic publishers, Amsterdam.
- Fernández-Tresguerres, M. E., M. Martín, D. García de Viedma, R. Giraldo, and R. Díaz-Orejas. (1995).** Host growth temperature and a conservative amino acid substitution in the replication protein of pPS10 influence plasmid host range. *J Bacteriol* **177**:4377-4384.
- Filutowicz, M. and S. A. Rakowski. (1998).** Regulatory implications of protein assemblies at the  $\gamma$  origin of plasmid. *Gene* **223**:195-204.
- Franklin, F. C., M. Bagdasarian, M. M. Bagdasarian, and K. N. Timmis. (1981).** Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. *Proc Natl Acad Sci USA* **78**:7458-7462.



- Frey, J. and M. Bagdasarian. (1989).** The Molecular Biology of IncQ Plasmids, p. 79-93. In C. M. Thomas (ed.), *Promiscuous Plasmids of Gram-Negative Bacteria*. Academic Press Limited, London.
- Frey, J., M. M. Bagdasarian, and M. Bagdasarian. (1992).** Replication and copy number control of the broad-host-range plasmid RSF1010. *Gene* **113**:101-106.
- Frick, D. N. and C. C. Richardson. (2001).** DNA primases. *Annu Rev Biochem* **70**:39-80.
- García de Viedma, D., R. Giraldo, M. J. Ruiz-Echevarría, R. Lurz, and R. Díaz-Orejas. (1995).** Transcription of *repA*, the gene of the initiation protein of the Pseudomonas plasmid pPS10, is autoregulated by interactions of the RepA protein at a symmetrical operator. *J Mol Biol* **247**:211-223.
- García de Viedma, D., R. Giraldo, G. Rivas, E. Fernández-Tresguerres, and R. Díaz-Orejas. (1996).** A leucine zipper motif determines different functions in a DNA replication protein. *EMBO J* **15**:925-934.
- Gardner, M.N., S.M. Deane and D.E. Rawlings (2001).** Isolation of a new broad-host-range IncQ-like plasmid, pTC-F14, from the acidophilic bacterium *Acidithiobacillus caldus* and analysis of the plasmid replicon. *J Bacteriol* **183**:3303-3309.
- Giraldo, R., J. M. Andreu, and R. Díaz-Orejas. (1998).** Protein domains and conformational changes in the activation of RepA, a DNA replication initiator. *EMBO J* **17**:4511-4526.
- Giraldo, R. and R. Díaz-Orejas. (2001).** Similarities between the DNA replication initiators of Gram-negative bacteria plasmids (RepA) and eukaryotes (Orc4p)/archaea (Cdc6p). *Proc Natl Acad Sci U S A* **98**:4938-4943.
- Grimwade, J. E., V. T. Ryan, and A. C. Leonard. (2000).** IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. *Mol Microbiol* **35**:835-844.
- Gruss, A. and S. D. Ehrlich. (1989).** The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol Rev* **53**:231-241.
- Guerry, P., J. Van Embden, and S. Falkow. (1974).** Molecular nature of two nonconjugative plasmids carrying drug resistance genes. *J Bacteriol* **117**:619-630.
- Hall, M. C. and S. W. Matson. (1999).** Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* **34**:867-877.
- Hallberg, K. B. and E. B. Lindström. (1994).** Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. *Microbiology* **140**:3451-3456.
- Haring, V., P. Scholz, E. Scherzinger, J. Frey, K. Derbyshire, G. Hatfull, N. S. Willetts, and M. Bagdasarian. (1985).** Protein RepC is involved in copy number control of the broad host range plasmid RSF1010. *Proc Natl Acad Sci USA* **82**:6090-6094.
- Haring, V. and E. Scherzinger. (1988).** The replication proteins of IncQ plasmid RSF1010 In C. Thomas and F. C. H. Franklin (eds.), *Molecular biology of broad host range plasmids*. Academic Press, London.



- Harley, C. B. and R. P. Reynolds. (1987).** Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* **15**:2343-2361.
- Henderson, D. and R. J. Meyer. (1996).** The primase of broad-host-range plasmid R1162 is active in conjugal transfer. *J Bacteriol* **178**:6888-6894.
- Henderson, D. and R. J. Meyer. (1999).** The MobA-linked primase is the only replication protein of R1162 required for conjugal mobilization. *J Bacteriol* **181**:2973-2978.
- Hiasa, H., H. Sakai, T. Komano, and G. N. Godson. (1990).** Structural features of the priming signal recognized by primase: mutational analysis of the phage G4 origin of complementary DNA strand synthesis. *Nucleic Acids Res* **18**:4825-4831.
- Higashi, A., H. Sakai, Y. Honda, K. Tanaka, D.-M. Miao, T. Nakamura, Y. Taguchi, T. Komano, and M. Bagdasarian. (1994).** Functional features of *oriV* of the broad host range plasmid RSF1010 in *Pseudomonas aeruginosa*. *Plasmid* **31**:196-200.
- Hockings, S. C. and A. Maxwell (2002).** Identification of four GyrA residues involved in the DNA breakage – reunion reaction of DNA gyrase. *J Mol Biol* **318**:351-359.
- Hoekema, A., P. R. Hirsch, P. J. J. Hooykaas, and R. A. Schilperoort. (1983).** A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**:179-180.
- Honda, Y., H. Sakai, and T. Komano. (1988).** Two single-strand DNA initiation signals located in the *oriV* region of plasmid RSF1010. *Gene* **68**:221-228.
- Honda, Y., H. Sakai, T. Komano, and M. Bagdasarian. (1989).** RepB' is required in *trans* for the two single-stranded DNA initiation signals in *oriV* of plasmid RSF1010. *Gene* **80**:155-159.
- Honda, Y., H. Sakai, H. Hiasa, K. Tanaka, T. Komano, and M. Bagdasarian. (1991).** Functional division and reconstruction of a plasmid replication origin: molecular dissection of the *oriV* of the broad-host-range plasmid RSF1010. *Proc Natl Acad Sci USA* **88**:179-183.
- Honda, Y., T. Nakamura, K. Tanaka, A. Higashi, H. Sakai, T. Komano, and M. Bagdasarian. (1992).** DnaG-dependent priming signals can substitute for the two essential DNA initiation signals in *oriV* of the broad host-range plasmid RSF1010. *Nucleic Acids Res* **20**:1733-1737.
- Honda, Y., T. Akioka, S. Takebe, K. Tanaka, D. Miao, A. Higashi, T. Nakamura, Y. Taguchi, H. Sakai, T. Komano, and M. Bagdasarian. (1993).** Mutational analysis of the specific priming signal essential for DNA replication of the broad host-range plasmid RSF1010. *FEBS Letters* **324**:67-70.
- Horiuchi, T. and M. Hidaka. (1988).** Core sequence of two separable terminus sites is a 20 bp inverted repeat. *Cell* **54**:515-523.
- Ilyina, T. V., A. E. Gorbalenya, and E. V. Koonin. (1992).** Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J Mol Evol* **34**:351-357.



- Ingmer, H., E. L. Fong, and S. N. Cohen. (1995).** Monomer-dimer equilibrium of the pSC101 RepA protein. *J Mol Biol* **250**:309-314.
- Iwasaki, K., H. Uchiyama, O. Yagi, T. Kurabayashi, K. Ishizuka, and Y. Takamura. (1994).** Transformation of *Pseudomonas putida* by electroporation. *Biosci Biotech Biochem* **58**:851-854.
- Jones, C. S., D. J. Osborne, and J. Stanley. (1993).** Molecular comparison of the IncX plasmids allows division into IncX1 and IncX2 subgroups. *J Gen Microbiol* **139 ( Pt 4)**:735-741.
- Katayama, T., T. Kubota, K. Kurokawa, E. Crooke, and K. Sekimizu. (1998).** The  $\beta$ -subunit sliding clamp of DNA polymerase III, the *E. coli* replicase, renders DnaA protein inactive for initiating chromosomal replication. *Cell* **94**:61-71.
- Keck, J. L., D. D. Roche, A. S. Lynch, and J. M. Berger. (2000).** Structure of the RNA polymerase domain of *E. coli* primase. *Science* **287**:2482-2486.
- Kelley, W. L., I. Patel, and D. Bastia. (1992).** Structural and functional analysis of a replication enhancer: separation of the enhancer activity from origin function by mutational dissection of the replication origin  $\gamma$  of plasmid R6K. *Proc Natl Acad Sci USA* **89**:5078-5082.
- Khatri, G. S., T. MacAllister, P. R. Sista, and D. Bastia. (1989).** The replication terminator protein of *E. coli* is a DNA sequence-specific contra-helicase. *Cell* **59**:667-674.
- Kim, K. and R. J. Meyer. (1986).** Copy-number of broad host-range plasmid R1162 is regulated by a small RNA. *Nucleic Acids Res* **14**:8027-8046.
- Kim, M. H., J. C. Hines, and D. S. Ray. (1981).** Viable deletions of the M13 complementary strand origin. *Proc Natl Acad Sci USA* **78**:6784-6788.
- Kim, S.-Y., S. Sharma, J. R. Hoskins, and S. Wickner. (2002).** Interaction of the DnaK and DnaJ chaperone system with a native substrate, P1 RepA. *J Biol Chem* **277**:44778-44783.
- Kim, Y.-J. and R. J. Meyer. (1991).** An essential iteron-binding protein required for plasmid R1162 replication induces localized melting within the origin at a specific site in AT-rich DNA. *J Bacteriol* **173**:5539-5545.
- Klenow, H. and I. Henningsen. (1970).** Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc Natl Acad Sci USA* **65**:168-175.
- Kobayashi, N. and M. J. Bailey. (1994).** Plasmids isolated from the sugar beet phyllosphere show little or no homology to molecular probes currently available for plasmid typing. *Microbiology* **140** :289-296.
- Konieczny, I. and D. R. Helinski. (1997[a]).** Helicase delivery and activation by DnaA and TrfA proteins during the initiation of replication of the broad host range plasmid RK2. *J Biol Chem* **272**:33312-33318.



- Konieczny, I. and D. R. Helinski. (1997[b]).** The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone. *Proc Natl Acad Sci USA* **94**:14378-14382.
- Konieczny, I., K. S. Doran, D. R. Helinski, and A. Blasina. (1997).** Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *J Biol Chem* **272**:20173-20178.
- Konieczny, I. and K. Liberek. (2002).** Cooperative action of *Escherichia coli* ClpB protein and DnaK chaperone in the activation of a replication initiation protein. *J Biol Chem* **277**:18483-18488.
- Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer. (1994).** Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* **58**:401-465.
- Kramer, M. G., S. A. Khan, and M. Espinosa. (1997).** Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *EMBO J* **16**:5784-5795.
- Kramer, M. G., S. A. Khan, and M. Espinosa. (1998).** Lagging-strand replication from the *ssoA* origin of plasmid pMV158 in *Streptococcus pneumoniae*: *in vivo* and *in vitro* influences of mutations in two conserved *ssoA* regions. *J Bacteriol* **180**:83-89.
- Krüger, R. and M. Filutowicz. (2000).** Dimers of  $\pi$  protein bind the A+T-rich region of the R6K  $\gamma$  origin near the leading-strand synthesis start sites: regulatory implications. *J Bacteriol* **182**:2461-2467.
- Kumar, C. C. and R. P. Novick. (1985).** Plasmid pT181 replication is regulated by two countertranscripts. *Proc Natl Acad Sci USA* **82**:638-642.
- L'Abée-Lund, T. M. and H. Sørum (2002).** A global non-conjugative Tet C plasmid, pRAS3, from *Aeromonas salmonicida*. *Plasmid* **47**:172-181
- Laing, E. and I. S. Pretorius. (1992).** Synthesis and secretion of an *Erwinia chrysanthemi* pectate lyase in *Saccharomyces cerevisiae* regulated by different combinations of bacterial and yeast promoter and signal sequences. *Gene* **121**:35-45.
- Learn, B. A., S.-J. Um, L. Huang, and R. McMacken. (1997).** Cryptic single-stranded-DNA binding activities of the phage  $\lambda$  P and *Escherichia coli* DnaC replication initiation proteins facilitate the transfer of *E. coli* DnaB helicase onto DNA. *Proc Natl Acad Sci USA* **94**:1154-1159.
- Lehman, I. R. (1974).** DNA ligase: structure, mechanism, and function. *Science* **186**:790-797.
- Levchenko, I., D. York, and M. Filutowicz. (1994).** The dimerization domain of R6K plasmid replication initiator protein  $\pi$  revealed by analysis of a truncated protein. *Gene* **145**:65-68.



- Lilley, A. K., M. J. Bailey, M. J. Day, and J. C. Fry. (1996).** Diversity of mercury resistant plasmids obtained by exogenous isolation from the bacteria of sugar beet in three successive years. *FEMS Microbiol Ecol* **20**:211-227.
- Lilley, A., P. Young, and M. Bailey. (2000).** Bacterial population genetics: do plasmids maintain bacterial diversity and adaptation, p. 287-300. In C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood academic publishers, Amsterdam.
- Lin, L.-S. and R. J. Meyer. (1987).** DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Res* **15**:8319-8331.
- Lin, L.-S., Y.-J. Kim, and R. J. Meyer. (1987).** The 20 bp, directly repeated DNA sequence of broad host range plasmid R1162 exerts incompatibility *in vivo* and inhibits R1162 DNA replication *in vitro*. *Mol Gen Genet* **208**:390-397.
- Lohman, T. M. and K. P. Bjornson. (1996).** Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* **65**:169-214.
- Maeser, S., P. Scholz, S. Otto, and E. Scherzinger. (1990).** Gene F of plasmid RSF1010 codes for a low-molecular-weight repressor protein that autoregulates expression of the *repAC* operon. *Nucleic Acids Res* **18**:6215-6222.
- Maestro, B., J. M. Sanz, M. Faelen, M. Couturier, R. Díaz-Orejas, and E. Fernández-Tresguerres. (2002).** Modulation of pPS10 host range by DnaA. *Mol Microbiol* **46**:223-234.
- Maestro, B., J. M. Sanz, R. Díaz-Orejas, and E. Fernández-Tresguerres. (2003).** Modulation of pPS10 host range by plasmid-encoded RepA initiator protein. *J Bacteriol* **185**:1367-1375.
- Manen, D., L. C. Upegui-Gonzalez, and L. Caro. (1992).** Monomers and dimers of the RepA protein in plasmid pSC101 replication: domains in RepA. *Proc Natl Acad Sci USA* **89**:8923-8927.
- Masai, H. and K. Arai. (1989).** Leading strand synthesis of R1 plasmid replication *in vitro* is primed by primase alone at a specific site downstream of *oriR*. *J Biol Chem* **264**:8082-8090.
- Mathews, C. K. and K. E. van Holde. (1990).** *Biochemistry*. The Benjamin/Cummings Publishing Company, Inc., New York.
- Matson, S. W. and K. A. Kaiser-Rogers. (1990).** DNA helicases. *Annu Rev Biochem* **59**:289-329.
- McCarty, J. S., A. Buchberger, J. Reinstein, and B. Bukau. (1995).** The role of ATP in the functional cycle of the DnaK chaperone system. *J Mol Biol* **249**:126-137.
- Messer, W., F. Blaesing, D. Jakimowicz, M. Krause, J. Majka, J. Nardmann, S. Schaper, H. Seitz, C. Speck, C. Weigel, G. Wegrzyn, M. Welzeck, and J. Zakrzewska-Czerwinska. (2001).** Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie* **83**:5-12.



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

**Meynell, E., G. G. Meynell, and N. Datta. (1968).** Phylogenetic relationships of drug-resistance factors and other transmissible bacterial plasmids. *Bacteriol Rev* **32**:55-83.

**Miao, D.-M., Y. Honda, K. Tanaka, A. Higashi, T. Nakamura, Y. Taguchi, H. Sakai, T. Komano, and M. Bagdasarian. (1993).** A base-paired hairpin structure essential for the functional priming signal for DNA replication of the broad host range plasmid RSF1010. *Nucleic Acids Res* **21**:4900-4903.

**Miao, D.-M., H. Sakai, K. Tanaka, Y. Honda, T. Komano, and M. Bagdasarian. (1995[a]).** Functional distinction among structural subsections in the specific priming signal for DNA replication of the broad host-range plasmid RSF1010. *Biosci Biotech Biochem* **59**:920-921.

**Miao, D.-M., H. Sakai, S. Okamoto, K. Tanaka, M. Okuda, Y. Honda, T. Komano, and M. Bagdasarian. (1995[b]).** The interaction of RepC initiator with iterons in the replication of the broad host-range plasmid RSF1010. *Nucleic Acids Res* **23**:3295-3300.

**Miller, J. H. (1972).** *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

**Miller, P. C. (1997).** The design and operating practice of bacterial oxidation plant using moderate thermophiles (The BacTech process), p. 81-102. In D. E. Rawlings (ed.), *Biomining: Theory, Microbes and Industrial processes*. Springer-Verlag, New York.

**Moreira, D. (2000).** Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol Microbiol* **35**:1-5.

**Morin, D. (1995).** Bacterial leaching of refractory gold sulfide ores, p. 25-62. In C. C. Gaylarde and H. A. Videla (eds.), *Bioextraction and biodeterioration of metals*. Cambridge University Press, Cambridge.

**Mukhopadhyay, G., K. M. Carr, J. M. Kaguni, and D. K. Chattoraj. (1993).** Open-complex formation by the host initiator, DnaA, at the origin of P1 plasmid replication. *EMBO J* **12**:4547-4554.

**Mukhopadhyay, S. and D. K. Chattoraj. (2000).** Replication-induced transcription of an autorepressed gene: the replication initiator gene of plasmid P1. *Proc Natl Acad Sci USA* **97**:7142-7147.

**Mulugu, S., A. Potnis, Shamsuzzaman, J. Taylor, K. Alexander, and D. Bastia. (2001).** Mechanism of termination of DNA replication of *Escherichia coli* involves helicase-contrahelicase interaction. *Proc Natl Acad Sci USA* **98**:9569-9574.

**Newman, G. and E. Crooke. (2000).** DnaA, the initiator of *Escherichia coli* chromosomal replication, is located at the cell membrane. *J Bacteriol* **182**:2604-2610.



- Niedenzu, T., D. Röleke, G. Bains, E. Scherzinger, and W. Saenger. (2001).** Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010. *J Mol Biol* **306**:479-487.
- Nieto, C., R. Giraldo, E. Fernández-Tresguerres, and R. Diaz. (1992).** Genetic and functional analysis of the basic replicon of pPS10, a plasmid specific for *Pseudomonas* isolated from *Pseudomonas syringae* patovar *savastanoi*. *J Mol Biol* **223**:415-426.
- Noble, C. G. and A. Maxwell. (2002).** The role of GyrB in the DNA cleavage-religation reaction of DNA gyrase: a proposed two metal-ion mechanism. *J Mol Biol* **318**:361-371.
- Nordström, K. (1990).** Control of plasmid replication - how do DNA iterons set the replication frequency? *Cell* **63**:1121-1124.
- Novick, R. P. (1987).** Plasmid incompatibility. *Microbiol Rev* **51**:381-395.
- Novick, R. P., S. Iordanescu, S. J. Projan, J. Kornblum, and I. Edelman. (1989).** pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **59**:395-404.
- Osborn, M., S. Bron, N. Firth, S. Holsappel, A. Huddleston, R. Kiewiet, W. Meijer, J. Seegers, R. Skurray, P. Terpstra, C. M. Thomas, P. Thorsted, E. Tietze, and S. L. Turner. (2000).** The evolution of bacterial plasmids, p. 301-361. In C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood academic publishers, Amsterdam.
- Pacek, M., G. Konopa, and I. Konieczny. (2001).** DnaA box sequences as the site for helicase delivery during plasmid RK2 replication initiation in *Escherichia coli*. *J Biol Chem* **276**:23639-23644.
- Pak, M. and S. Wickner. (1997).** Mechanism of protein remodeling by ClpA chaperone. *Proc Natl Acad Sci USA* **94**:4901-4906.
- Pal, S. K. and D. K. Chattoraj. (1988).** P1 plasmid replication: initiator sequestration is inadequate to explain control by initiator-binding sites. *J Bacteriol* **170**:3554-3560.
- Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, and V. A. Stanisich. (1994).** Complete nucleotide sequence of Birmingham IncP $\alpha$  plasmids: compilation and comparative analysis. *J Mol Biol* **239**:623-663.
- Park, K., S. Mukhopadhyay, and D. K. Chattoraj. (1998).** Requirements for and regulation of origin opening of plasmid P1. *J Biol Chem* **273**:24906-24911.
- Park, K., E. Han, J. Paulsson, and D. K. Chattoraj. (2001).** Origin pairing ('handcuffing') as a mode of negative control of P1 plasmid copy number. *EMBO J* **20**:7323-7332.
- Park, K. and D. K. Chattoraj. (2001).** DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. *J Mol Biol* **310**:69-81.



- Patel, P. H., M. Suzuki, E. Adman, A. Shinkai, and L. A. Loeb. (2001).** Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J Mol Biol* **308**:823-837.
- Patel, S. S. and K. M. Picha. (2000).** Structure and function of hexameric helicases. *Annu Rev Biochem* **69**:651-697.
- Paulsson, J. (2002).** Multileveled selection on plasmid replication. *Genetics* **161**:1373-1384.
- Perri, S. and D. R. Helinski. (1993).** DNA sequence requirements for interaction of the RK2 replication initiation protein with plasmid origin repeats. *J Biol Chem* **268**:3662-3669.
- Persson, C. and K. Nordström. (1986).** Control of replication of the broad host range plasmid RSF1010: The incompatibility determinant consists of directly repeated sequences. *Mol Gen Genet* **203**:189-192.
- Peters, M., E. Jögi, I. Suitso, T. Punnisk, and A. Nurk . (2001).** Features of the replicon of plasmid pAM10.6 of *Pseudomonas fluorescens*. *Plasmid* **46**:25-36
- Polesky, A. H., T. A. Steitz, N. D. Grindley, and C. M. Joyce. (1990).** Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. *J Biol Chem* **265**:14579-14591.
- Rasooly, A. and R. P. Novick. (1993).** Replication-specific inactivation of the pT181 plasmid initiator protein. *Science* **262**:1048-1050.
- Ratnakar, P. V., B. K. Mohanty, M. Lobert, and D. Bastia. (1996).** The replication initiator protein  $\pi$  of the plasmid R6K specifically interacts with the host-encoded helicase DnaB. *Proc Natl Acad Sci USA* **93**:5522-5526.
- Rawlings, D. E., I. Pretorius, and D. R. Woods. (1984).** Expression of a *Thiobacillus ferrooxidans* origin of replication in *Escherichia coli*. *J Bacteriol* **158**:737-738.
- Rawlings, D. E. and D. R. Woods. (1985).** Mobilization of *Thiobacillus ferrooxidans* plasmids among *Escherichia coli* strains. *Appl Environ Microbiol* **49**:1323-1325.
- Rawlings, D. E., R. A. Dorrington, J. Rohrer, and A.-M. Clennel. (1993).** A molecular analysis of the replication and mobilization regions of a broad-host-range plasmid isolated from *Thiobacillus ferrooxidans*. *FEMS Microbiol Rev* **11**:3-7.
- Rawlings, D. E. and T. Kusano. (1994).** Molecular genetics of *Thiobacillus ferrooxidans*. *Microbiol Rev* **58**:39-55.
- Rawlings, D. E. and S. Silver. (1995).** Mining with microbes. *Bio/Technology* **13**:773-778.
- Rawlings, D. E. (1997).** Mesophilic, autotrophic bioleaching bacteria: description, physiology and role, p. 229-245. In D. E. Rawlings (ed.), *Biomining: Theory, Microbes and Industrial processes*. Springer-Verlag, New York.
- Rawlings, D. E. (1998).** Industrial practice and the biology of leaching metals from ores. *J Ind Microbiol Biotech* **20**:268-274.



- Rawlings, D. E., N. J. Coram, M. N. Gardner, and S. M. Deane. (1999[a]).** *Thiobacillus caldus* and *Leptospirillum ferrooxidans* are widely distributed in continuous flow biooxidation tanks used to treat a variety of metal containing ores and concentrates, p. 777-786. In R. Amils and A. Ballester (eds.), *Biohydrometallurgy and the environment toward the mining of the 21st century*. Elsevier, Amsterdam, The Netherlands.
- Rawlings, D. E., H. Tributsch, and G. S. Hansford. (1999[b]).** Reasons why 'Leptospirillum'-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. *Microbiology* **145**:5-13.
- Rawlings, D. E. and E. Tietze. (2001).** Comparative biology of IncQ and IncQ-like plasmids. *Microbiol Mol Biol Rev* **65**:481-96.
- Rawlings, D. E. (2002).** Heavy metal mining using microbes. *Annu Rev Microbiol* **56**:65-91.
- Rawlings, D. E., D. Dew, and C. du Plessis. (2003).** Biomineralization of metal-containing ores and concentrates. *Trends Biotechnol* **21**:38-44.
- Rohrer, J. and D. E. Rawlings. (1993).** Regulation of mobilization of the broad-host-range plasmid pTF-FC2. *Mol Microbiol* **9**:1051-1059.
- Ryan, V. T., J. E. Grimwade, C. J. Nievera, and A. C. Leonard. (2002).** IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli oriC* by two different mechanisms. *Mol Microbiol* **46**:113-124.
- Sakai, H. and T. Komano. (1996).** DNA replication of IncQ broad-host-range plasmids in gram-negative bacteria. *Biosci Biotech Biochem* **60**:377-382.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989).** *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schaper, S. and W. Messer. (1995).** Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* **270**:17622-17626.
- Scherzinger, E., M. M. Bagdasarian, P. Scholz, R. Lurz, B. Rückert, and M. Bagdasarian. (1984).** Replication of the broad host range plasmid RSF1010: requirement for three plasmid-encoded proteins. *Proc Natl Acad Sci USA* **81** :654-658.
- Scherzinger, E., V. Haring, R. Lurz, and S. Otto. (1991).** Plasmid RSF1010 DNA replication *in vitro* promoted by purified RSF1010 RepA, RepB and RepC proteins. *Nucleic Acids Res* **19**:1203-1211.
- Scherzinger, E., G. Ziegelin, M. Bárcena, J. M. Carazo, R. Lurz, and E. Lanka. (1997).** The RepA protein of plasmid RSF1010 is a replicative DNA helicase. *J Biol Chem* **272**:30228-30236.
- Schippers, A. and W. Sand. (1999).** Bacterial leaching of metal sulfides proceeds by two indirect mechanisms via thiosulfate or via polysulfides and sulfur. *Appl Environ Microbiol* **65**:319-321.



- Scholz, P., V. Haring, E. Scherzinger, R. Lurz, M. M. Bagdasarian, H. Schuster, and M. Bagdasarian. (1985).** Replication determinants of the broad-host-range plasmid RSF1010, p. 243-259. *In* D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender (eds.), *Plasmids in Bacteria*. Plenum Press, New York.
- Scholz, P., V. Haring, B. Wittmann-Liebold, K. Ashman, M. Bagdasarian, and E. Scherzinger. (1989).** Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* **75**:271-288.
- Seitz, H., C. Weigel, and W. Messer. (2000).** The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Mol Microbiol* **37** :1270-1279.
- Sharma, R., A. Kachroo, and D. Bastia. (2001).** Mechanistic aspects of DnaA-RepA interaction as revealed by yeast forward and reverse two-hybrid analysis. *EMBO J* **20**:4577-4587.
- Skarstad, K. and E. Boye. (1994).** The initiator protein DnaA: evolution, properties and function. *Biochim Biophys Acta* **1217**:111-130.
- Smalla, K., H. Heuer, A. Götz, D. Niemeyer, E. Krögerrecklenfort, and E. Tietze. (2000[a]).** Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity of IncQ-like plasmids. *Appl Environ Microbiol* **66**:4854-4862.
- Smalla, K., M. Osborn, and E. M. H. Wellington. (2000[b]).** Isolation and Characterisation of plasmids from bacteria, p. 207-248. *In* C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood academic publishers, Amsterdam.
- Smith, A. S. and D. E. Rawlings. (1997).** The poison-antidote stability system of the broad-host-range *Thiobacillus ferrooxidans* plasmid pTF-FC2. *Mol Microbiol* **26**:961-970.
- Smith, A. S. and D. E. Rawlings. (1998).** Efficiency of the pTF-FC2 pas poison-antidote stability system in *Escherichia coli* is affected by the host strain, and antidote degradation requires the lon protease. *J Bacteriol* **180**:5458-5462.
- Snyder, L. and W. Champness. (2003).** *Molecular genetics of bacteria*. 2<sup>nd</sup> Edition. ASM Press, Washington, D.C.
- Sobecky, P. A., T. J. Mincer, M. C. Chang, and D. R. Helinski. (1997).** Plasmids isolated from marine sediment microbial communities contain replication and incompatibility regions unrelated to those of known plasmid groups. *Appl Environ Microbiol* **63**:888-895.
- Stenzel, T. T., T. MacAllister, and D. Bastia. (1991).** Cooperativity at a distance promoted by the combined action of two replication initiator proteins and a DNA bending protein at the replication origin of pSC101. *Genes Dev* **5** :1453-1463.
- Sutton, M. D. and G. C. Walker. (2001).** Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc Natl Acad Sci U S A* **98**:8342-8349.
- Sýkora, P. (1992).** Macroevolution of plasmids: A model for plasmid speciation. *J Theor Biol* **159**:53-65.



- Takechi, S., H. Matsui, and T. Itoh. (1995).** Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. *EMBO J* **14** :5141-5147.
- Tanaka, K., K. Kino, Y. Taguchi, D.-M. Miao, Y. Honda, H. Sakai, T. Komano, and M. Bagdasarian. (1994).** Functional difference between two oppositely oriented priming signals essential for the initiation of the broad host-range plasmid RSF1010 replication. *Nucleic Acids Res* **22**:767-772.
- Thomas, C. M., D. M. Stalker, and D. R. Helinski. (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.
- Thomas, C. M., M. A. Cross, A. A. Hussain, and C. A. Smith. (1984).** Analysis of copy number control elements in the region of the vegetative replication origin of the broad host range plasmid RK2. *EMBO J* **3**:57-63.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**:4673-4680.
- Tietze, E. (1998).** Nucleotide sequence and genetic characterization of the novel IncQ-like plasmid pIE1107. *Plasmid* **39**:165-181.
- Toukdarian, A. E., D. R. Helinski, and S. Perri. (1996).** The plasmid RK2 initiation protein binds to the origin of replication as a monomer. *J Biol Chem* **271**:7072-7078.
- Tsutsui, H., A. Fujiyama, T. Murotsu, and K. Matsubara. (1983).** Role of nine repeating sequences of the mini-F genome for expression of F-specific incompatibility phenotype and copy number control. *J Bacteriol* **155**:337-344.
- Urh, M., J. Wu, J. Wu, K. Forest, R. B. Inman, and M. Filutowicz. (1998).** Assemblies of replication initiator protein on symmetric and asymmetric DNA sequences depend on multiple protein oligomerization surfaces. *J Mol Biol* **283** :619-631.
- Versalovic, J. and J. R. Lupski. (1993).** The *Haemophilus influenzae* *dnaG* sequence and conserved bacterial primase motifs. *Gene* **136**:281-286.
- Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. (1982).** Distantly related sequences in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**:945-951.
- Wang, P.-Z., S. J. Projan, V. Henriquez, and R. P. Novick. (1993).** Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element. *EMBO J* **12**:45-52.
- Whittle, G., M. E. Katz, E. H. Clayton, and B. F. Cheetham. (2000).** Identification and characterization of a native *Dichelobacter nodosus* plasmid, pDN1. *Plasmid* **43**:230-234.
- Wickner, S., D. Skowyra, J. Hoskins, and K. McKenney. (1992).** DnaJ, DnaK, and GrpE heat shock proteins are required in *oriP1* DNA replication solely at the RepA monomerization step. *Proc Natl Acad Sci USA* **89**:10345-10349.



- Wu, F., I. Levchenko, and M. Filutowicz. (1994).** Binding of DnaA protein to a replication enhancer counteracts the inhibition of plasmid R6K  $\gamma$  origin replication mediated by elevated levels of R6K  $\pi$  protein. *J Bacteriol* **176** :6795-6801.
- Wu, J., M. Sektas, D. Chen, and M. Filutowicz. (1997).** Two forms of replication initiator protein: positive and negative controls. *Proc Natl Acad Sci USA* **94**:13967-13972.
- Xu, H., G. Ziegelin, W. Schröder, J. Frank, S. Ayora, J. C. Alonso, E. Lanka, and W. Saenger. (2001).** Flavones inhibit the hexameric replicative helicase RepA. *Nucleic Acids Res* **29**:5058-5066.
- Yamaguchi, K. and M. Yamaguchi. (1984).** The replication origin of pSC101: the nucleotide sequence and replication functions of the ori region. *Gene* **29** :211-219.
- Yanisch-Perron, C., J. Vieira, and J. Messing. (1985).** Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103-119.
- Yoda, K. and T. Okazaki. (1991).** Specificity of recognition sequence for *Escherichia coli* primase. *Mol Gen Genet* **227**:1-8.
- Yuzhakov, A., Z. Kelman, and M. O'Donnell. (1999).** Trading places on DNA - a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell* **96**:153-163.
- Zavitz, K. H. and K. J. Marians. (1991).** Dissecting the functional role of PriA protein-catalysed primosome assembly in *Escherichia coli* DNA replication. *Mol Microbiol* **5**:2869-2873.
- Zhang, S. and R. Meyer. (1997).** The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol Microbiol* **25**:509-516.
- Zheng, W., Z. Li, K. Skarstad, and E. Crooke. (2001).** Mutations in DnaA protein suppress the growth arrest of acidic phospholipid-deficient *Escherichia coli*. *EMBO J* **20**:1164-1172.
- Zhou, H. and R. J. Meyer. (1990).** Deletion of sites for initiation of DNA synthesis in the origin of broad host-range plasmid R1162. *J Mol Biol* **214**:685-697.
- Zhou, H., C. Byrd, and R. J. Meyer. (1991).** Probing the activation of the replicative origin of broad host-range plasmid R1162 with Tus, the *E. coli* anti-helicase protein. *Nucleic Acids Res* **19**:5379-5383.