

Plasmid-Mediated Restriction Evasion Mechanisms

Thesis submitted for the degree of
Doctor of Philosophy
at the
University of Leicester

by

Nicola Jayne Althorpe BSc

Department of Genetics
University of Leicester

September 1999

UMI Number: U594550

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U594550

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Plasmid-Mediated Restriction Evasion Mechanisms

Nicola Jayne Althorpe

Plasmid ColIb-P9 (IncI1) encodes mechanisms which allow it to avoid destruction by type I and type II restriction enzymes during transfer by conjugation between strains of *Escherichia coli*. A genetic system was developed to analyse these mechanisms. The system relied on measuring ColIb-mediated rescue of the restriction-sensitive plasmid R751 (IncP β) from destruction by *EcoKI* (type I) and *EcoRI* (type II). One ColIb mechanism was known to involve a plasmid-encoded antirestriction gene known as *ardA*, the product of which is active against type I enzymes. Tests for alleviation of *EcoKI* restriction of R751, showed strong protection by a co-transferring ColIb (Ard⁺) plasmid, slight protection when ColIb was resident in the recipient and no effect when ColIb was immobilised in the donor by removal of its *nic* site. Hence, expression of *ardA* is activated in the recipient cell following transfer; no detectable transfer of the ArdA protein occurs from the donor to the recipient.

The *ardA* gene is found in the leading region of ColIb, which is defined as the first segment of the plasmid to enter the recipient cell during conjugation. Nucleotide sequencing of 11.7 kb of this region identified ten open reading frames. Furthermore, the region also contains three dispersed repeat sequences homologous to a novel single-stranded DNA promoter described by Masai and Arai (1997, *Cell* 89, 897-907). It is proposed that these secondary structures form in the transferring T-strand of ColIb and function as promoters for transcription of genes encoded on the unique plasmid strand transferred during conjugation.

Another mechanism, which acts independently of *ardA*, alleviates restriction of both type I and type II enzymes in the recipient in second or subsequent rounds of transfer. Two separate mechanisms appear to be operating since alleviation of type I restriction occurs *in trans* and is constitutive. In contrast, alleviation of type II restriction is by a *cis*-acting mechanism. The 'substrate saturation' hypothesis, whereby increasing amounts of transferred DNA saturates the restriction system (Read *et al.* 1992, *Mol Microbiol* 6, 1933-1941) is ruled out by data presented.

Acknowledgements

I would like to thank the following;

- Brian Wilkins,
- Angela Thomas,
- Matthew Thrower,
- Emma 'Chick' Swanston,
- Lynn Cook,
- Kerry 'Kesser' Byrne,
- Pat Wilson,
- Stevey Bates,
- Bill Brammar,
- Steve Saville,
- The Medical Research Council,
- and all the Yeasty Boys,

I would also like to thank any friends and colleagues not mentioned above and my family for putting up with me during the past four years.

Abbreviations

Ala	Alanine
Arg	Arginine
Asn	Aspartic acid
Asp	Asparagine
ATP	Adenosine triphosphate
Ap ^R	Ampicillin resistance
bp	Base pair
BSA	Bovine Serum Albumin
Collb	Collb-P9
CCC	Covalently closed circle
Cm ^R	Plasmid-encoded chloramphenicol resistance
Clm ^R	Chromosomal-encoded chloramphenicol resistance
Cys	Cysteine
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dATP	Deoxyribo-adenosine triphosphate
dCTP	Deoxyribo-cytosine triphosphate
dGTP	Deoxyribo-guanidine triphosphate
dTTP	Deoxyribo-thimidine triphosphate
<i>drd</i>	Derepressed for transfer
ds	Double-stranded
EDTA	Diaminoethanetetra-acetic acid
e.o.p	Efficiency of plating
Gln	Glutamic acid
Glu	Glutamine
Gly	Glycine
HGT	High gelling temperature
His	Histidine
HTH	Helix-turn-helix
Ile	Iso-leucine
Inc	Incompatibility group
IPTG	Isopropylthio- β -D-galactoside
IR	Inverted repeat
kb	Kilobase
kDa	Kilodalton
Km ^R	Plasmid-encoded kanamycin resistance
Kan ^R	Chromosome-encoded kanamycin resistance

Leu	Leucine
LPS	Lipopolysaccharide
Lys	Lysine
m ⁺ m ⁻	Modification proficient Modification deficient
Met	Methionine
Nal ^R	Chromosome-encoded Naladixic acid resistance
nt	Nucleotide
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
Phe	Phenylalanine
Pro	Proline
r ⁺ r ⁻	Restriction proficient Restriction deficient
rbs	Ribosome binding site
Rif ^R	Chromosome-encoded rifampicin resistance
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Rotations per minute
SDS	Sodium dodecyl sulphate
Ser	Serine
Sm ^R	Plasmid-encoded streptomycin resistance
Sp ^R	Plasmid-encoded spectomycin resistance
ss	Single-stranded
SSB	Single-stranded DNA-binding protein
ssi	Single-stranded initiation sequence
Str ^R	Chromosome-encoded streptomycin resistance
TAE	Tris-acetate EDTA buffer
Tc ^R	Plasmid-encoded tetracycline resistance
Tet ^R	Chromosome-encoded tetracycline resistance
Thr	Threonine
Tn	Transposon
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultraviolet
Val	Valine
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Contents

Abstract

Acknowledgements

Abbreviations

Chapter One	General Introduction	1
1.1	Prologue	1
1.2	Enterobacterial Plasmids	3
1.2.1	Incl1 plasmids	4
1.3	Enterobacterial Conjugation	6
1.3.1	Conjugative cell-cell interactions	6
1.3.2	DNA processing reactions	9
1.3.3	Plasmid-encoded exclusion	15
1.4	Barriers to Horizontal DNA Transfer by Plasmids	18
1.5	Bacterial Restriction-Modification Systems	20
1.5.1	Type I R-M systems	21
1.5.2	Type I enzyme activity	24
1.5.3	Regulation of expression of type I R-M genes	25
1.5.4	Type II R-M systems	29
1.5.5	Regulation of type II R-M genes	31
1.5.6	Ecological roles of bacterial R-M systems	32
1.6	ColIb Leading Region Genes and Their Roles in Plasmid Establishment	34
1.6.1	Single-stranded DNA-binding protein (SSB)	35
1.6.2	Plasmid SOS inhibition (PsiB)	36
1.6.3	Alleviation of restriction of DNA (ArdA)	37
1.6.4	Zygotic Induction of leading region genes	38
1.6.5	Hypotheses for zygotic Induction	40
1.6.5.1	Modes of positive regulation	40

	1.6.5.2 Modes of negative regulation	41
	1.7 Restriction-Avoidance by Plasmids	41
	1.8 Aims and Direction of this Study	43
Chapter Two	Materials and Methods	
	2.1 Plasmids and Bacterial Strains	45
	2.2 Media and Chemicals	46
	2.2.1 Media	46
	2.2.2 Antibacterial agents	48
	2.2.3 11.1 x PCR buffer	48
	2.3 Phenotypic Characterisation of Bacterial Strains	48
	2.3.1 Colicin production	48
	2.3.2 <i>recA</i> mutants	49
	2.3.3 <i>recD</i> mutants	49
	2.3.4 Sensitivity to bacteriophage	49
	2.4 Strain Manipulations	50
	2.4.1 Bacterial conjugation	50
	2.4.2 Transformation of bacterial strains	50
	2.4.3 Bacteriophage propagation	51
	2.4.4 Transposon insertion mutagenesis	51
	2.5 DNA Manipulations	52
	2.5.1 Small scale plasmid DNA isolation	52
	2.5.2 Large scale plasmid DNA isolation	53
	2.5.3 Analysis of the concentration and purity of DNA samples	53
	2.5.4 Restriction enzyme analysis of DNA	53
	2.5.5 Isolation of DNA restriction fragments	54
	2.5.6 Dephosphorylation of DNA fragments	54
	2.5.7 Ligation of DNA fragments	54
	2.5.8 Primers	55

2.5.9	DNA sequencing reactions	55
2.6	Standard and Inverse PCR	55
2.6.1	Optimisation of the inverse-PCR protocol for amplification of template pNA2 using primer 5 and primer 6	56

Chapter Three

	Genetic Analysis of a Plasmid-Mediated Restriction Evasion Mechanism	
3.1	Introduction	58
3.2	Results and Discussion	59
3.2.1	Effects of restriction on the conjugative transmission of plasmids R751 and <i>Collbdrd</i>	59
3.2.2	Effects of co-transferring R751 with <i>Collbdrd</i> to a restricting recipient	61
3.2.3	Rescue of R751 from restriction by <i>Collbdrd</i> resident in the restricting recipient	63
3.2.4	Is conjugative transfer of <i>Collbdrd</i> a prerequisite for alleviation of restriction?	65
3.3	Summary	68

Chapter Four

	Construction of an <i>oriT</i> Mutant of <i>Collbdrd</i> and its Conjugative Properties	
4.1	Introduction	70
4.2	Results	71
4.2.1	Cloning and determination of the nucleotide sequence of the <i>Collb oriT</i> locus	71
4.2.2	Strategy employed to delete the <i>nic</i> sequence from the cloned <i>Collb oriT</i>	74
4.2.3	Construction of pNA8 containing the omega (Ω) intersposon	75
4.2.4	Construction of a <i>Collbdrd</i> mutant lacking the <i>nic</i> region	76

	4.2.5	Conjugative properties of pNA11 and pNA12	76
	4.3	Discussion	78
Chapter Five		Analysis of the Alleviation of Type I and Type II Restriction Using Non-Transferable Mutants pNA11 and pNA12	
	5.1	Introduction	80
	5.2	Results	80
	5.2.1	Rescue of R751 from <i>EcoRI</i> restriction by <i>ColIbdrd oriT</i> mutant pNA11	80
	5.2.2	Mobilisation of <i>EcoRI</i> producing plasmid NTP14 by <i>ColIbdrd</i>	81
	5.2.3	A non-transferable <i>ColIbdrd</i> confers no <i>EcoKI</i> protection on R751	82
	5.2.4	<i>ColIbdrd</i> transfer does not mediate retrotransfer of NTP14	83
	5.3	Discussion	84
Chapter Six		Construction of a <i>mob</i> Mutant of Plasmid NTP14	
	6.1	Introduction	86
	6.2	Results	
	6.2.1	Restriction analysis of plasmid NTP14	88
	6.2.2	Construction of pNA16 and pNA17	89
	6.2.3	Characterisation of pNA16 and pNA17	89
	6.2.4	<i>ColIb</i> does not alleviate <i>EcoRI</i> restriction <i>in trans</i> .	91
	6.2.5	Conjugative transfer of the <i>EcoRI</i> R-M genes causes some killing of recipient cells	92
	6.3	Discussion	93

Chapter Seven	Determination and Analysis of the Nucleotide Sequence Of the ColIb Leading Region	
7.1	Introduction	96
7.2	Results and Discussion	97
7.2.1	Strategy for determining the nucleotide sequence of the ColIb leading region	97
7.2.2	Identification of putative protein coding regions (<i>orfs</i>) in the ColIb leading region	98
7.2.3	Similarities between DNA and protein sequences found in the leading region to those present in the databases	99
7.2.4	G + C content of the ColIb leading region	103
7.2.5	Similarity of the ColIb leading region to other conjugative plasmids	104
7.2.6	Examination of the ColIb leading region for type I restriction enzyme cleavage sites	104
7.2.7	Repeat sequences within the leading region	105
7.3	Summary	106
Chapter Eight	General Discussion	108
Appendices	Appendix I	
Literature cited		

Chapter One

General Introduction

1.1 Prologue

Extrachromosomal genetic elements known as plasmids are commonly found in bacteria and determine functions involved in their stable replication-maintenance. Many plasmids also possess transfer genes (*tra*) that support horizontal transfer of the element between bacteria by the process of conjugation. The first conjugative plasmid to be recognised was the F fertility factor of *Escherichia coli* (Lederberg and Tatum, 1946). Since the discovery of F, the majority of research has focused on developing a molecular understanding of the conjugation system of F and other enterobacterial plasmids (Wilkins and Lanka, 1993). To date, the most documented transfer systems belong to plasmids of the incompatibility (Inc) complexes F, I, P, N and W. Essential features of these transfer systems, including strategies for mating-pair formation and transmission of genetic material, are conserved between an extensive collection of plasmids representing more than 25 Inc groups isolated from Gram-negative bacteria (Wilkins and Lanka, 1993).

These conserved features apply less to the conjugation systems of plasmids isolated from Gram-positive bacteria. In fact, the similarities that exist are that they require physical contact between donor and recipient cells and that conjugation is mediated by plasmids or a special class of conjugative transposon. Thus, a global definition of conjugation is that it is a “process whereby a DNA molecule is transferred from a donor to a recipient bacteria via a specialised protein complex called the conjugation apparatus” (Zechner *et al.*, 1999).

In addition to housekeeping functions involved in replication-maintenance and transfer, many plasmids carry an array of specialised genes of high environmental and medical importance. The most interesting of these

genes are those that encode resistance to antibiotics and virulence determinants. One feature of plasmid transfer that has brought it much attention is its so-called promiscuity; that is, the ability of certain types of plasmids to mediate transfer between bacteria exceeding their maintenance-range. During such transfer, there is the potential for the exchange of genetic material by the process of recombination. Therefore, plasmids make a significant contribution to the gene pool and hence, play an important role in the ecology of bacteria (Datta and Hughes, 1983; Mazodier and Davis, 1991; Amàbile-Cuevas and Chicurel, 1992).

One potential barrier to horizontal transfer of DNA is that imposed by bacterial restriction-modification (R-M) systems. These are ubiquitous in nature and are therefore likely to be encountered frequently by transferring DNA. The restriction endonuclease component operates to cleave infecting DNA recognised as foreign by its lack of cognate methylation at target sites for the enzyme. Plasmids transferring and replicating within a single host strain will acquire the appropriate methylation and avoid destruction by the restriction enzymes of the host. In contrast, plasmids transferring to new host strains will not carry the appropriate methylation. Therefore, promiscuous transfer raises the intriguing question of how plasmids avoid destruction from newly encountered R-M systems.

Bacteriophages, like plasmids, are also subject to destruction by host-encoded restriction enzymes during infection of a new strain. However, bacteriophages are known to encode a variety of anti-restriction mechanisms. It is now becoming clear that certain plasmids also encode such protective mechanisms. This thesis focuses on one such plasmid, ColIb-P9 of the IncI1 group of plasmids, and attempts to characterise further the plasmid's two restriction-evasion mechanisms. Data described in this thesis have been published in the papers Althorpe *et al.* (1999) and Bates *et al.* (1999).

1.2 Enterobacterial Plasmids

The majority of large plasmids (> 20 kb) isolated from Gram-negative bacteria carry transfer genes that promote transmission of the element to a new bacterial cell by the process of conjugation. Such plasmids are described as conjugative or self-transmissible. Most plasmids are generally classified according to incompatibility (Inc) relationships based initially on the inability of two closely related or identical plasmid replicons to be propagated in the same cell line. The phenomenon of incompatibility reflects the identity between the two plasmids and their replication control and active partitioning systems (Couturier *et al.*, 1988). Further classification of plasmids may involve criteria such as i) replicon-typing, ii) DNA hybridisation tests, iii) morphological and serological typing of conjugative pili and iv) sensitivity to pilus-specific bacteriophages (Bradley, 1980). As a result of these tests, it is the trend that plasmids from the same Inc group often share similar if not identical features between their transfer systems.

The subject of this thesis is the 93-kb enterobacterial plasmid ColIb-P9 (Fig. 1.1), which along with closely related plasmid R64 represent the IncI1 group of enterobacterial plasmids. ColIb-P9 was originally isolated in *Shigella sonnei* P9 and like F was one of the first conjugative plasmids discovered (Frédéricq, 1965). Consequently, ColIb-P9 has featured extensively in the literature regarding the fundamentals of conjugation. However, the complex details regarding the conjugation system of this plasmid and F are beyond the scope of this thesis. Therefore, the reader is directed to a number of extensive reviews that exist for further details (Willetts and Wilkins, 1984; Willetts and Skurray, 1987; Wilkins and Lanka, 1993; Lanka and Wilkins, 1995; Firth *et al.*, 1996; Pansegrau and Lanka, 1996 and Zechner *et al.*, 1999). To this end, only a general overview of the details regarding I1-like conjugation will be described in the following sections. Where appropriate, analogies with F- and P-like conjugation will be made.

Fig. 1.1 Map of IncI1 plasmid ColIb.

Regions Tra1 and Tra2 contain the plasmids transfer genes. Loci within Tra1 include the shufflon (*shf*), shufflon recombinase (*rci*), the flexible pilus structural gene (*pilV*) and genes encoding an EDTA-resistant nuclease (*nuc*). Tra2 includes DNA primase (*sog*), entry exclusion (*eex*) and genes involved in rigid pilus assembly. Adjacent to the leading region is the origin of transfer (*oriT*) and genes encoding relaxosome proteins (*nikA*, *nikB*). The ColIb Rep region contains the origin of vegetative replication (*oriV*), I1-type incompatibility locus (*inc*) and replication controlling factor (*repZ*). Leading region genes shown are those that encode the single-stranded DNA binding protein (*ssb*), plasmid SOS-inhibition (*psiB*) and alleviation of restriction of DNA (*ardA*). Other loci include the colicin Ib gene (*cib*) and colicin Ib immunity gene (*imm*), fertility inhibition of F (*FinQ*) and genes determining abortive phage infection (*ibfA*, *ibfB*). Indicated on the map by internal lines are *EcoRI* cleavage sites. Triangles (Δ) indicate *SalI* cleavage sites. Map represents data from studies with both ColIb and closely related IncI1 plasmid R64 (Rees *et al.*, 1987; Howland *et al.*, 1989; Komano *et al.*, 1990; Furuya *et al.*, 1991 and Chilley and Wilkins, 1995). Map is reproduced from Read (1993).

Fig. 1.1

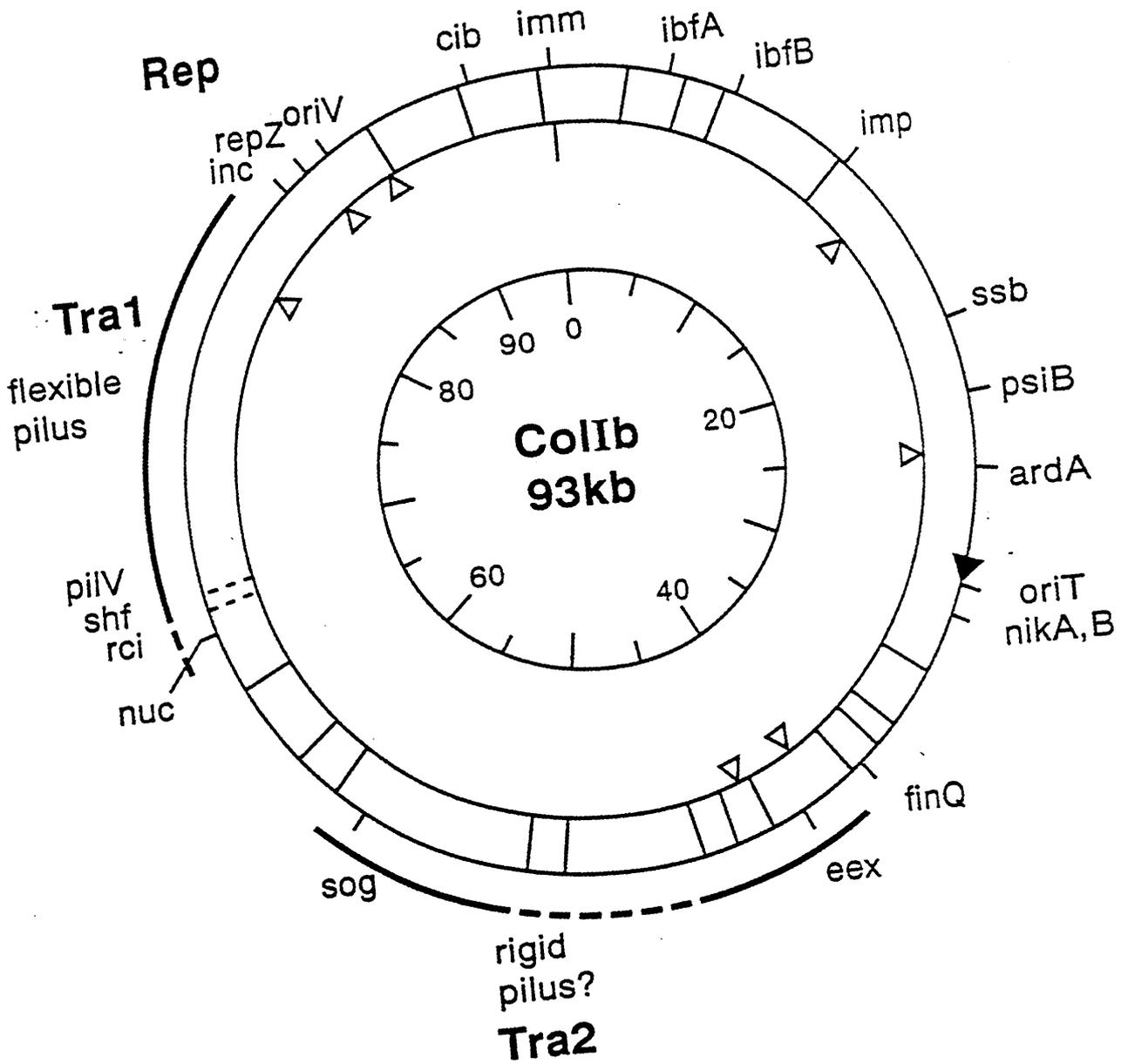
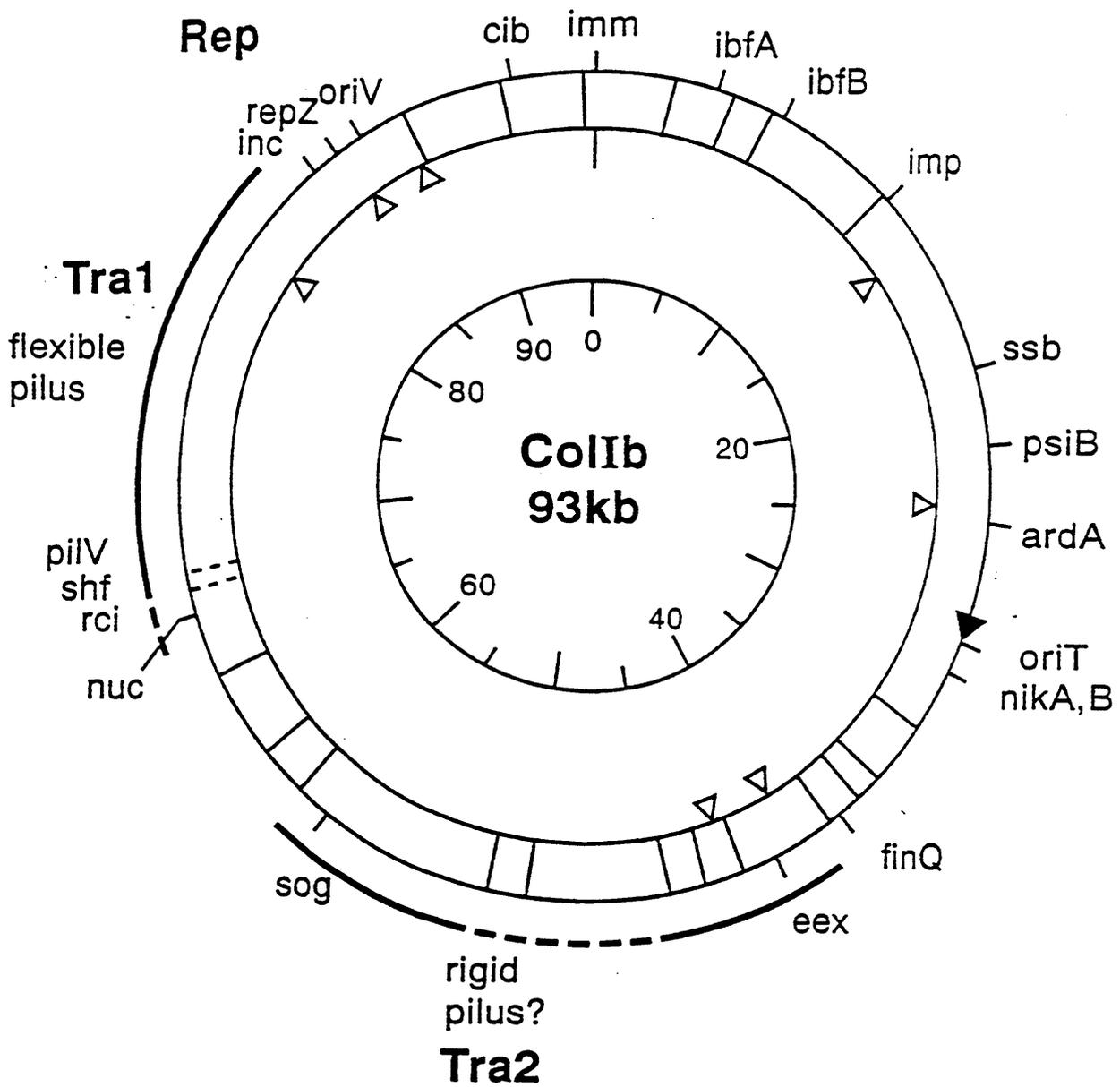


Fig. 1.1



1.2.1 IncI1 plasmids

Plasmids belonging to the I complex are typified by those which specify homologous thin conjugative pili and include plasmids from the six Inc groups I1 (I α), I2, I5, B, K and Z. Some of these plasmids can be further classified into one of two groups based on replicon-typing and serological relationships that exist between the conjugative apparatus they specify. These include the plasmids from the Inc groups I1-B-K and the IncI2 plasmids. Despite the relationships between these pili, plasmids from the I complex display different patterns of sensitivity to I-pilus specific bacteriophages If1 and PR64FS (Walia and Duckworth, 1986; Bradley, 1984).

All I-type plasmids determine two types of conjugative pilus. The thick rigid pili are one type, which are all morphologically similar and bear comparisons to those specified by plasmids of IncP and IncN groups as they support efficient conjugation between cells on semi-solid surfaces. However, these thick rigid pili are also required in liquid matings and are thought to be essential for DNA transfer (Bradley, 1980; Bradley, 1984; Rees *et al.*, 1987; Horiuchi and Komano, 1998). The second type is the thin flexible pilus, which is required to support efficient conjugation between cells in liquid media. Through DNA sequence analysis of the R64 genes responsible for thin pili biosynthesis, these pili were inferred to belong to the type IV pilus family (Kim and Komano, 1997). R64 thin pili consist of the major subunit PilS and one of seven PilV products (Horiuchi and Komano, 1998).

The I1-B-K plasmids have a narrow replication-maintenance range and can be only maintained in strains of *Escherichia*, *Salmonella*, *Shigella* and *Klebsiella* (Walia and Duckworth, 1986; Wilkins, 1995). Through replicon typing tests, the replication region of the IncI1 plasmids along with those of the groups IncFII, IncFI and IncFIV, have been shown to belong to the RepFIC family of replicons. Apparently, the narrow maintenance range of this replicon-type is due to the requirement for host encoded enzymes (Jones, 1991; Couturier *et al.*, 1988; Saadi *et al.*, 1987). Despite the narrow transmission range

of the IncI1 plasmids they are still able to promote low-level transfer to a broad range of bacteria (Wilkins, 1995).

A large proportion of the backbone of the IncI1 plasmid is assigned to the transfer system of the plasmid (~50 kb). The transfer genes of ColIb are organised into two separate modules, Tra1 and Tra2, as determined by genetic mapping studies of the plasmid [Fig. 1.1] (Uemura and Mizobuchi, 1982; Rees *et al.*, 1987).

Tra1 is located between Tra2 and the *rep* region and was shown by transposon mutagenesis to carry genes that are required for the assembly of the thin flexible conjugative pilus (Rees *et al.*, 1987). The composition of the thin pilus can be changed by rearrangements in a DNA sequence known as the shufflon (Komano *et al.*, 1990). This highly mobile DNA sequence consists of three 300 bp segments separated by six 19 bp repeat sequences and is located at the 5' terminus of the thin-pilus gene, *pilV*. Site-specific recombination mediated by the product of the associated *rci* gene selects a different C-terminal segment of *pilV*. Consequently, the conjugation efficiency in liquid environments is changed, possibly by affecting the recognition of surface components of different recipients (Komano *et al.*, 1994; Horiuchi and Komano, 1998).

The Tra2 segment contains genes involved in the assembly of the thick rigid conjugative pilus, entry exclusion (*eex*) and a gene encoding a DNA primase known as *sog* (Hartskeerl and Hoekstra, 1984; Rees *et al.*, 1987).

Expression of these transfer genes, with the exception of *eex*, are subject to the plasmid's negative *trans*-acting fertility inhibition system which is analogous to that of F-like plasmids (Meynell and Datta, 1967).

Finally, the *oriT* and *oriT* operon, which are both homologous to IncP counterparts can be found located adjacent to the plasmid's leading region (Rees *et al.*, 1987; Furuya and Komano, 1991). This and the corresponding regions of other I1, B and K plasmids are found to carry highly conserved zygotically inducible genes which are not essential for conjugation but are

believed to be involved in establishment of the plasmid in the newly infected recipient cell. These genes are discussed in more detail in section 1.6.

1.3 Enterobacterial Conjugation

The model of enterobacterial conjugation is divided into three crucial stages. The first stage involves the formation of stable contacts between donor and recipient bacteria and is commonly referred to as mating-pair formation. The second stage involves a series of reactions that result in the transfer of a unique strand of the plasmid to the recipient cell. The final stage involves the establishment of the immigrant plasmid in the recipient, which may be facilitated by plasmid proteins transferred from the donor to recipient or those synthesised early in the infected cell (Wilkins and Lanka, 1993).

1.3.1 Conjugative cell-cell interactions

Interactions between donor and recipient cells at surface levels are an essential first stage to conjugation. The pilus specified by all conjugative plasmids isolated from *Escherichia coli* is an essential feature in initiating contact between mating bacteria. These donor-specific surface appendages, which are also known as sex pili, tend to fall into one of three morphological groups including thin flexible, thick flexible and rigid filaments or rods (Bradley, 1980).

The biological significance of the conjugative pilus as an organelle for mediating contact between donor and recipient bacteria is established and its importance is emphasised by the number of transfer genes that are assigned to its assembly. However, certain aspects of this initial stage of conjugation have yet to be resolved, the most interesting of which is the exact route for ssDNA transfer from the donor to the recipient bacterium after contact is initiated.

From studies with F-mediated conjugations in liquid media, the process of conjugative cell-cell interactions is proposed to proceed in several stages (Anthony *et al.*, 1994). First, the donor cells interact with recipients via the tip

of the pilus and a receptor on the surface of the recipient cell. The exact nature and identity of these cell surface receptors are not yet known. However, early experiments with ConF⁻ and ConI⁻ mutants (defective in receiving F or I plasmids by conjugation, respectively) of *E. coli* K-12 suggested that the surface lipopolysaccharide (LPS) and outer membrane protein OmpA are the most likely candidates (Havekes *et al.*, 1977; Achtman *et al.*, 1978b; Anthony *et al.*, 1994). However, whether LPS or OmpA are the actual receptors recognised by conjugative pili or are involved in the biosynthesis or structural integrity of the authentic receptor is not yet understood (Havekes *et al.*, 1977; Anthony *et al.*, 1994).

After contact is initiated, the second stage of conjugative cell-cell interactions involves the retraction of the pilus which brings the donor and recipient cells into wall-wall contact (Achtman *et al.*, 1978a). Assembly of the extended F-pilus is a complicated process that is poorly understood. However, what is known is that the process involves the products of 15 transfer genes, which form a membrane-spanning complex traversing the bacterial cell envelope from the which the pilus is extended and retracted (Manning *et al.*, 1981; Schandel *et al.*, 1992). Retraction of the F-pilus involves depolymerisation of its subunits into the cell membrane (Dürrenberger *et al.*, 1991; Frost *et al.*, 1994).

Mating aggregates in F-mediated conjugations are at first sensitive to shear forces. However, once the mating cells are in wall-wall contact the aggregate is rapidly stabilised by a process requiring the products of *traG* and *traN* (Manning *et al.*, 1981; Panicker and Minkley, 1985). TraN can be found within the outer membrane of the cell envelope of the donor and contains an exposed surface domain, which may interact with the lipopolysaccharide of the recipient (Dürrenberger *et al.*, 1991).

The exact route supporting transfer of single-stranded DNA from the donor to the cytoplasm of the recipient is unknown. Early studies with F-mediated conjugations suggested that DNA transfer might occur through the

extended pilus (Ou and Anderson, 1970). Such a view was supported by the findings of Harrington and Rogerson (1990) who found that Hfr-C donors separated from F⁻ recipients by a filter 6 µm thick with a pore size of 0.01-0.1 µm were still able to produce transconjugants. However, the favoured model to date is that the donor and recipient cell envelopes, once within intimate contact, fuse forming a transmembrane pore which allows the transfer of DNA and any proteins involved in DNA metabolism to the recipient. In matings mediated by plasmid F, TraD is found traversing the inner and outer membrane of the donor (Panicker and Minkley, 1985) and appears to interact with OmpA forming a putative pore between the donor and the outer membrane of the recipient for DNA transfer (Dürrenberger *et al.*, 1991). No evidence of a plasma bridge formed between the two cells has yet been found, which is a notion supported by Dürrenberger *et al.* (1991) who examined mating cells within wall-wall contact by electron microscopy (Willetts and Wilkins, 1984; Dürrenberger *et al.*, 1991). The model is lacking in that there is no evidence for how the plasmid DNA translocates across the inner membrane of the recipient. However, it has been proposed that this may be a passive process similar to that of infecting bacteriophage λ, which upon injection into the periplasmic space, λ DNA is transported across the inner membrane by the sugar uptake mechanism for mannose (Dürrenberger *et al.*, 1991).

The putative signal for initiation of DNA transfer after contact between the donor and recipient has been established is proposed to be the product of the F gene *traM*. TraM which is found at the base of the F pilus is required to be in close contact with transfer replication proteins and binds extensive regions of the origin of transfer [*oriT*] (Dürrenberger *et al.*, 1991; Zechner *et al.*, 1999). TraM also provides the physical link between the TraD protein forming part of the mating bridge and the relaxase protein TraI that is covalently bound to the transferring DNA. Finally, once DNA transfer has occurred conjugation is terminated and mating aggregates separate (Achtman *et al.*, 1978a).

From studies with IncP plasmids, the assembly of RP4 surface-specific pili is reasonably well understood (for review see Zechner *et al.*, 1999). In addition, the transfer genes required for RP4-mediated matings between *E. coli* and from *E. coli* to strains of yeast are known (Lessl *et al.*, 1993; Haase *et al.*, 1995; Wilkins and Bates, 1997; Bates *et al.*, 1998). However, interactions at the cellular level between donor and recipient bacteria on surfaces have been neglected and to date no Con⁻ mutants have been isolated. This is unfortunate considering that most natural communities of bacteria are attached to surfaces or grow as micro-colonies (Simonsen, 1990). Previous studies with ConF⁻ mutants isolated from liquid matings were shown to display wild type behaviour when mated with appropriate donors on surfaces, which suggests that the interactions that occur during surface matings are different from those that occur in liquid (Achtman *et al.*, 1978b)

1.3.2 DNA processing reactions

The reactions required for transfer of a unique DNA strand of the plasmid to the recipient cell during enterobacterial conjugation have been best studied for plasmid F, IncP plasmid RP4 and some mobilisable plasmids from the IncQ group. Transfer of the plasmid strand to the recipient cell during conjugation is proposed to occur in number of stages (Lanka and Wilkins, 1995).

The first stage of plasmid transfer is initiated at the *nic* site within the origin of transfer (*oriT*) by specific endonucleolytic cleavage of a single strand of the plasmid DNA (Lanka and Wilkins, 1995). Cleavage is mediated by a plasmid-encoded transfer gene product called relaxase. The relaxase enzyme binds to the *oriT* in the presence of accessory proteins to form a complex known as the relaxosome. This complex can be isolated and upon denaturation with SDS releases an open circular plasmid that is an ideal substrate for determining the *nic* site of the *oriT* by nucleotide sequencing (Fürste *et al.*, 1989; Furuya and Komano, 1991; Lanka and Wilkins, 1995).

Cleavage of *oriT* breaks a specific phosphodiester bond between two chemically different nucleotides present at the *nic* site. From studies with plasmid F, RP4, R64, RSF1010 and ColE1, the 3' terminus of *nic* has been shown to carry an unmodified hydroxyl group that is susceptible to primer extension by DNA polymerase I of *E. coli* and a 5' terminus that remains covalently attached to the relaxase enzyme via the hydroxyl side of either a seryl, threonyl or tyrosyl residue of the enzyme. The location of these active centre amino acids in the relaxase vary and can be found within different domains of the enzyme depending on which plasmid encodes them (Lanka and Wilkins, 1995). The biochemistry of these reactions at the *oriT* is best understood for RP4 relaxosomes (Pansegrau *et al.*, 1990b).

Whether cleavage of the *oriT* is influenced by an unidentified signal synthesised during cellular interactions between the donor and recipient cell is still widely debated. From studies with plasmid F, the product of the *traM* gene has been implicated in this signalling process. TraM is able to interact with TraD of the mating bridge and the *oriT*-protein complex (section 1.3.1) suggesting that the protein is involved in initiating the transfer process (Di Laurenzio *et al.*, 1991).

An alternative view to that described above is that cleavage of the *oriT* is a reversible process that occurs in the absence of recipient cells. It is envisaged that the relaxase enzyme covalently bound to the 5' terminus of the plasmid strand is also able to mediate a cleaving-joining reaction involving the hydroxyl free end of the 3' terminus. Such a property of the relaxase suggests that the enzyme is also involved in recircularisation of the transferred plasmid strand in the recipient cell and consequently in termination of conjugation (Wilkins and Lanka, 1993; Pansegrau *et al.*, 1994b). Examination of RP4 relaxosomes in non-conjugating cells revealed that TraI possesses cleaving-joining activity, which resembles that of a type I topoisomerase (Pansegrau *et al.*, 1990; Pansegrau *et al.*, 1994b). The model proposes that the energy produced from the cleaved phosphodiester bond and the formation of

the covalent TraI-*oriT* intermediate is conserved and utilised in the re-joining reaction. The superhelical covalently closed circle (CCC) state of the plasmid, which is believed to be required in the nicking reaction, is retained by the relaxase forming a non-covalent clamp to the 3' end at the *nic* site (Pansegrau *et al.*, 1990a; Matson and Morton, 1991). In addition, this CCC state may also be required so that when the plasmid is unwound the energy is provided to drive transfer of the plasmid strand to the recipient cell.

The *oriT* is the only *cis*-acting site required for DNA transfer. The location of the *oriT* is usually at one end of the transfer complex of the plasmid and can take up to 500 bp in length. The *oriT* can generally be divided into four domains including, i) the nicking domain, which is also known as the *nic* sequence, ii) an AT-rich region, which may facilitate strand separation in negatively supercoiled DNA, iii) one promoter, responsible for outward reading transcription of either the *tra* or *mob* genes and iv) a region with patterns of direct and indirect repeats that may facilitate recognition and binding of specific proteins associated with the *oriT* (Di Laurenzio *et al.*, 1991). One interesting feature of the *oriT* of plasmid F and RP4 is the requirement for the formation of secondary structures that alter the topology of the associated region and appear to enhance protein-*oriT* complex formation (Wilkins and Lanka, 1993; Pansegrau *et al.*, 1994b).

Extensive nucleotide sequence similarities are only found between the *oriT* of closely related plasmids, although significant similarities are found between the *nic* regions (~10 bp) of a more diverse set of conjugative plasmids (Furuya and Komano, 1991; Fig. 1.2). There are four families of a particular type of *nic* region that are identifiable, one of which includes the *nic* sequences from mobilisable plasmids (Pansegrau and Lanka, 1991; Lanka and Wilkins, 1995; Zechner *et al.*, 1999).

The sequence of the ColIb *oriT* was unavailable at the start of this research. However, nucleotide sequences of IncI1 plasmid R64 suggests that the I1-*oriT* bears more similarity to those of IncP plasmids RP4 and R751 than that

of plasmid F. Both the I1 and P *oriT* have a 19-bp and 17-bp inverted repeat sequence downstream of the *nic* site. In addition, the 12-bp *nic* region of R64 shares six conserved nucleotides with the *nic* region of RP4 and R751 (Komano *et al.*, 1988; Furuya and Komano, 1991; Fig. 1.2).

The relaxase of R64 is the product of the *nikB* gene, which together with *nikA* is found directly adjacent to the *oriT* as the *oriT* operon (Furuya *et al.*, 1991). NikB is responsible for cleavage possibly via the N-terminal domain of the protein, although the entire protein is required for relaxation and mobilisation of R64. NikA (15 kDa) is involved in recognition of *oriT* and directing relaxosome assembly (Furuya *et al.*, 1991; Furuya and Komano, 1991). In agreement with the finding that the amino acid sequences of certain DNA relaxases share striking similarities, both NikA and NikB have been shown to resemble transfer counterparts of IncP plasmids RP4 and R751. NikA shares 30% similarity to TraJ and the N-terminal region of NikB shares 26% identity in a 276 amino acid overlap with TraI, which is required for relaxation (Furuya and Komano, 1991; Pansegrau and Lanka, 1991).

The formation of relaxosomes at the IncI1 *oriT* is a poorly understood process. However, through *in vitro* studies involving IncP plasmids the relaxation of RP4 is proposed to occur in a sequential order, which is enhanced by the formation of a bend upstream of the *nic* site involving TraK (Pansegrau *et al.*, 1990b; Wilkins and Lanka, 1993). RP4 TraJ binds to the cognate *oriT* by recognising the right arm of the 19-bp inverted repeat sequence. RP4 TraI, which is the relaxase, binds only to the TraJ-*oriT* complex and not to the *oriT* alone (Pansegrau *et al.*, 1990b). The whole structure is subsequently stabilised by protein-protein interactions between RP4 TraH and the protein-*oriT* complex. Mutations in the *traI* gene suggest that binding of TraI to the *oriT* has an additional effect of reducing relaxase gene expression (Balzer *et al.*, 1994). For a more detailed review of RP4 and R751 relaxosome assembly see Zechner *et al.* (1999). The similarities between NikA and NikB to their RP4 and R751 counterparts and between the IncI1 *oriT* and IncP *oriT* suggests a

Fig. 1.2. Similarities between the *nic* regions of conjugative plasmids.

Figure shows the alignment of several *nic* regions from the *oriTs* of various conjugative plasmids. The nucleotide sequences of these *nic* regions generally fall into one of three groups. The *nic* cleavage site is indicated (▼). Identical nucleotides are indicated by a black box, conserved nucleotides by a grey box. The consensus sequence generated for each group is also shown. (Figure is reproduced from Lanka and Wilkins, 1995)

Fig. 1.2

Plasmid	Nucleotide sequence of <i>nic</i> region																					
RP4	C	T	T	C	A	C	C	T	A	T	C	C	T	G	▼	C	C	C	G	G	C	
R751	C	T	T	C	A	C	A	C	A	T	C	C	T	G	▼	C	C	C	G	C	C	
pTF-FC2	C	A	A	C	G	G	T	C	A	T	C	C	T	G	▼	T	A	T	T	G	C	
R64	A	A	T	T	G	C	A	C	A	T	C	C	T	G	▼	T	C	C	C	G	T	
pTiC58 RB	C	A	C	A	A	T	A	T	A	T	C	C	T	G	▼	C	C	A	C	C	A	
pTiC58 LB	G	C	C	A	A	T	A	T	A	T	C	C	T	G	▼	T	C	A	A	A	C	
									Y	A	T	C	C	T	G						Y	
F	T	T	T	G	C	G	T	G	G	G	G	T	G	T	▼	G	G	T	G	C	T	
P307	T	T	T	G	C	G	T	A	G	G	G	T	G	T	▼	G	G	T	G	C	T	
R100	T	T	T	G	C	G	T	A	G	T	G	T	G	T	▼	G	G	T	G	C	T	
pED208	T	T	T	G	C	G	A	C	G	G	G	T	G	T	▼	G	G	G	A	C	T	
R46	G	C	T	G	C	G	T	T	A	G	G	T	G	T	▼	A	A	T	A	G	C	A
R388	G	G	T	G	C	G	T	A	T	T	G	T	C	T	▼	A	C	A	G	C	C	
			T	G	C	G	W	N	D	K	G	T	S	T		R	K	D	R	C		
RSF1010	A	A	C	C	G	G	T	A	A	G	T	G	C	G	▼	C	C	C	T	C	C	
R1162	A	A	C	C	G	G	T	A	A	A	T	G	C	G	▼	C	C	C	T	C	C	
pTF1	T	T	A	C	T	C	T	A	A	G	T	G	C	G	▼	C	C	C	T	T	G	
pTiC58 <i>oriT</i>	C	G	A	G	T	A	T	A	A	T	T	G	C	G	▼	C	C	C	T	T	G	
pSC101	A	A	A	G	T	C	T	A	A	G	T	G	C	G	▼	C	C	C	T	G	A	
pIP501	T	G	C	G	T	A	T	A	A	G	T	G	C	G	▼	C	C	C	T	T	A	
pGO1	T	T	C	G	C	A	T	A	A	G	A	G	C	G	▼	C	C	C	T	T	A	
							T	A	A	D	W	G	C	G	C	C	C	T				

similar mechanism of relaxosome formation and cleavage reaction operating for this group of plasmids. Recent studies with R64 NikA have shown that in a similar way to RP4 TraJ, NikA recognises and binds to a repeat sequence 17-bp in length found 8-bp from the *nic* site. Such binding of NikA was also shown to induce DNA bending within the *oriT* sequence (Furuya and Komano, 1995; Furuya and Komano, 1997).

The next stage of DNA transfer involves the unwinding of the cleaved plasmid DNA strand and transfer to the recipient cell in a unidirectional manner (Howland and Wilkins, 1988; Lanka and Wilkins, 1995). Unwinding is mediated by one or more DNA helicases, some of which may be encoded by the plasmid. Studies with plasmid F have shown that the TraI protein has the additional properties of a helicase, however, the function of F TraI as the helicase responsible for unwinding the strand destined for transfer has yet to be shown genetically (Wilkins and Bates, 1997). F TraI unwinds the DNA in a 5'-3' direction relative to the strand that is bound to the enzyme (Matson and Morton, 1991; Wilkins and Lanka, 1993). From comparisons between the *in vitro* rate of unwinding by DNA helicase I (1 kb per sec at 37°C) and the rate of chromosomal DNA transfer from Hfr donor cells (0.75 kb per sec at 37°C), it is proposed that unwinding may provide the motive force for DNA transfer as long as the transfer complex is anchored to the cell envelope allowing displacement of the unwound strand. Possible candidates for anchoring of plasmid F TraI-DNA complex to the cell envelope are TraD and TraM (Silverman, 1987).

It is generally believed that nicking at the *oriT* relaxes the plasmid completely. However, examination of plasmid F after random nicking by irradiation showed that domains of the plasmid retain negative supercoiling (Wilkins and Lanka, 1993). If the DNA strand that is nicked by the relaxase is restrained from free rotation, DNA helicase activity would overwind the duplex ahead of the enzyme and thus prevent further unwinding. This problem could be overcome by a topoisomerase that prevents accumulation of

positive torsional stress. One possible candidate is DNA gyrase, which is required during bacterial conjugation as shown through a number of experiments involving naladixic acid, an inhibitor of the A subunit of the enzyme, and the use of temperature sensitive *gyr* mutants (Matson and Morton, 1991; Wilkins and Lanka, 1993).

During transfer of plasmid F only one strand is transferred to the recipient cell (Cohen *et al.*, 1968; Vapnek and Rupp, 1971). Upon examination of which λ prophage strand is transferred during Hfr matings, it was found that the plasmid F DNA strand is transmitted in the 5'-3' orientation (Willetts and Wilkins, 1984). The orientation of transfer for other plasmids has not yet been demonstrated. However, the similarities between plasmid protein-*oriT* interactions to those of plasmid F and the orientation of *oriT* and gene entry to recipient cells suggests that most enterobacterial plasmids transfer in the same orientation (Howland and Wilkins, 1988; Lanka and Wilkins, 1995).

The third stage in DNA transfer involves the recircularisation of the plasmid strand and the regeneration of a replacement strand in the donor by a rolling circle mode of replication and synthesis of a complementary strand in the recipient (Traxler and Minkley, 1987; Waters and Guiney, 1993; Wilkins and Lanka, 1993). Informative experiments involving manipulated mobilisable plasmids were able to show that the 3' terminus created at the *nic* site can be elongated continuously by a DNA polymerase present in the donor cell (see Wilkins and Bates, 1997). The model further suggests that the plasmid strand destined to be transferred is an intermediate of greater than unit length, which upon exposure of the regenerated *nic* site to the relaxase molecule linked covalently to the 5' terminus of the concatomer is cleaved (Wilkins and Bates, 1997). Experiments involving ColIb and plasmid F provide no evidence that the plasmid is transferred to the recipient as a concatomer; since sedimentation-rate studies indicate transferred DNA accumulates in the recipient cell as unit length molecules (Falkow *et al.*, 1971; Boulnois and Wilkins, 1978).

Generation of the complementary strand in the recipient is believed to involve the product of the *E. coli dnaE* gene, DNA polymerase III holoenzyme (Boulnois and Wilkins, 1978). However, due to the orientation of transfer of the plasmid strand (5'-3'), replication proceeds discontinuously and requires a series of primers synthesised *de novo* (Willetts and Wilkins, 1984). A feature found for IncI1 and IncP plasmids is that the transferred strand is escorted to the recipient with multiple copies of plasmid-specific DNA primase molecules generated by the plasmid-encoded primase genes *sog* and *traC*, respectively (Boulnois and Wilkins, 1978; Lanka and Barth, 1981; Chatfield *et al.*, 1982; Rees and Wilkins, 1990). Both the Sog and TraC proteins can be transferred along with the plasmid DNA from the donor to the recipient cell (Rees and Wilkins, 1989; Rees and Wilkins, 1990). Plasmid F relies on host-encoded replication machinery for the generation of these primers (Merryweather *et al.*, 1986; Wilkins and Lanka, 1993).

1.3.3 Plasmid-encoded exclusion

Transmissible plasmids usually carry entry- or surface-exclusion genes, the products of which reduce the ability of the host cell to receive plasmids of the same or related type by conjugation. Exclusion should not be confused with incompatibility, which is mediated from an entirely separate locus on the plasmid and prevents the co-existence of two similar plasmids within the same cell (Finlay and Paranchych, 1986). Stages of a plasmid's conjugation cycle that can be influenced by exclusion are mating-pair formation and/or DNA transfer (Rashtchian *et al.*, 1983).

One possible ecological role of exclusion may be similar to that of plasmid fertility inhibition, as a means of reducing wasteful and energy consuming conjugations. In addition, exclusion may also serve to reduce entry of plasmids that are likely to displace the resident element by the process of incompatibility (Zechner *et al.*, 1999).

The effects of exclusion were first noticed when the viability of F⁻ recipient cells was greatly reduced in comparison to F⁺ recipients when mated with an excess of Hfr donors (Alfoldi *et al.*, 1957; Skurray and Reeves, 1973). The phenomenon requires the production of pili by the donor and close physical contact between the donor and recipient cells (Skurray and Reeves, 1973). This recipient killing event was named lethal zygotis which is believed to be the result of intense mating conditions and multiple conjugation events. Lethal zygotis is associated with membrane damage and leakage of membrane constituents and cytoplasmic components, which presumably kill the cell (Zechner *et al.*, 1999). The role of exclusion in preventing lethal zygotis is highlighted by the fact that F⁻ recipient cells are not killed when mated with an excess of F⁺ donor cells. This is due to a functional exclusion system encoded by the F plasmid (Zechner *et al.*, 1999).

The most extensively studied exclusion systems (*sfx*) belong to the F-like plasmids. Two genes, *traS* and *traT*, have been shown to provide exclusion by two independent mechanisms. Both *traS* and *traT* are located within the major F transfer operon between *traG* and *traD*. TraS (TraSp) is a 16.9 kDa protein associated with the inner membrane of the donor cell and functions by blocking DNA entry either by affecting the entry system of the donor or by interfering with the initiation signal generated following successful mating-pair formation (Manning *et al.*, 1981; Finlay and Paranchych, 1986; Frost *et al.*, 1994; Firth *et al.*, 1996). TraT (TraTp) is a 26 kDa lipoprotein found in abundance in the outer membrane of the donor cell, which serves to impair the formation of stable mating aggregates, possibly by blocking access of the F-pilus tip to its putative receptor or by interfering with the F products TraG and/or TraN which are required to stabilise the mating-aggregates (Manning *et al.*, 1981; Minkley and Willetts, 1984; Dreiseikelmann, 1994; Frost *et al.*, 1994; Firth *et al.*, 1996). TraT alone is not sufficient for full expression of the exclusion system (Rashtchian *et al.*, 1983).

Extensive sequence homologies exist between the exclusion genes of plasmid F and the functionally related genes of other F-like plasmids (Finlay and Paranchych, 1986). The exclusion genes of R100 have been shown to have their own strong promoters which are not subject to regulation by the transcriptional activator TraJ (Rashtchian *et al.*, 1983). Also, downstream of *traT* a termination signal can be found. However, it is believed that these genes are only transcribed from such promoters under certain conditions that have yet to be characterised. Otherwise, they are transcribed from the main F Tra operon promoter (P_Y) and are therefore subject to control by TraJ (Ham *et al.*, 1989; Frost *et al.*, 1994; Firth *et al.*, 1996).

The first exclusion determinant studied for an I-type plasmid was that of plasmid R144. The exclusion (*exc*) locus was shown to consist of two overlapping open reading frames that encode two polypeptides, 13 kDa and 19 kDa in size (Hartskeerl *et al.*, 1986). However, it was later found that the smaller polypeptide was the result of translation being reinitiated in the reading frame of the larger polypeptide. Mutagenesis of the larger polypeptide revealed that this protein is essential for exclusion (Hartskeerl *et al.*, 1985). These polypeptides are now referred to as ExcA (19 kDa) and ExcB (13 kDa) and the exclusion locus as *eex*.

Comparisons between the physical and genetic maps of IncI1 plasmids R144, R64 and ColIb indicate that the exclusion locus of each is located at a similar corresponding position on the plasmid (Furuya and Komano, 1994; Rees *et al.*, 1987). The nucleotide sequences of the *eex* genes of R144 and R64 have been determined and extensive homology exists between them (Hartskeerl *et al.*, 1986; Furuya and Komano, 1994). The nucleotide sequence for the exclusion genes of ColIb only recently became available when the complete sequence of the plasmid was determined. These genes are 100% identical to those of plasmid R144 (EMBL: AB021078).

From cell fractionation experiments the larger polypeptide (ExcA) encoded by R144 was found bound to the inner membrane of the donor on the

periplasmic side but can also be found in a soluble form in the cytoplasm (Hartskeerl *et al.*, 1985a; Hartskeerl and Hoekstra, 1985). The second polypeptide ExcB can be found in the inner membrane. The stage of conjugation that is inhibited by I-type exclusion is unknown, although mating-aggregate formation is unaffected (Hartskeerl and Hoekstra, 1984). Measurements of the amount of radio-labelled DNA transferred from a donor carrying an I-type plasmid to an R144 excluding recipient suggest that I-type exclusion operates by inhibiting DNA transfer rather than mating-pair formation (Hartskeerl and Hoekstra, 1985).

1.4 Barriers to Horizontal DNA Transfer of Plasmids

Productive plasmid exchange not only requires a mode of transfer but also establishment of the DNA in the recipient cell. Such exchange is achieved primarily by the process of conjugation and subsequent autonomous replication in the new host. This section focuses on barriers that may be encountered by plasmids transferring by the process of conjugation. A barrier can be described as any biological process, active or passive, which can limit the level of gene exchange between organisms (Matic *et al.*, 1996).

One barrier is the physical boundary that exists between bacteria separated in nature into different microhabitats. Another barrier, which operates between different species of bacteria, is referred to as maintenance- or host-range (Matic *et al.*, 1996). The replication-maintenance range of a plasmid is subject to a number of factors, which will affect the process of establishment of the plasmid and as a consequence influence whether the plasmid persists within a given population of bacteria (Summers, 1996).

Some of these factors include, first, the ability of the autonomous plasmid to replicate at a rate that matches the growth and division rate of its host cell. If the plasmid replicates too slowly then plasmid free cells will begin to appear in the population. If the plasmid replicates too fast its copy number

will rise and the metabolic load imposed on its host will have detrimental affects. The plasmid generally overcomes this difficulty by encoding a copy number control mechanism (Summers, 1996).

Second, if the plasmid has a low copy number the requirement for an active partitioning system is essential to ensure even distribution of plasmids to daughter cells upon cell division. Third, the range of host replication enzymes that can be utilised by the plasmid will influence whether the plasmid is maintained. It should be noted that most natural plasmids commonly employ additional mechanisms that serve to maintain their presence within the bacterial population. One such mechanism involves a plasmid-encoded toxin, which is neutralised by a short lived plasmid-encoded antidote. Typically, these toxins have a long half-life. If the cell continues to divide following the loss of the plasmid the antidote becomes insufficient to neutralise the lethal effects of the toxin and the cell becomes exposed (Summers, 1996). Common examples include colicin production by colicin producing plasmids (reviewed in Luria and Suit, 1987).

The transfer range of a plasmid is generally much broader than its host range, as shown by either the recovery of plasmid-borne transposons in recipient cells after a round of conjugation or the use of shuttle vectors as reporters for the transfer range of a plasmid (Wilkins, 1995). Evidently, the potential for exchange of genetic material among a different variety of bacterial species is enormous.

Finally, another barrier to plasmid transfer is that imposed by bacterial restriction-modification systems. These are ubiquitous in nature and are therefore likely to be encountered frequently by transferring DNA. R-M systems are centrally important to this thesis and will be discussed in more detail below.

1.5 Bacterial Restriction-Modification Systems

R-M systems are commonly detected in bacteria found in every ecological niche and taxonomic group (Wilson and Murray, 1991). These systems generally consist of two enzymes with opposing intracellular activities, able to recognise and react with short specific sequences of DNA. The target sequences of these enzymes vary in length, continuity, and symmetry and whether they are unique or degenerate. The restriction endonuclease recognises its target sequence and catalyses cleavage of DNA either within the sequence or away from it. The cognate modification enzyme recognises the same target sequence to its endonuclease counterpart and catalyses the addition of a methyl group to either an adenosyl or cytosyl residue in each strand of the sequence. Methylation renders DNA refractory to cleavage by the restriction endonuclease (Wilson and Murray, 1991).

Both strands of the target sequence are methylated. However, hemimethylation, where only one strand of the sequence is methylated, is usually adequate to prevent cleavage. Such a property is especially important during replication of the bacterial chromosome where the daughter strand of the duplex will be unmethylated (Wilson and Murray, 1991).

The phenomenon of restriction and modification was first described 50 years ago when it was observed that certain strains of bacteria could restrict the growth of infecting bacterial viruses. However, an explanation of events at the molecular level was not available for another 10 years (Arber and Dussoix, 1962; Bickle and Krüger, 1993). It was these early findings which led to the view that the biological function and ecological role of R-M systems was to protect the host cell from invasion by foreign DNA, namely viruses. Over recent years, the ecological role of R-M systems has come under debate. The discovery that many bacteriophages often escape the effects of restriction, particularly if they possess few target sites or encode antirestriction systems, has led to a number of conflicting views as to what the biological role of these systems are (Matic *et al*, 1996; See section 1.5.6).

The classic R-M systems can be classified into one of three distinct groups based on co-factor requirement, subunit composition of enzymes, structure of recognition sequence and the position of cleavage by the restriction endonuclease enzyme (Yuan and Hamilton, 1984). However, it should be noted that several R-M systems identified do not meet these criteria for classification. One such example includes restriction enzymes that only recognise and cleave modified DNA (Redaschi and Bickle, 1996). It is beyond the scope of this thesis to discuss all types of R-M systems that exist, therefore, the reader is directed to reviews by Wilson and Murray (1991), Bickle and Krüger (1993) and Redaschi and Bickle (1996) for further details. The classic type I and type II R-M systems are of relevance to this thesis and therefore the following sections will focus on these two types only.

1.5.1 Type I R-M systems

Type I R-M systems are the most complex systems discovered so far (Redaschi and Bickle, 1996). Three subunits, R (restriction), M (modification) and S (specificity) encoded by the genes *hsdR*, *hsdM* and *hsdS*, form a multifunctional enzyme that can catalyse both cleavage and methylation. In addition, these enzymes have also been shown to have DNA topoisomerase activity (Yuan and Hamilton, 1984; Redaschi and Bickle, 1996). Co-factors required by type I enzymes for activity include S-adenosylmethionine (AdoMet), Mg^{2+} and ATP. Although the S subunit dictates the sequence recognised in both the methylation and restriction reactions, an enzyme can be isolated containing only S and M subunits. This product has been found to function as a modification methylase. The R subunit, on the other hand, cannot function independently of either of the other two subunits (Hadi *et al.*, 1975; Redaschi and Bickle, 1996).

Type I enzymes recognise asymmetric bipartite sequences that include an internal non-specific spacer of six to eight bp (Yuan and Hamilton, 1984). In addition, parts of the recognition sequence outside the internal spacer may also

be degenerate. The nature of the ends of the restriction fragments created by type I enzymes are elusive, although they are resistant to 5' end phosphate labelling by polynucleotide kinase (Endlich and Linn, 1985b). Cleavage by type I restriction enzymes usually occurs at variable distances of 400 to 7000 bp away from the recognition sequence (Redaschi and Bickle, 1996). Due to the nature of cleavage reactions, detection of type I R-M systems by *in vitro* digestion of DNA using bacterial cell extract is difficult. Such systems are generally detected *in vivo* by a series of molecular techniques often including restriction sensitive bacteriophages (Wilson and Murray, 1991). Other criteria used to characterise type I R-M systems include chromosomal location of *hsd* genes, enzyme subunit structure, genetic organisation, co-factor requirement, reaction mechanism and recognition sequence (Bickle and Krüger, 1993).

Type I R-M systems were classically associated with the Enterobacteriaceae family of bacteria as most systems discovered were identified in strains of *E. coli* and species of *Citrobacter*, *Salmonella* and *Klebsiella* (Redaschi and Bickle, 1996). The discovery of type I systems in *Mycoplasma pulmonis* (Dybvig and Yu, 1994) and *Haemophilus influenzae* Rd (Fleischmann *et al.*, 1995) suggests that these systems might be more prevalent in nature than original findings imply (Barcus and Murray, 1995; Chilley and Wilkins, 1995). Unfortunately, the complex characteristics of these systems hinder their identification. Type I systems can be organised into one of four genetically related families based on a series of molecular approaches involving complementation tests, DNA hybridisations, immunological tests and DNA sequence comparisons (Daniel *et al.*, 1988; Cowan *et al.*, 1989). These families include IA (K); *EcoBI*, *EcoKI*, *EcoDI*, *StySBI*, *StySPI*, *StySQI* and *StySJI*^b; IB (A), *EcoAI*, *EcoEI* and *CfrAI*; the plasmid-encoded IC (R124) family including *EcoR124* and *EcoDXXI* and finally ID; *StySBLI* (Bickle and Krüger, 1993; Titheradge *et al.*, 1996).

The *hsd* genes of each family have been characterised by DNA sequencing and DNA hybridisation studies. Homologies only exist within

families between either the *hsdM* or *hsdR* gene. No homology exists between the genes of different families apart from short sequence motifs common to DNA adenine methylases and ATP-binding proteins (Bickle and Krüger, 1993; Redaschi and Bickle, 1996). A shared feature of all these *hsd* genes is that they are arranged into two continuous transcriptional units. The *hsdM* and *hsdS* genes form an operon, which is transcribed from the *Pmod* promoter and *hsdR* is transcribed from its own promoter, *Pres*. Systems belonging to the IA and IB families have the same gene organisation with *hsdR* gene proceeding *hsdM* and *hsdS*. These genes also have the same location within the bacterial chromosome. Most of the type IC systems are encoded by plasmids where the gene organisation is similar to that of the ID family, *hsdM* and *hsdS* proceeding *hsdR* (Redaschi and Bickle, 1996; Titheradge *et al.*, 1996).

The S subunit of the type I enzyme can be divided into three functional domains; the amino terminus, which recognises the trinucleotide half of the bipartite recognition sequence; the central domain, which recognises and determines the number of nucleotides within the non-specific spacer; and the carboxy terminus, which recognises the other half of the recognition sequence. A certain degree of homology exists between the *hsdS* genes from different families, particularly the region thought to be responsible for S-mediated protein-protein interactions with M and R subunits. The *hsdS* gene contains two regions of non-homology, which are believed to be responsible for specific sequence recognition. Homology is only ever detected in these regions if the recognition sequences of the enzymes are the same (Cowan *et al.*, 1989; Redaschi and Bickle, 1996)

Rearrangement of parts of the *hsdS* genes by recombination has been demonstrated in the laboratory and is believed to occur in nature. Recombinations alter the three domains of the S subunit involved in sequence recognition, which in turn results in new sequence specificities being formed. Such a phenomenon is believed to put type I enzymes at an advantage over

other R-M systems as they can randomly change their sequence specificity (Bickle and Krüger, 1993).

1.5.2 Type I enzyme activity

Most of the following information about type I enzyme activity comes from studies with *EcoKI* and *EcoBI*. The type I three-component enzyme complex can function as either an endonuclease, a DNA methylase, an ATP hydrolase coupled to restriction or a DNA helicase (Burckhardt *et al.*, 1981a). The type I enzyme devoid of any co-factors has little affinity for DNA. However in the presence of Ado-Met the enzyme becomes activated and will bind DNA regardless of its sequence or if it is methylated. If the recognition sequence is present, then the enzyme forms a more stable recognition complex that can react to three different methylation states including fully modified, fully unmodified or hemimethylated (heteroduplex). If the recognition sequence is fully modified, the presence of ATP leads to the dissociation of the enzyme from the DNA (Burckhardt *et al.*, 1981b). In contrast, if the sequence is unmethylated then the enzyme binds to the DNA as a tightly bound complex, a process that requires the presence of ATP (Wilson and Murray, 1991). This tightly bound complex is different from that enzyme state which binds any sequence of DNA suggesting that a further conformational change in the enzyme has occurred (Bickle *et al.*, 1978). In fact, this conformational change is due to the enzyme losing bound Ado-Met in the presence of ATP.

One hypothesis to explain why these enzymes cleave away from their recognition sequences is that once the enzyme is bound tightly to its unmethylated recognition sequence it translocates DNA in both directions simultaneously. The *EcoBI* enzyme has only ever been shown to translocate DNA in one direction (Endlich and Linn, 1985a). During translocation, the enzyme remains attached to its recognition sequence and DNA is looped past in a process requiring ATP hydrolysis. In fact, ATP hydrolysis continues after cleavage has gone to completion, which in circular molecules is believed to be

associated with stalled translocating enzymes, due to the topological barriers these enzymes face if two sites are far apart and the DNA molecule has not been cleaved by other enzymes (Bickle *et al.*, 1978). Cleavage occurs some distance from the recognition sequence when two similar complexes collide (Studier and Bandyopadhyay, 1988; Wilson and Murray, 1991). Further evidence to support this hypothesis is that when cleavage reactions are synchronised, the sizes of resulting restriction fragments correspond to cleavage occurring midway between two adjacent recognition sequences (Burckhardt *et al.*, 1981a; Burckhardt *et al.*, 1981b; Wilson and Murray, 1991; Redaschi and Bickle, 1996).

Finally, if the sequence is hemimethylated then ATP induces the formation of the methylase complex using free Ado-Met as a methyl donor to methylate the other strand. In the case of type I enzymes, the base methylated is always an adenosyl residue. Not only does Ado-Met function as an allosteric effector but is also the methyl donor and only co-factor required in the methylation reaction. Once methylation is completed, the enzyme dissociates from DNA in a process that does not require ATP hydrolysis. Interestingly, the methyltransferases of the *EcoKI* enzyme are the only prokaryotic enzymes to show a strong preference for hemimethylated DNA. Such enzymes are often referred to as maintenance methylases as their presence ensures methylation following DNA replication. In contrast, the *EcoAI* methyltransferase methylates DNA *de novo* (Kelleher *et al.*, 1991).

1.5.3 Regulation of expression of type I R-M genes

There are circumstances when regulation of restriction and modification activities within the bacterial cell is essential. These include either during establishment of an R-M system in a new strain or during changes in the physiology of the cell, which might lead to undermodification of the host genome. The latter situation is known to arise during cell starvation (Bickle and Krüger, 1993; Redaschi and Bickle, 1996).

As already mentioned, the genetic organisation of the type I *hsd* genes is such that *hsdM* and *hsdS* form an operon and *hsdR* is expressed from a separate promoter. However, examination of specific *hsd_k-lacZ* operon fusions under conditions similar to those described above showed no altered expression patterns at the transcriptional level for these genes when transferred to a naive cell (Prakash-Cheng *et al.*, 1993).

Despite there being no obvious control of these *hsd* genes at the transcriptional level, transmission of novel type I R-M genes is surprisingly efficient (Suri and Bickle, 1985; Redaschi and Bickle, 1996). Such findings suggest that during transfer of the *hsd* genes, the modification methyltransferase is functional in the recipient before the restriction endonuclease. As early as 1965, Glover and Coulson described experiments which showed that the restriction phenotype of the *EcoKI* R-M system is delayed following conjugative transfer. These experiments involved the conjugative transfer of the *EcoKI* R-M genes from an *E. coli* K-12 Hfr donor to an *E. coli* B F⁻ recipient. In other experiments, the appearance of the restriction phenotype was delayed in the recipient for 15 generations whilst modification was observed immediately (Prakash-Cheng and Ryu, 1993). Since there is no evidence for transcriptional control of expression of *hsd_k* genes, these findings suggest that control must be at the translational or post-translational level (Prakash-Cheng *et al.*, 1993).

A spontaneous mutant of *E. coli* C has been isolated, which is sensitive to the receipt of *hsd_k* genes. The mutant gene responsible, *hsdC*, must somehow mediate control of type I R-M systems in *E. coli* K-12 (Prakash-Cheng *et al.*, 1993). It was postulated that the HsdC peptide might be required to influence assembly of the *EcoKI* enzyme (Dryden *et al.*, 1997). Analyses of the pathway responsible for the assembly of *EcoKI* suggests that at a particular stage the HsdR subunit may become susceptible to proteases (Dryden *et al.*, 1997). In a screening of strains of *E. coli* K-12 that are deficient in proteases, Makovets *et al.* (1998) identified two possible candidates ClpX and ClpP that make up the *E.*

coli ClpXP protease. At least one of these products, ClpX or ClpP is required in the recipient for the efficient transmission of *Eco*KI. These results imply that the ClpXP protease is involved in modulating restriction activity and that delayed restriction during conjugative transfer occurs at the post-translational level (Makovets *et al.*, 1998).

Over recent years much emphasis has been placed on understanding the complexities of transcriptional and translational control in the regulation of protein availability within the cell. The roles of certain proteases in the regulation of gene expression in both prokaryotic and eukaryotic organisms is becoming increasingly more evident. It has long been known that proteases, present within all cells, degrade proteins that are unstable or abnormal. However, during regulatory proteolysis, specific proteases have the ability to adjust and regulate the amounts of available protein within the cell by targeting specific proteins for degradation, in particular, regulatory proteins and key metabolites. Therefore, in addition to control at the level of transcription and translation, regulatory proteolysis plays an essential role in allowing the cell to respond to external stresses and developmental signals. Cleavage by the protease renders the protein biologically inactive and may also result in exposing the protein to further degradation by other cellular proteases. Cells that are mutants in such initiating proteases tend to accumulate an abundance of the target protein (Gottesman and Maurizi, 1992; Gottesman, 1996).

Proteins susceptible to degradation by proteases generally fall into three classes. The first class includes those proteins that are either unstable, damaged or abnormal due to mutations within the corresponding gene. External stresses that lead to the induction of a particular stress response within the cell often result in the accumulation of abnormal or damaged proteins. Therefore, it is not surprising that a large proportion of genes induced during such stress responses encode proteases. This finding is particularly true for the heat shock response (Gottesman, 1996).

The second class includes proteins that are required within the cell for limited times and are quite often involved in complex regulatory pathways. These proteins have often been referred to in the literature as 'timing proteins' (Gottesman, 1996).

The final class includes proteins that form part of a multi-complex structure, where under certain conditions rapid degradation of the uncomplexed protein may occur (Gottesman, 1996). This is believed to be the case during assembly of the type I restriction enzyme, *EcoKI* (Makovets *et al.*, 1998). It is envisaged that under normal conditions the HsdM subunit of the type I restriction enzyme protects the HsdR subunit from attack by proteases. However, one idea is that during transmission of type I R-M genes, the ClpXP protease competes with the HsdM subunit for interaction with HsdR, which delays the production of the active endonuclease complex (Makovets *et al.*, 1998). The effect of a *clpX* mutation on the transmission of the *EcoKI hsd* genes by conjugation is more extreme than the effect of a *clpP* mutation. Such results imply that ClpX may be involved in initially targeting the HsdR component, which in turn makes it a substrate for further degradation by the ClpXP protease.

ClpP can associate with two components to form a protease. These components include either ClpA or ClpX as molecular chaperones. Both ClpA and ClpX form the ATPase subunit of the complex, which is essential, as most proteolytic reactions require ATP hydrolysis. When complexed with the protease, these ATPase components regulate protein degradation by binding, unfolding and translocating substrates into the proteolytic chamber of ClpP. Following proteolysis, the products of the degradation reaction diffuse out of the ClpXP or ClpAP complex (Wickner and Maurizi, 1999). The genes responsible for ClpX and ClpP lie within the same operon, which is regulated by a heat shock promoter. In addition to recent findings that the ClpXP protease is involved in degrading the HsdR component of *EcoKI* and *EcoAI* restriction endonuclease enzymes, the protease is also responsible for

degradation of λ O protein, P1 addiction protein, Mu repressors, and RpoS, which is a sigma factor required for gene expression during stationary phase (Gottesman, 1996).

1.5.4 Type II R-M systems

The type II R-M systems are the least complex and have been discovered in virtually every class of bacteria over 150 of which were isolated in *E. coli* and *Salmonella* strains (Roberts and Macelis, 1994; Redaschi and Bickle, 1996).

Type II systems typically comprise separate restriction and modification enzymes that act independently of each other. The only co-factors required are Mg^{2+} ions by the endonucleases and Ado-Met by the methyltransferases. The ability of these enzymes either to cleave or to methylate DNA at fixed positions within their recognition sequences has led to their exploitation in many branches of molecular biology and to the wide screening of bacteria for new systems. Several thousand type II R-M systems have been discovered, but only a few have been fully characterised. It should be noted that most new specificities discovered are duplicates of other systems and are referred to as isoschizomers.

The recognition sequences of type II enzymes generally consist of between four to eight specific nucleotides, quite often the sequences are symmetric and may be continuous or interrupted (Wilson and Murray, 1991). Most endonucleases act as homodimers where each subunit interacts with one half of the recognition sequence and cleavage of both DNA strands is coordinated. In contrast, the methyltransferases normally act as monomers, which transfer one methyl group to one strand per DNA binding event (Newman *et al.*, 1981). The *EcoRI* endonuclease is commonly isolated as a homodimer. Examination of the *EcoRI*-DNA complex indicated that the active sites of the enzyme are placed directly against the phosphodiester bonds that are hydrolysed during the cleavage reaction (McClarín *et al.*, 1986). Methylation at these sites presumably prevents such a complex forming due to

steric hindrance. Cleavage of DNA by type II enzymes can produce either 5' or 3' overhangs or blunt ends.

Plasmids transferring by the process of enterobacterial conjugation do so in single-stranded DNA form. Whether type II restriction enzymes are able to cleave single-stranded DNA is still debatable. Evidence has been reported to suggest that some type II restriction enzymes cleave single-stranded DNA, although whether such enzymes act on single-stranded DNA or temporary duplex structures that form in the DNA strand is unknown (Nishigaki *et al.*, 1985). In addition, it has been found that *EcoRI* is capable of cleaving ssDNA that is immobilised on cellulose by ligation. However, the cleavage reaction is less efficient than that which cleaves dsDNA (Bischofberger *et al.*, 1987).

Many type II R-M systems have been cloned and the nucleotide sequence of the structural genes determined. Surprisingly, no homologies exist between the amino acid sequences of cognate restriction and modification enzymes, which suggests that they must have evolved independently of each other (Chandrasegaran and Smith, 1988; Wilson, 1991; Redaschi and Bickle, 1996). Furthermore, it is rare for any homology to exist between restriction endonucleases and their isoschizomers, despite the two enzymes recognising the same target sequences (Wilson and Murray, 1991; Bickle and Krüger, 1993). It is envisaged that the methyltransferase enzyme evolved first, methylating the bacterial genome, upon which the restriction endonuclease could then invade (Bickle and Krüger, 1993).

The endonuclease and methyltransferase of a cognate type II system possess separate target recognition domains (TRD) and as a consequence, each enzyme employs a different strategy for recognition of their target sequences. Base analogue substitution experiments were able to show that each enzyme recognises a different base in the target sequence that is common to both enzymes (Newman *et al.*, 1981). Such a feature of type II R-M systems probably limits these enzymes from developing new specificities during their evolution (Wilson and Murray, 1991).

A subclass, classified as type II systems due to co-factor requirement consists of those restriction enzymes that recognise and cleave at a precise distance one-20 base pairs from their recognition sequence. These are referred to as type IIS systems, where the S stands for shifted cleavage (Szybalski *et al.*, 1991).

1.5.5 Regulation of type II R-M systems

For reasons described in section 1.5.3, expression of type II restriction and modification activities need to be regulated. Type II R-M systems vary in their genetic organisation and their gene expression patterns. Consequently, it is difficult to generalise; therefore, the mechanisms described can only be used as examples, which may apply to some known systems.

One way in which type II R-M reactions can be modulated is by a passive mechanism that relies on the restriction enzyme functioning in a multimeric form. Most type II restriction enzymes function as homodimers and cleavage requires the accumulation of R subunits. In contrast, most of the corresponding methyltransferases function as monomers. Consequently, there is a lag between the two opposing enzyme activities. This mechanism of control can be applied to the *EcoRI* system (Greene *et al.*, 1981).

Evidence that the restriction genes of some type II R-M systems can be cloned without the corresponding modification gene suggests that *E. coli* is able to repair a certain amount of DNA damage created by these enzymes. The restriction genes that have been successfully cloned include those of *PaeR7I*, *TaqI*, *AvaII*, *HaeII*, *HinFI*, *PstI* and *XbaI*. However, colonies of strains differ from others in that they appear translucent and are susceptible to high frequency loss of the plasmid which carries the restriction gene. In addition, synthesis of the restriction enzyme is greatly reduced compared to that in a modified strain (Lunnen *et al.*, 1988; Redaschi and Bickle, 1996).

Regulation of restriction and modification activities by some type II systems involves the product of another tightly linked ORF (Tao *et al.*, 1991).

The predicted amino acid sequences of these C (controller) proteins have potential DNA binding properties such that they may function as transcriptional repressors or activators. Mutations in the C gene of the *PvuII* system results in a restriction-less phenotype, which in turn can be complemented *in trans* by the wild-type gene. Disruption of the C gene of *BamHI* results in decreased restriction and increased methylase activity (Bickle and Krüger, 1993). Nine R-M systems that have been sequenced have one of these ORFs; five share sequence homology and are able to complement each other *in trans*. Such findings suggest that these C genes may have evolved as a family independently of RM genes as no homology is found between them (Redaschi and Bickle, 1996).

1.5.6 Ecological roles of bacterial R-M systems

Early studies with R-M systems and the finding that they are widespread led to the model that their main ecological role was to provide bacteria with protection against infection by bacteriophages. Such a concept is now referred to as the 'cellular defence' hypothesis. There are situations in nature where the carriage of a novel R-M system is seen as advantageous to the host cell. One of these situations is when bacteria specifying a R-M system invade microhabitats containing phages that can infect the strain (Korona and Levin, 1993; Bickle and Krüger, 1993; Korona *et al.*, 1993).

A problem with the 'cellular defence' hypothesis is that restriction is not an absolute barrier to transferring DNA and that a fraction will escape destruction and acquire methylation. Hence the R-M barrier is only transient and consequently, the 'cellular defence' is not the only role that can be assigned to R-M systems (Korona and Levin, 1993). A further short coming in the 'cellular defence' hypothesis is that it does not explain restriction enzymes that have long recognition sequences of 8 bp, which are not likely to be present in the genomes of many bacterial viruses (Naito *et al.*, 1995). Also, the 'cellular defence' hypothesis does not explain the extreme diversity and specificity of

sequence recognition (O'Neill *et al.*, 1997). Hence, restriction-modification is regarded as being able to slow the rate at which DNA transmission occurs rather than stop it (Matic *et al.*, 1996). Bacteria probably develop more long term protection from phage infection by spontaneous mutation of genes encoding phage receptor sites (Korona *et al.*, 1993).

Another potential role assigned to bacterial R-M systems is that restriction potentiates genetic recombination in natural populations of bacteria by cleaving entrant DNA molecules into fragments that have recombinogenic properties (Wilson and Murray, 1991; Korona and Levin, 1993). Most DNA fragments generated by restriction enzyme cleavage *in vivo* will be subject to degradation by the RecBCD nuclease [exonuclease V] (Dixon and Kowalczycki, 1993). However, recombination is possible if these fragments contain chi sites (χ), which are short specific 8-bp sequences able to stimulate the recombination properties of the RecBCD enzyme. Furthermore, experiments have been conducted to show that recombination of fragments generated by restriction can take place in the absence of RecBCD, relying completely on the recombination pathways of either Red of phage λ or RecE of the cryptic prophage *rac*. However, these results are not indicative of events in nature.

Recombination during interspecific DNA transmission is unlikely as the extent and degree of sequence divergence that exists will not only delimit the activity of enzymes that control the initial stage of recombination, strand separation, but recombination will also be inhibited by the mismatch-repair system (Matic *et al.*, 1996). Nevertheless, 0.1-1% of natural populations of *E. coli* carry mutations in their mismatch-repair systems. Evidence to suggest that restriction-modification may influence recombination in *E. coli* is from the mosaic patterns observed in the nucleotide sequences of *trp* operons of naturally occurring strains of *Escherichia coli* (ECOR-collection) (Milkman, 1997; Matic *et al.*, 1996).

A more recent role, originally assigned to type II R-M systems, is that these paired R-M genes act as selfish symbionts, which serve to force their

maintenance on the host cell. Plasmids carrying type II R-M systems have increased stability in *E. coli* and attempts to displace the resident plasmid by another that does not carry the same R-M or M genes results in host cell death (Naito *et al.*, 1995; Kulakauskas *et al.*, 1995). Such data are consistent with the 'toxin-antidote' hypothesis, where the restriction enzyme acts as the toxin and the methylase as the antidote. Any cells that continue to divide after losing the plasmid are unable to protect themselves from attack by remaining restriction enzyme (Kusano *et al.*, 1995). It is envisaged that the ability of these plasmid-borne type II R-M systems to act as selfish symbionts probably contributed to their evolution as gene pairs (Naito *et al.*, 1995).

Mutations in both *hsdR* and *hsdM* genes of type I R-M systems are easily isolated without detrimental effects on the host cell. Such findings suggest that the 'toxin-antidote' hypothesis applied to type II R-M systems cannot be applied to type I systems. In addition, the ability of the host cell to modulate restriction activity by type I enzymes also suggests that this particular type of R-M system is not selfish in nature (O'Neill *et al.*, 1997; section 1.5.3).

1.6 ColIb Leading Region Genes and their Functions

As described in section 1.3.2, conjugative transfer of a plasmid starts at *oriT* by specific cleavage of the *nic* site by the relaxase enzyme. A unique DNA strand is released called the T-strand, which is transferred to the recipient cell with a 5'-3' polarity (Wilkins and Lanka, 1993). The *oriT* is orientated such that the transfer genes of the plasmid enter the recipient cell last (Howland and Wilkins, 1988). Whether delayed transfer of the Tra genes to the recipient cell reflects the need for their continued expression in the donor cell or their delayed expression in the recipient is unknown. However, one possible explanation is that their delayed transfer gives priority to entry of genes that are expressed early in the recipient and promote establishment of the plasmid in the newly infected cell. Such genes are located within the leading region of

the plasmid, which can be defined as the first segment of the plasmid to enter the recipient cell during conjugation. Leading region genes that have been identified on ColIb include *ssb*, *psiB* and *ardA*, the functions of which are described in the following sections.

The leading regions of plasmid F and ColIb are believed to be inessential for conjugation as disruption of known genes within this sector result in only a slight decrease in conjugation efficiency under laboratory conditions (Loh *et al.*, 1989; Jones *et al.*, 1992). The length of the ColIb leading region is unknown. However, it is estimated to extend 15 kb from the *nic* site of the plasmid, although these numbers are purely arbitrary (Rees *et al.*, 1987).

1.6.1 Single-stranded DNA-binding protein (SSB)

Single-stranded DNA-binding protein (SSB) found in *E. coli* is known to play an important role in DNA replication, recombination and repair, in particular mismatch repair (Chase *et al.*, 1983; Golub and Low, 1986). SSB proteins have a high affinity for single-stranded DNA and bind predominantly in a tetrameric form with no sequence specificity (Howland *et al.*, 1989). *In vitro* studies with SSB have shown that some of the roles of the protein include i) preventing reassociation of unwound DNA, ii) protecting ssDNA from nucleases, iii) enhancing the processivity of DNA polymerases, iv) stabilisation and organisation of replication origins, and v) promoting binding of DNA polymerase to template DNA (reviewed in Meyer and Laine, 1990). Furthermore, in the presence of SSB, the helix-coil transition temperature of dsDNA is lowered enabling the DNA molecule to melt more easily (Williams *et al.*, 1983).

ColIb and a number of other enterobacterial plasmids including F, encode SSB proteins which are very closely related in their amino acid composition to the *E. coli* counterpart (Chase *et al.*, 1983; Golub and Low, 1985; Howland *et al.*, 1989). 19 plasmids from 12 incompatibility groups are known to carry *ssb* genes that can complement defects caused by an *ssb-1* mutation in

the corresponding *E. coli* gene (Golub and Low, 1985; Golub and Low, 1986). In addition, the *ssb* genes of these plasmids were found to share extensive sequence homology with the *ssb* gene of plasmid F. These *ssb* genes were found to be co-ordinately regulated with the plasmids fertility genes (Golub and Low, 1986). Such findings suggest that the role of SSB during conjugation is possibly to prevent drainage of the cellular counterpart during replication of the complementary plasmid strand. However, mutations in plasmid *ssb* genes have very little effect on the ability of the plasmid to form transconjugants (Golub and Low, 1986; Howland *et al.*, 1989).

The ColIb *ssb* gene maps in the leading region of the plasmid 11 kb from the *oriT* (see Fig. 1.1). The gene was originally isolated by its ability to rescue the temperature and UV sensitivity of *E. coli* chromosomal *ssb-1* mutant. In addition, the gene has been shown to be homologous to the F *ssb* gene by cross-hybridisation studies and sequence comparisons. Both share 84% sequence similarity at the nucleotide sequence level and 83% at the predicted amino acid level. The amino acid sequence of ColIb SSB contains two signature motifs characteristic of single-stranded DNA binding proteins. One of these signatures is the DNA binding motif that has been characterised for the *E. coli* SSB protein, which is perfectly conserved in ColIb SSB (Howland *et al.*, 1989).

1.6.2 Plasmid SOS inhibition (PsiB)

The product of the ColIb *psiB* gene acts during conjugation to prevent induction of the bacterial *E. coli* SOS response in recipient cells as a function of its intracellular concentration (Jones *et al.*, 1992). The SOS response is triggered by certain conditions that either cause damage to DNA or interfere with the process of DNA replication. Long-lived regions of single-stranded DNA that are created during such damage or those that exist near a stalled replication fork result in activation of RecA. Activated RecA in the presence of co-factors such as ATP cleaves the LexA transcriptional repressor and a number of phage

repressors, which leads to the induction of a diverse set of specialised genes known collectively as the SOS regulon (Walker, 1987).

One possibility is that the SOS response is induced by ssDNA transferring in conjugation or short tracts of ssDNA that are exposed when the second strand of the plasmid is generated in the recipient during discontinuous lagging strand DNA synthesis (see section 1.3.2). Plasmid *psiB* genes, therefore, may have evolved to permit conjugative transfer in single-stranded form without the additional complication of inducing the SOS response and host physiology (Jones *et al.*, 1992). The PsiB protein of IncFII plasmid R6-5 has been shown to prevent induction of the SOS response by inhibiting activation of the RecA co-protease rather than preventing expression of genes associated with the SOS regulon (Bailone *et al.*, 1988).

The ColIb *psiB* gene was identified through its cross-hybridisation to the *psiB* gene detected on R6-5 (Bagdasarian *et al.*, 1986; Jones *et al.*, 1992). ColIb PsiB shares 85.4% and 84.6% identity to the corresponding proteins of F and R6-5, respectively (Jones *et al.*, 1992). Through cross-hybridisation studies conducted by Chilly and Wilkins (1995) it was found that plasmids carrying *ssb* genes more often than not also carry a *psiB* gene. Transposon mutagenesis of plasmid *ssb* genes leads to the generation of a strong Psi⁺ phenotype, which is presumably the result of the polar effects on transcription caused by the antibiotic resistance gene promoter present in the Tn insertion (Dutreix *et al.*, 1988; Jones *et al.*, 1992). Such findings suggest that expression and location of both *ssb* and *psiB* are linked and that no transcriptional terminator exists between the two (Jones *et al.*, 1992).

1.6.3 Alleviation of restriction of DNA (ArdA)

The product of the ColIb *ardA* gene acts as an anti-restriction protein, which is active in alleviating restriction by three families of type I restriction enzymes (Kotova *et al.*, 1988; Delver *et al.*, 1991; Read *et al.*, 1992). The ArdA protein alleviates restriction mediated by type I restriction enzymes, possibly by

interfering with the functional integrity of the R-M enzyme complex. The ArdA protein of ColIb is unable to alleviate the effects of restriction mediated by type II and type III enzymes (Read *et al.*, 1992; Belogurov and Delver, 1995).

The *ardA* gene like *ssb* and *psiB* is also known to have a counterpart in the leading regions of plasmids belonging to other Inc groups. Homologues have been identified in the leading region of enterobacterial plasmids belonging to the IncFV group, the B-I1-K set of the I complex and the IncN group. All *ardA*⁺ plasmids identified so far are enterobacterial plasmids, which may be significant as type I restriction enzymes are classically associated with members of the Enterobacteriaceae (Chilley and Wilkins, 1995). Furthermore, these *ardA* genes also have the same transcriptional orientation within their respective leading regions (Delver *et al.*, 1991; Read *et al.*, 1992; Belogurov *et al.*, 1993; Chilley and Wilkins, 1995), although ColIb does not contain the equivalent *ardB* antirestriction gene or the *ardK* and *ardR* regulators that are found in the leading region of IncN plasmid pKM101 (Belogurov *et al.*, 1993).

The ArdA protein operates specifically during conjugation to protect the immigrant plasmid from destruction by type I restriction enzymes that might be present in the recipient cell. The function of ArdA and its interaction with type I restriction enzymes will be discussed further in section 1.7.

1.6.4 Zygotic induction of leading region genes

Detectable levels of expression of both *ssb* and *psiB* is only observed when such genes are carried by plasmids that are derepressed for transfer functions or if the cell is exposed to SOS-inducing treatment (Golub and Low, 1986; Jones *et al.*, 1992). Such findings suggest that these genes are under the control of the plasmid's fertility inhibition system (Howland *et al.*, 1989).

When ColIb is derepressed for transfer functions through mutation of the Fin system, *ssb* and *psiB* are expressed at low levels in an established strain. Transcriptional activity of the leading region of naturally derepressed plasmid F was also only detected at low levels, which suggests that genes contained

within this region are only fully expressed under unique conditions (Cram *et al.*, 1984). During the process of conjugation, the products of these *ColIb* genes accumulate in a transient burst in the conjugatively infected cell. Such enhanced expression was originally demonstrated by a technique involving the insertion of a promoterless *lacZ* operon fusion into *psiB* and *ssb* on *ColIbdrd*. Conjugation experiments were performed using the specially constructed *ColIb::lacZ* plasmids and *ssb* or *psiB* gene product was quantified by measuring the level of β -galactosidase specific activity. Shortly after initiation of conjugation a 10-20 fold burst of expression of both *ssb* and *psiB* was observed. Rapid expression starts within 10 minutes and begins to plateau out at 20-30 minutes (Jones *et al.*, 1992).

To determine whether enhanced expression of *ssb* and *psiB* occurs in the recipient or donor cell a selective phage lysis technique was used. During a series of conjugation experiments using either T6 bacteriophage sensitive or resistant donor or recipient cells, it was found that accumulation of β -galactosidase could be found only in recipient cells. Such results clearly show that both *psiB* and *ssb* genes are induced upon entry to the recipient cell during conjugation. It was these findings which led to the application of the term zygotic induction (Jones *et al.*, 1992). However, it should be noted that the term zygotic induction was originally used to describe the induction of λ prophage during Hfr mating experiments, which resulted in lysis of F⁻ recipient cells (Hayes, 1968).

The *ardA* gene has not yet been shown to be zygotically induced during conjugation. However, indications that *ardA* expression is also enhanced during conjugation comes from the findings that optimal alleviation of type I restriction occurs during the process of conjugative transfer of *ColIbdrd* *ArdA*⁺ as opposed to transfer by transformation (Read *et al.*, 1992).

1.6.5 Hypotheses for zygotic induction

The potential roles of *ssb*, *psiB* and *ardA* in the establishment of ColIb shortly after conjugative transfer are understood. However, one intriguing question that remains to be answered is how are these leading region genes regulated so that they are expressed early and transiently in the recipient cell. Several hypotheses have been generated, some of which have been examined experimentally (Jones *et al.*, 1992; Roscoe, 1996). The following two lists (1.6.5.1, 1.6.5.2) summarise these hypotheses.

1.6.5.1 Possible modes of positive regulation

i) Activation is by transfer-associated plasmid topology changes.

Expression of genes from certain types of promoter can be initiated during changes in localised topology of DNA, such that the angle of the -35 and -10 sites of the promoter form an optimal alignment that is recognised by RNA polymerase (Wang and Syvanen, 1992; Roongta, 1990). During the process of conjugation, transfer of DNA to the recipient cell is associated with significant changes in the topology of the plasmid. However, treatment of ColIb containing cells with an inhibitor of DNA gyrase, coumermycin-A1, which reduces the negative supercoiling density of the plasmid does not result in any significant enhancement of *ssb::lacZ* or *psiB::lacZ* expression (Roscoe, 1996).

ii) Activation occurs when genes are transferred in single-stranded form.

There is evidence that during enterobacterial conjugation the plasmid DNA is transferred to the recipient cell in single-stranded form (see section 1.3.2). The usual action of RNA polymerase on single-stranded DNA is the production of primers for DNA replication. However, it is envisaged that transcription of leading region genes might be initiated on the transferring single plasmid-strand during conjugation.

iii) Activation is by a plasmid- or chromosomal-encoded positive regulator.

Leading region gene expression may be controlled by a regulator that is induced only during the process of conjugation. Subjection of *E. coli* to any stimuli that alter the physiological state of the cell often result in the induction of a co-ordinated set of genes, known collectively as a stress-response. If conjugation induces any such responses, the regulators or gene products could act to control leading region gene expression. In *E. coli* the heat shock stress response can be induced by stimuli other than heat. However, conjugative transfer of ColIb does not stimulate this stress response as shown by its inability to induce a heat shock reporter construct in recipient cells during a 60 minute conjugation (Roscoe, 1996).

1.6.5.2 Possible modes of negative regulation

i) Transcription is regulated by a *trans*-acting repressor that accumulates in the newly infected recipient cell.

It is proposed that such a repressor inhibits transcription of leading region genes in the newly infected cell. This hypothesis was tested by transferring a ColIb plasmid that carries a promoterless *lacZ* operon fusion in *psiB* to a recipient cell harbouring a ColIb plasmid that is defective in encoding a functional exclusion system. However, the hypothesis was ruled out because under such conditions zygotic induction of *psiB* was still observed (Roscoe, 1996).

1.7 Restriction-Avoidance Strategies Employed by Plasmids

It is known that bacteriophages have evolved a number of strategies for avoiding the effects of restriction during their life cycles. Examples of phage anti-restriction mechanisms include i) inhibition of restriction enzymes, ii) virus-coded DNA modifying enzymes, iii) stimulation of host modification functions, iv) incorporation of unusual nucleotides into phage DNA, v) coinjection of phage restriction inhibitory head proteins with DNA, vi) destruction of restriction endonuclease co-factors and vii) counterselection

against restriction sites in phage (reviewed in Krüger and Bickle, 1983 and Bickle and Krüger, 1993).

Plasmids that have been found to encode restriction-avoidance mechanisms include members of the IncI and IncF group of enterobacterial plasmids. Restriction-avoidance is not a feature common to all transferring plasmids. For example, the IncP plasmids are extremely sensitive to restriction. The lack of a restriction-evasion mechanism may explain why IncP replicons are deficient in restriction enzyme recognition sites. In a similar way to some phages, these plasmids have presumably been exposed during their evolution to strong selection for the elimination of restriction enzyme target sequences by mutation and selection (Wilkins *et al.*, 1996).

Plasmid ColIb is remarkably resistant to destruction by type I and type II restriction enzymes during its transfer by conjugation, even though the plasmid carries multiple enzyme target sites (Read *et al.*, 1992). The resistance of ColIb to type I restriction enzyme involves a specialised plasmid-encoded antirestriction mechanism known as ArdA (see section 1.6.3). How ArdA specifically acts to alleviate restriction by type I enzymes is not clear. However, it seems unlikely that the protein interacts with the recognition sites of the enzymes as the protein is highly acidic and therefore unlikely to bind to DNA (Delver *et al.*, 1991). It seems more likely that ArdA interferes with the type I enzyme complex in a way that prevents restriction activity and not methylation (Belogurov and Delver, 1995). The ArdA protein of ColIb and a number of other plasmids share a common amino acid sequence motif also found in the T7 0.3 anti-restriction protein. The motif is only nine amino acid residues in length and is believed to be the interaction site for anti-restriction proteins and restriction endonucleases (Belogurov *et al.*, 1993). The motif also shares similarity with a conserved sequence known as the Argos repeat, which is found in the DNA sequence specificity (S) polypeptide of type I enzymes. These motifs in anti-restriction proteins might allow the protein to compete

with the S subunit, which plays a role in the assembly of the type I enzyme complex (Belogurov and Delver, 1995).

ColIb encodes another restriction-avoidance mechanism that acts independently of *ardA*. The process becomes manifest after several minutes of conjugation and alleviates restriction by both type I and type II enzymes during second or subsequent rounds of transfer (Read *et al.*, 1992). The molecular basis of this second evasion process has yet to be elucidated.

1.8 Aims and Direction of this Study

The initial aim of my studies was to further the understanding of the restriction-avoidance mechanisms encoded by ColIb which are active against type I and type II restriction enzymes and operates independently of *ardA*. The first hypothesis tested is that transfer of multiple copies of ColIb during conjugation overwhelms the restriction system in the recipient cell by substrate saturation (Read *et al.*, 1992). A second hypothesis is that multiple transfers induce a change in the physiology of the recipient cell, which in turn leads to a transient breakdown of the restriction barrier. A third possibility is that ColIb possesses a gene that encodes an anti-restriction function active against both classes of restriction enzyme.

An experimental priority in the testing of these hypotheses was to establish whether ColIb could confer cross-protection on a restriction-sensitive plasmid when co-transferred together. The first two hypotheses predict alleviation of restriction by ColIb *in trans*. Consequently, a restriction-sensitive plasmid that is co-transferred with ColIb should acquire some protection from destruction by restriction. For this purpose a genetic system was developed, which relied on measuring ColIb-mediated rescue of plasmid R751 (IncP β) from destruction by *EcoKI* and *EcoRI*. One outcome of these experiments was the discovery of a *cis*-acting restriction protection phenomenon.

An important experimental tool was the construction of a specific mutant of Collb that was deleted of 14 bp from the *nic* region. The resulting plasmid was completely defective in self-transfer by conjugation, yet proficient in forming mating-pairs and in expressing a full complement of plasmid proteins. This mutant has proved invaluable in various genetic investigations.

Also examined was the classic ArdA system. The underlying question was whether *ardA* is transiently overexpressed (zygotically induced) in the recipient cell during conjugation like other leading region genes such as *psiB* and *ssb*. This question was addressed using the R751 system and the Collb *nic* mutant to measure ArdA-mediated alleviation of restriction during different stages of conjugation. Data obtained suggested that *ardA* is subject to zygotic induction. To investigate potential regulatory mechanisms governing enhanced expression of leading region genes during conjugation, the nucleotide sequence of 11.7 kb of the leading region was determined in a collaboration with Dr. Steven Bates working in our laboratory. Such an approach was invaluable in the generation of a model of zygotic induction and the development of the concept that plasmids carry 'early genes' expressed by transcription of the incoming single-stranded DNA.

Chapter Two

Materials and Methods

2.1 Plasmids and Bacterial Strains

The genotypes, sources and relevant restriction-modification phenotypes of bacterial strains and plasmids used in this work are described in Tables 2.1.1 and 2.1.2. Bacterial strains used were all *Escherichia coli* K-12 derivatives. Plasmids constructed during this work are described in Table 2.1.3. For the purpose of publication, some recombinant plasmids were assigned alternate names to those used in this thesis. Where this has occurred, the alternative name has been indicated in Table 2.1.3. Plasmid pHP45 Ω contains an omega (Ω) intersposon; this 2.0 kb fragment contains a streptomycin/spectomycin resistance gene flanked by inverted repeats carrying transcription and translation termination signals and polylinkers (Prentki and Krisch, 1984).

Spontaneous naladixic acid and rifampicin variants of strains GI65, NM654 and NM816 are designated with the suffix N or R. All were used in conjugation experiments and are not shown in Table 2.1.1. These resistant strains were isolated by plating 1.0 ml of overnight culture, concentrated to 0.1 ml, onto nutrient agar plates containing either naladixic acid or rifampicin at 25 $\mu\text{g ml}^{-1}$.

Plasmids R751 and pHP45 Ω and bacterial strain DL307 were gifts courtesy of Peter Thorsted, Chris Thomas and David Leach, respectively. Bacterial strains NM654, NM816 and NM840 were kindly provided by Noreen Murray.

Table. 2.1.1 *Escherichia coli* Strains

Strain	Description or genotype	Source or reference
DH5 α	F f80 Δ LacZ Δ M15 <i>recA1 endA1</i> <i>gyrA96 thi-1 hsdR17</i> (<i>r_{KI}⁻ m_{KI}⁺</i>) <i>supE44 relA1</i> <i>deoR</i> Δ (<i>lacZYA-argF</i>) U169	Hanahan (1983)
DL307 (FS1585)	C600 <i>recD1009 Tet^R</i>	Stahl <i>et al.</i> (1986)
GI65 (C600)	<i>thr-1 leu-6 thi-1 lacY1 fhuA21</i> (<i>r_{KI}⁺ m_{KI}⁺</i>)	Arber and Dussoix (1962)
JO8 (BW103)	<i>dna⁺ leu deoB rpsL cir recA1</i>	Merryweather <i>et al.</i> (1986)
NM654	C600 Δ <i>hsdRM</i> (<i>r_{KI}⁻ m_{KI}⁻</i>)	Makovets <i>et al.</i> (1998)
NM816	C600 Δ <i>hsdR</i> (<i>r_{KI}⁻ m_{KI}⁺</i>)	N. Murray
NM840	C600 <i>gyrA96</i> Δ <i>hsdRM</i> (<i>r_{KI}⁻ m_{KI}⁻</i>)	Makovets <i>et al.</i> (1998)
W3110	Prototrophic (<i>r_{KI}⁺ m_{KI}⁺</i>)	Backman, (1972)

Table. 2.1.2 Plasmids

Plasmid	Description	Source or reference
<u>InC11 plasmids</u>		
ColIb <i>drd1</i> pLG221	InC11 derepressed for transfer ColIb <i>drd1 cib::Tn5 Km^R</i>	Laboratory stock Howland <i>et al.</i> (1989)
pLG272	ColIb <i>cib::Tn5 Km^R</i>	Howland <i>et al.</i> (1989)
pLG292	ColIb <i>drd1 ardA::aphA-1 Km^R</i>	Read <i>et al.</i> (1992)
<u>Other plasmids</u>		
pBR328	4.9 kb Ap ^R Cm ^R Tc ^R	Soberon <i>et al.</i> (1980)
pCRS3	pBR328Ω(ColIb <i>oriT⁺ SalI</i> 15.9 kb) Ap ^R Cm ^R	Rees <i>et al.</i> (1987)
pCRS4	pBR328Ω(ColIb <i>SalI</i> 10.1 kb) Ap ^R Cm ^R	Rees (1986)
pHP45Ω	4.3 kb Ap ^R Sm ^R Sp ^R	Prentkl and Krisch (1984)
pIC19H	2.7 kb Ap ^R	Marsh <i>et al.</i> (1984)
pJH16	pACYC184Ω(<i>EcoRI m_{RI}⁺</i>) Cm ^R	Heitman <i>et al.</i> (1989)
pLG290	pUC19Ω(ColIb <i>ardA⁺ SalI-PstI</i> 2.7 kb) Ap ^R	Read <i>et al.</i> (1992)
NTP14	17 kb <i>cea⁺ r_{RI}⁺ m_{RI}⁺ Ap^R</i>	Smith <i>et al.</i> (1976)
pRR2-1	pRR1Ω[ColIb <i>drd1</i> E1 (20.4 kb) + E16 (2 kb)] Ap ⁺	Roscoe (1996)
pRR4	pBR328Ω(ColIb <i>eex⁺ EcoRI</i> 3.4 kb) Ap ^R Cm ^R	Roscoe (1996)
R751	56 kb IncPβ, Tn402/5090, Tn4321Ω (RK2 <i>terR, terA, BglIII</i> ~3.8 kb) Tc ^R Tp ^R	Thorsted <i>et al.</i> (1998)
pSBCP1	pIC19HΩ(ColIb <i>ssb PstI-ClaI</i> 1.1 kb) Ap ^R	Laboratory stock

pSBCP2	pIC19HΩ(Collb <i>Clal-PstI</i> 1.6 kb) Ap ^R	Laboratory stock
pSBSP2	pIC19HΩ(Collb <i>psiB PstI-SalI</i> 1.6 kb) Ap ^R	Laboratory stock
pUC19	2.7 kb Ap ^R	Yanisch-Perron <i>et al.</i> (1985)

Table. 2.1.3 Plasmids constructed during this work

Name	Description	Published as
pNA1/pNA1a	pUC19Ω(CollIb <i>oriT</i> ⁺ <i>nic</i> ⁺ <i>PstI</i> 1.6 kb) Ap ^R	
pNA2	pIC19HΩ(CollIb <i>oriT</i> ⁺ <i>nic</i> ⁺ <i>PstI</i> - <i>ClaI</i> 1.1kb) Ap ^R	pLG2059
pNA3	pIC19HΩ(CollIb <i>oriT</i> ⁺ <i>nic</i> ⁺ <i>Sall</i> - <i>EcoRI</i> 8.0 kb) Ap ^R	
pNA4	pNA2Δ <i>nic</i> :: <i>SphI</i> - <i>KpnI</i> Ap ^R	
pNA5	pNA2Δ <i>nic</i> :: <i>KpnI</i> - <i>SphI</i> Ap ^R	
pNA6	pIC19HΩ(CollIb <i>ardA</i> ⁺ <i>BglII</i> - <i>Sall</i> 6.1 kb) Ap ^R	
pNA7	pNA2Δ <i>nic</i> :: <i>EcoRI</i> Ap ^R	pLG2060
pNA8	pNA7Ω(Δ <i>nic</i> ::omega intersposon <i>EcoRI</i> 2.0 kb) Ap ^R Sm ^R Sp ^R	pLG2061
pNA9	pIC19HΩ(CollIb <i>BglII</i> - <i>PstI</i> 0.62 kb) Ap ^R	
pNA10	pIC19HΩ(CollIb <i>PstI</i> 2.0 kb) Ap ^R	
pNA11	pLG221Δ <i>nic</i> ::Ω Km ^R Sm ^R Sp ^R	pLG2062
pNA12	pLG292Δ <i>nic</i> ::Ω Km ^R Sm ^R Sp ^R	pLG2063
pNA13	pIC19HΩ(CollIb <i>PstI</i> 0.76 kb) Ap ^R	
pNA14	NTP14Ω(Ω intersposon <i>ClaI</i> 2.0 kb) (<i>r_{RI}</i> ⁺ <i>m_{RI}</i> ⁺) Ap ^R Sm ^R Sp ^R	
pNA15	pIC19HΩ(Ω intersposon <i>ClaI</i> 2.0 kb) Ap ^R Sm ^R Sp ^R	
pNA16	pBR328Ω(NTP14 <i>Sall</i> - <i>ClaI</i> 9.7 kb) (<i>r_{RI}</i> ⁻ <i>m_{RI}</i> ⁻) Ap ^R Cm ^R	
pNA17	pBR328Ω(NTP14 <i>Sall</i> - <i>ClaI</i> 7.3 kb) (<i>r_{RI}</i> ⁺ <i>m_{RI}</i> ⁺) Ap ^R Cm ^R	

Plasmids pLG2059, pLG2060, pLG2061, pLG2062, pLG2063 are also described in Althorpe *et al.* (1999).

Table 2.1.4 Oligonucleotide primers used during this work

Primer	Sequence (5'-3')	Template	Reference
M13/pUC Universal Forward	CCCAGTCACGACGTTGTAAAACG	M13/pUC	GIBCOBRL (1994)
M13/pUC Universal Reverse	AGCGGATAACAATTCACACAGG	"	"
Primer B	CTGCATAAGACTATGATGC	pNA2	Althorpe <i>et al.</i> , (1999)
Primer C	GCGTATTCCTTGTCACCG	"	"
Primer 5	CGGAATTCGTTTTTCAGGCCATTATAGCC	"	"
Primer 6	CGGAATTCGGCAATTGTAATACCGTCCC	"	"
Lead 1	CAGCCAGGAAAGATGGCATCG	pNA9	Chapter 7
Lead 2	CATCAATTGTTGTTTCATATCGATGC	"	"
Lead 3	GGTACATCACCTCCAGTAGCCG	pLG290	"
Lead 33	CCTTCTGCCGTCCACTTCTTCG	"	"
Lead 331	GACATCGGACCGTAGCGGCAGC	"	"
Lead 332	GACCGTGCAATAAAGGTGAAGG	"	"
Lead 333	GGCAGGCGGCGAAAAAGTCG	"	"
Lead 334	GAGGGACCGGTGACCGCAGC	"	"
Lead 335	GATATCGCGGTCGTGCTCTGC	"	"
Lead 4	GCGTAAATCCGTGCGACAGACG	"	"
Lead 44	CTCCGCTTTTACCCGGCCATGC	"	"
Lead 43	CCACTGAAAGCACTGAAGAGGTGC	"	"
Lead 432	CGCTCTATTTGACTATGAGG	"	"
Lead 443	CATCGCCGAAGCCTGGTTTGACC	"	"
Lead 433	GGACAGCTGCCCCGTGACGTGG	"	"
Lead 431	CGACAGGCAGCTGCTGGCGCACC	"	"
Lead 434	GTCTCTACCCTCCAACCTCCC	"	"
Lead 5	GGAACTGCATCTTCCTGCGG	pNA13	"
Lead 55	CCGCTGGTTGATATCACTGTCG	"	"
Lead 6	CGTATAAACATAAACCATCTTCACCC	"	"
Lead 66	CTTCTGCTCACTGCAGGTGACG	"	"
Lead 7	GCACAAATCTGTCAAGCTGACC	pNA10	"
Lead 77	CGTGCCATGTTCTCTGTGTCACG	"	"
Lead 771	GTAAGGCCACTGCGCCTGATGG	"	"
Lead 772	CAATGTGATGAATCAGACTTTACC	"	"
Lead 8	CGTATCCGTCATCAGGCGGACC	"	"
Lead 88	GTTTTGTGGCTGGTACATCTCTCC	"	"
Lead 881	CTTTTTCAAAGGTGTTTTATGACC	"	"
Lead 882	GCATTTTGCCGGCCCTTTTGG	"	"

2.2 Media and Chemicals

2.2.1 Media

Unless otherwise stated, bacterial strains were routinely grown in nutrient broth (NB) and on nutrient agar (NA) plates. Suspensions of bacterial strains were made in phosphate buffer (PB). Bacteriophage λ was always resuspended and diluted in λ buffer (λ B). For long term storage of strains, overnight cultures of bacteria were streaked onto NA plates supplemented with the appropriate antibacterial agent and incubated at 37°C for 12 hours. Frozen stocks were made by resuspending the bacteria in NB + 20% glycerol, which were then stored in vials at -20°C. Frozen cultures made in this way have a shelf-life of ~10 years. All media and chemicals were resuspended in distilled water unless indicated otherwise.

Nutrient agar (NA)

Oxoid No.2 nutrient broth	25 g l ⁻¹
Bio-gene agar	16 g l ⁻¹

Luria agar (LA)

Oxoid tryptone	10 g l ⁻¹
Oxoid yeast extract	5 g l ⁻¹
NaCl	5 g l ⁻¹
Bio-gene agar	15 g l ⁻¹

Phosphate buffer (PB)

KH ₂ PO ₄	3 g l ⁻¹
NaHPO ₄	7 g l ⁻¹
NaCl	4 g l ⁻¹
MgSO ₄ ·7H ₂ O	0.1 g l ⁻¹
	pH 7.0

λ buffer (λ B)

Tris-HCl	0.06 M
MgSO ₄ ·7H ₂ O	2.5 g l ⁻¹
Gelatin	0.05 g l ⁻¹
	pH 7.2

Soft nutrient agar (SNA)

BBL trypticase	10 g l ⁻¹
NaCl	5 g l ⁻¹
Bio-gene agar	6.5 g l ⁻¹

Bottom layer agar (BLA)

BBL trypticase	10 g l ⁻¹
NaCl	5 g l ⁻¹
Bio-gene agar	10 g l ⁻¹

Minimal media (MM)

Minimal salts*	100 ml l ⁻¹
20% glucose	50 ml l ⁻¹
0.1% Thiamine HCl	1 ml l ⁻¹
Ca-Mg salts**	10 ml l ⁻¹
Bio-gene agar	16.2 g l ⁻¹

*Minimal salts:

Na₂HPO₄ (anhydrous) 79 g l⁻¹, KH₂PO₄ (anhydrous) 30 g l⁻¹, NaCl 5 g l⁻¹, NH₄Cl 10 g l⁻¹.

** Ca-Mg salts:

CaCl₂, 1.47 g l⁻¹ (0.01M), MgSO₄·7H₂O 4.6 g l⁻¹ (0.1M).

Thiamine HCl was prepared in distilled water, filter sterilised and stored at 4°C.

2.2.2 Antibacterial agents

Unless otherwise stated, antibacterial agents were added to media at the following concentrations: ampicillin (Ap), 100 $\mu\text{g ml}^{-1}$; chloramphenicol (Cm), 25 $\mu\text{g ml}^{-1}$; kanamycin (Km), 25 $\mu\text{g ml}^{-1}$; naladixic acid (Nal), 25 $\mu\text{g ml}^{-1}$; rifampicin (Rif), 25 $\mu\text{g ml}^{-1}$; streptomycin (Sm), 50 $\mu\text{g ml}^{-1}$; spectomycin (Sp), 50 $\mu\text{g ml}^{-1}$ and tetracycline (Tc), 7.5 $\mu\text{g ml}^{-1}$.

Chemicals used for α -complementation tests (blue and white colony selection) were added to LA Ap plates at the following concentrations; X-gal, 40 $\mu\text{g ml}^{-1}$ and IPTG, 0.1 mM ml^{-1} (Sambrook *et al.*, 1989). Stock solutions of X-gal were always made up in dimethylformamide.

2.2.3 11.1 x PCR buffer

Tris-HCl pH 8.8	476 mM ml^{-1}
(NH ₄) ₂ SO ₄	121.9 mM ml^{-1}
MgCl ₂	49 mM ml^{-1}
2-mercaptoethanol (100%)	5 $\mu\text{l ml}^{-1}$
EDTA pH 8.0	50 $\mu\text{M ml}^{-1}$
dATP	11.1 mM ml^{-1}
dCTP	11.1 mM ml^{-1}
dGTP	11.1 mM ml^{-1}
dTTP	11.1 mM ml^{-1}
BSA (DNAse free)	1.25 $\mu\text{g ml}^{-1}$

2.3 Phenotypic Characterisation of Bacterial Strains

2.3.1 Colicin production

Strains to be tested for colicin production were grown overnight in NB. One μl of culture was spotted onto a NA plate which was incubated for a further 12 hours at 37°C. The bacteria were then killed by exposure to chloroform vapour for 10 minutes, allowing another 15 minutes for evaporation of the vapour. The plate was overlaid with 3.5 ml of SNA

containing 0.1 ml of a 1:20 dilution of an overnight culture of colicin sensitive cells. After 5 hours incubation at 37°C, colicin-producing strains were indicated by a clear zone of killing of the indicator strain around the test strain.

2.3.2 *recA* mutants

To test strains for the carriage of a *recA* mutation, the bacteria were streaked onto NA plates and treated with a series of timed exposures to shortwave UV light (0.5 J sec⁻¹ m⁻²). The growth of the test strain that had been irradiated for 0, 10, 15, 30 and 60 seconds was compared to that of a wild type strain, GI65 and a *recA* strain, JO8. RecA mutants have increased sensitivity to UV irradiation.

2.3.3 *recD* mutants

The RecD phenotype of DL307 was tested using the bacteriophage λ MM5885(χ° Red⁻ Gam⁻) and λ MM5659(χ^{+} Red⁻ Gam⁻) supplied as gifts from David Leach. The screen relies on the ability of both the χ° and χ^{+} phage to form indistinguishable large plaques on *recD* strains whilst on *rec⁺* strains, the χ° phage form small plaques and the χ^{+} phage large plaques (Russel *et al.*, 1989)

2.3.4 Sensitivity to bacteriophage

Bacterial strains were routinely tested for their ability to restrict a virulent strain of bacteriophage λ . Overnight cultures of strains to be tested were diluted in NB and 10 mM MgSO₄ to an A₆₀₀ absorbance value of 0.05. The bacteria were then grown at 37°C with vigorous shaking to an A₆₀₀ of 0.5 (mid-exponential phase). The bacteria were resuspended in 3 ml of SNA (~2 × 10⁸ cells) and used to overlay a BLA plate. λ *vir.0*, propagated on NM654, λ *vir.K* on GI65 and λ *vir.RI* on GI65(NTP14) were used, depending on the restriction phenotype of the strain being tested. The phage were diluted in λ B and 10 μ l spotted onto the top layer of the BLA plate. The phage suspension was allowed to adsorb into the agar for 20 minutes and the plate then incubated at 37°C,

overnight. Comparison of the efficiency of plating (e.o.p) of the phage on the test strain to that of a control strain allowed the restriction phenotype to be determined.

2.4 Strain Manipulations

2.4.1 Bacterial conjugation

Overnight cultures were diluted in NB to an A_{600} of 0.05 using a Bausch and Lomb 501 spectrophotometer and grown with vigorous shaking at 37°C to an A_{600} of 0.5 (2×10^8 cells ml^{-1}). Conjugation mixtures contained equal volumes of donor and recipient bacteria. All interrupted filter matings involved filtering 0.5 ml of mating mixture onto a 25 mm cellulose acetate filter (Whatman[®], pore size 0.45 μm) for each time point, which was then placed on the surface of a prewarmed nutrient agar plate for the conjugation period. Bacteria were released from each filter by vigorous blending in 5 ml phosphate buffer. The titre of the transconjugants were determined by serial dilution and 100 μl samples spread onto nutrient agar plates containing the appropriate antibacterial agent for selection. For liquid matings equal amounts of donor and recipients were added to a 200 ml conical flask and incubated at 37°C in a water bath with gentle rotation (49 rpm). A 0.3 ml sample was recovered and added to 2.7 ml of PB in a glass tube. The tube was agitated violently in a blender for 10 seconds to terminate conjugation. Again, the titres of transconjugants were determined in the same way as for filter matings.

2.4.2 Transformation of bacterial strains

The method used to transform bacterial cells with DNA was by a chemical method based on that of Cohen *et al.* (1972).

Competent cells were made by diluting overnight cultures of bacteria in NB to an A_{600} of 0.05 and grown with vigorous shaking at 37°C to an A_{600} value of 0.5 (mid-exponential phase). The cells were pelleted at 4°C in an MSE

Microcentaur and then washed twice by centrifugation in 5 ml of ice cold 0.1 M CaCl₂. The pellet was finally resuspended in 1 ml of 0.1 M CaCl₂ giving a concentration of approximately 2×10^9 cells ml⁻¹. Competent cells made this way can be stored for up to 24 hours at 4°C.

1-5 µg of DNA was added to 250 µl of competent cells in a sterile Eppendorf tube. Incubation of the cells for 1 hour on ice was followed by a 4 minute heat-shock at 42°C. The cells were then returned to ice for a further 2 minutes before being transferred to 5 ml of prewarmed NB. After 1 hour incubation at 37°C with vigorous shaking the cells were pelleted, resuspended in PB, and diluted to give single colonies when samples were spread onto selective NA plates.

2.4.3 Bacteriophage propagation

The bacteriophage to be propagated was plated on the appropriate bacterial strain at a titre that formed a number of plaques after overnight growth. The SNA overlay was scraped from the BLA plate and transferred to a Sorvall centrifuge tube. Addition of 1.5 ml λB and 0.5 ml of chloroform to the tube was followed by the mix being thoroughly mashed with a sterile spatula. The tube was allowed to stand for 30 minutes at 37°C before being centrifuged at 8000 g in a Sorvall SS-34 fixed angle rotor for 5 minutes. The supernatant was removed to a sterile test-tube containing 0.1 ml of chloroform, vortexed, then incubated for a further 15-20 minutes at 37°C. The aqueous layer was removed to a screw cap tube and the phage suspension stored at 4°C.

2.4.4 Transposon insertion mutagenesis

Tn5 mutants were obtained in *Escherichia coli* strain W3110 (Winans and Walker, 1983).

2.5 DNA Manipulations

2.5.1 Small scale plasmid DNA isolation

For isolation of high quality DNA samples for use in sensitive applications such as DNA sequencing, cloning or PCR, a QIAprep spin minikit (Qiagen) was used according to the manufacturers instructions.

For standard small scale preparations of plasmid DNA an alkaline lysis method based on that of Birnboim and Doly (1979) was used.

1.5 ml of an overnight culture of plasmid containing cells, grown with appropriate antibiotic selection, were centrifuged at 13,000 g for 3 minutes in a MSE Microcentaur. The pellet was resuspended in 100 µl of lysis buffer (25 mM Tris-HCl, pH 8.0; 10 mM EDTA, 50 mM sucrose) and incubated on ice for 5 minutes. Addition of 200 µl of alkaline SDS (0.2 M NaOH; 1 % Sodium dodecyl sulphate) to the cells was followed by gentle mixing and 5 minutes incubation on ice. 150 µl of 5 M potassium acetate (pH 4.8) was then added to the tube, the contents mixed carefully before a final 5 minute incubation period on ice.

The lysate was centrifuged for 10 minutes to remove the cell debris and the supernatant transferred to a clean Eppendorf tube containing 400 µl of a phenol/chloroform mixture [50% phenol, 48% chloroform, 2% isoamyl alcohol] (Sambrook *et al.*, 1989). After mixing, the tube was centrifuged for 3 minutes, the upper phase was removed and transferred to an equal volume of iso-propanol. The mixture was incubated on ice for 15 minutes and then centrifuged for 30 minutes to pellet the DNA. The supernatant was discarded and 500 µl of 70 % ethanol was added to wash the pellet. The tube was centrifuged for 5 minutes and the pellet air dried before being resuspended in 20 µl distilled water. The concentrations of DNA isolated by this method were estimated by electrophoresis on an agarose gel along side a control DNA sample of known concentration. RNA was removed from DNA samples by incubating at 37°C for 1 hour in the presence of RNaseA (0.25 µg).

2.5.2 Large scale plasmid DNA isolation

All large scale preparations of plasmid DNA were carried out using Qiagen mini kits according to the instructions.

2.5.3 Analysis of the concentration and purity of DNA samples

DNA yields and purity were determined from the A_{260}/A_{280} values measured in a Hitachi U-2000 spectrophotometer. DNA concentrations were calculated where: $\sim 50 \mu\text{g ml}^{-1}$ of double-stranded DNA, $\sim 40 \mu\text{g ml}^{-1}$ of single-stranded DNA or RNA and $\sim 20 \mu\text{g ml}^{-1}$ of single-stranded oligonucleotide have an A_{260} value of 1. The purity of DNA samples was confirmed by the A_{260}/A_{280} values of the sample having a ratio of 1.8 to 2.0 (Sambrook *et al.*, 1989).

2.5.4 Restriction enzyme analysis of DNA

Restriction endonuclease enzymes, supplied by GIBCOBRL, were routinely used to analyse DNA. Conditions and reaction buffers used in digests were as recommended by the manufacturers. Reactions were generally carried out in a total reaction volume of 20 μl consisting of restriction enzyme, 1 x reaction buffer, $\sim 1 \mu\text{g}$ of DNA and if necessary 0.25 μg of RNaseA.

Samples, to be analysed by gel electrophoresis, were mixed with 2/10 loading dye (5 mM Tris-HCl (pH 7.5), 100 mg ml^{-1} glycerol, 0.01 mg ml^{-1} bromophenol blue). Restriction fragments were separated on 0.6-1.0% (w/v) Seakem HGT agarose (FMC Corp.) horizontal gels in TAE buffer (40 mM Tris-acetate (pH 7.4), 1 mM EDTA) containing ethidium bromide (10 $\mu\text{g ml}^{-1}$), at 20-100 volts for 2-8 hours. DNA fragments were visualised on a short wavelength UV transilluminator and photographed using either a Polaroid MP4 land camera with 15 seconds exposure or a UV visual analyser (Sony imagestore 5000, version 7.2). The sizes of the DNA fragments were

determined by comparisons with λ Hind III molecular weight markers (23.13, 9.4, 6.56, 4.36, 2.32, 2.03, 0.56 kb).

2.5.5 Isolation of DNA restriction fragments

All DNA bands were cut with a scalpel from a 0.8% Seakem HGT agarose gel (FMC Corp.) and transferred to an Eppendorf tube. Purification of the DNA fragments from the agarose was achieved using a QIAEX II Gel extraction kit (Qiagen) according to the manufacturers instructions. The yield and purity of DNA isolated was calculated as described in section 2.5.3.

2.5.6 Dephosphorylation of DNA fragments

To prevent re-ligation of single-stranded 5' protruding ends created by some restriction enzymes, the 5' phosphate residues of the DNA were removed by treatment with calf intestinal alkaline phosphatase (CIAP-Pharmacia Biotech Ltd.) according to the manufacturers instructions. Removal of the CIAP enzyme was by phenol : chloroform extraction and isolation of the DNA was by ethanol precipitation.

2.5.7 Ligation of DNA fragments

Vector and plasmid DNA to be ligated were digested with the appropriate restriction enzymes. The corresponding DNA fragments from the digestion were isolated from an agarose gel after separation by gel electrophoresis (section 2.5.5). Ligations were carried out in a total reaction volume of 20 μ l including 5 units of T4 DNA ligase and T4 ligation buffer (GIBCOBRL). The concentration of vector: insert DNA was adjusted so that a ratio of 1 to 3 was achieved. The ligation reaction was incubated overnight at room temperature. For blunt-end ligations the same conditions were used except the units of T4 DNA ligase enzyme was increased to 25 per reaction and the vector: insert DNA ratio was changed to 1 to 8.

2.5.8 Primers

Oligonucleotide primers were made by the PNAACL facility, University of Leicester, using an Applied Biosystems 394-8 synthesiser. Primers arrived in a 35% ammonia solution and were precipitated on ice for 2 hours in 1 volume of 3 M $\text{NH}_4\text{C}_2\text{H}_3\text{O}_2$ (pH 7.0) and 3 volumes of ethanol. The primers were pelleted by centrifugation in a MSE Microcentaur for 30 minutes and washed in 70 % ethanol, dried and dissolved in distilled H_2O . The concentrations of single-stranded oligonucleotide were calculated as described in section 2.5.3. The number of pico moles of primer were calculated and in general 20 pmol ml^{-1} of primer used in PCR reactions and 3.2 pmol ml^{-1} used in DNA sequencing reactions.

2.5.9 DNA sequencing reactions

All DNA sequencing reactions were carried out using the ABI PRISM™ Dye Terminator cycle sequencing ready reaction kit according to the manufacturers instructions. Samples were analysed by the PNAACL facility, Leicester University, on a PE-ABI 377 automated DNA sequencer. Sequence data obtained was then analysed using the abi (Applied Biosystems Inc.) sequence analysis program "FACTURA", version 1.2.

Further analysis of the sequence data was carried out using the Genetics Computer Group (GCG) Wisconsin Package Suite of programs (version 9.1). FASTA (Pearson and Lipman, 1988) and BLAST (Altschul *et al.*, 1990) searches were carried out through the European Bioinformatics Institute site on the World Wide Web (<http://www.ebi.ac.uk>).

2.6 Standard and inverse PCR

Primers used for PCR reactions are as described in section 2.5.8. All PCR-amplifications were carried out in a total reaction volume of 50 μl containing,

5 μ l of 11.1 x PCR buffer; 50 ng of template DNA; 20 pmol ml⁻¹ of each primer and 5 units of *Taq* DNA polymerase (Advanced Biotechnologies Ltd.) or when the PCR product was to be used in further applications, 5 units of the proof-reading polymerase BIO X-Act™ (Bioline (UK) Ltd) was used. All reactions were carried out in a hybaid omnigene PCR machine. Example programs include: 95°C-30 sec (Hot start), 1 cycle; 95°C-45 sec, 49-60°C-30 secs, 72°C – 1 min / kb amplified, 20-35 cycles; 72°C – 1 min / 1 kb, 1 cycle.

2.6.1 Optimisation of the inverse-PCR protocol for amplification of template pNA2 using primer 5 and primer 6

The first step involved establishing the optimal temperature at which primer 5 and primer 6 anneal to template pNA2. Two sets of 25 cycle PCR-amplifications were set up involving reactions with primer 5 and universal forward (UF) and with primer 6 and universal reverse (UR). The annealing sites for each of these primers are indicated on the genetic map of pNA2 presented in Fig. 4.3 (A). A range of annealing temperatures (49-56°C) were tested and the resulting DNA fragments produced in each reaction examined by agarose gel electrophoresis. High annealing temperatures reduce the amount of non-specific binding of primers to their template which in turn increases the amount of specific DNA product yielded (Saiki, 1989). Primer 5 and primer 6 were inferred to have an optimal annealing temperature of 53°C as only the desired DNA product was detected in each reaction at this temperature (Fig. 4.3, B lane numbers 3 and 4).

Further optimisation of the PCR protocol involved examining the time required for both primer 5 and primer 6 to be fully extended by the *Taq* polymerase enzyme and result in the amplification of the 3.8 kb pNA2 template. It is generally accepted that one minute is the maximum time required for *Taq* polymerase to extend 1 kb of DNA (Saiki, 1989). Based on this principle, a 35 cycle PCR-amplification of pNA2 was set up with primer 5 and primer 6, allowing 4 minutes extension time. The extension product produced

by this reaction was analysed by agarose gel electrophoresis, the result of which is presented in Fig. 4.3 (B), lane number 5. These findings indicate that 4 minutes is adequate time for the amplification of 3.8 kb of DNA.

Chapter Three

Genetic Analysis of a Plasmid-Mediated Restriction Evasion Mechanism

3.1 Introduction

This chapter describes a series of genetic experiments aimed at analysing the mechanism employed by *ColIb_{drd}* during conjugation, which enables the plasmid to avoid destruction by two different classes of restriction enzyme.

ColIb_{drd} is remarkably resistant to destruction by type I and type II restriction enzymes during its transfer by conjugation, even though the plasmid carries multiple restriction enzyme target sites [\sim 7 *Eco*KI sites; 20 *Eco*RI sites] (Read *et al.*, 1992; Wilkins *et al.*, 1996). Such resistance is not a feature common to all transferring plasmids. For example, IncP plasmids are extremely sensitive to these enzymes during their transfer, despite a distinct lack of restriction enzyme recognition sites in the IncP replicon [R751: 12 *Eco*KI sites; 5 *Eco*RI sites] (R751 EMBL accession number U67194). *ColIb*'s resistance to type I enzymes involves a specialised plasmid encoded anti-restriction mechanism known as *ardA*. This system is incapable of alleviating restriction by type II enzymes (Delver *et al.*, 1991; Read *et al.*, 1992). Interestingly, after the *ardA* gene has been insertionally inactivated, *ColIb_{drd}* still displays marked resistance to type I and type II restriction enzymes during its transfer (Read *et al.*, 1992). Evidence suggesting a second *ardA*-independent restriction evasion process encoded by *ColIb* has been provided by Read *et al.* (1992). This process becomes manifest after several minutes of conjugation and alleviates restriction of *ColIb_{drd}* by type I and type II enzymes during second or subsequent rounds of transfer. The molecular basis of this second evasion process has yet to be elucidated.

One hypothesis proposed by Read *et al.* (1992) to explain this second restriction-avoidance system of *ColIb_{drd}* is that transfer of multiple copies of the plasmid may overwhelm the restriction enzymes in the recipient cell by substrate saturation. A second hypothesis is that multiple transfers of *ColIb_{drd}* induce a change in the physiology of the recipient cell, which in turn leads to a transient breakdown of the restriction barrier. Another possibility is that *ColIb_{drd}* possesses a gene(s) that encodes an anti-restriction function active against both classes of restriction enzyme. In all cases, the expected phenotype would be alleviation of restriction allowing the next round of plasmid transfer to be successful.

Four independent yet related conjugation experiments are described in this chapter, which are all based on examining whether plasmid *ColIb_{drd}* can confer cross-protection on IncP β plasmid R751 from destruction by type I and type II restriction enzymes. To demonstrate the relative resistance of *ColIb_{drd}* to these enzymes during conjugation the first experiment involved a re-examination of the effects of type I and type II restriction on the transmission of plasmids R751 and *ColIb_{drd}*. The second experiment analysed the effect of co-transferring R751 with *ColIb_{drd}* from the same donor strain to a restricting recipient. Third, the effect of transferring unmodified R751 DNA to a restricting recipient harbouring *ColIb_{drd}* was explored. Finally, *ColIb* wild type and the cloned *ColIb* exclusion genes (*eex*) were utilised to investigate whether expression of the plasmids restriction-evasion mechanism is triggered by the act of *ColIb* transfer.

3.2 Results and Discussion

3.2.1 Effects of restriction on the conjugative transmission of plasmids R751 and *ColIb_{drd}*

To demonstrate that *ColIb_{drd}* encodes an *ardA*-independent restriction evasion mechanism, the effects of restriction on the conjugative transmission of unmodified R751 and *ColIb_{drd}* were examined. The standard experiment

involved an interrupted mating of a donor strain harbouring R751 or ColI*bdrd* and a restricting recipient strain. All conjugations were done on filters as plasmid R751 specifies only a surface obligatory mating system. The effects of restriction on the transmission of each plasmid was monitored by comparing the number of R751 or ColI*bdrd* transconjugants produced in the restricting recipients to those produced in the non-restricting counterpart. To simplify the genetic selections, a rifampicin-resistant (Rif^R) donor strain and a naladixic acid-resistant (Nal^R) recipient strain were used. One advantage of using a Nal^R recipient strain is that naladixic acid inhibits conjugative DNA transfer by sensitive donor cells. Therefore, any additional conjugations that might occur on the surface of the NA plates are prevented by the addition of Nal.

The donors and recipients used in each mating experiment are described within the figure legends. The R751 plasmid used carries two selectable antibiotic resistance genes one of which confers resistance to tetracycline (Tc^R), the other trimethoprim (Tp^R). ColI*bdrd* plasmids used were pLG221 and pLG292 both of which confer resistance to kanamycin (Km^R) through the carriage of a Tn5 insertion in the *cib* (colicin Ib) locus of pLG221 and a non-transposable Tn903 *aphA-1* insertion in the *ardA* gene of pLG292 (Rees *et al.*, 1987; Read *et al.*, 1992).

The two R-M systems chosen for study were *EcoKI* and *EcoRI* as examples of type I and type II systems, respectively. The type I system is encoded by the *E. coli* K-12 chromosome and the type II system is encoded by the natural ColE1-like plasmid NTP14 (Arber and Dussoix, 1962; Smith *et al.*, 1976). Both these R-M systems were originally isolated in strains of *Escherichia coli* and are therefore presumed to be native to the Enterobacteriaceae family. Therefore, any plasmid whose transfer range overlaps this family of bacteria is likely to encounter these R-M systems frequently.

The results of the mating experiments are presented in Fig. 3.1 and Fig. 3.2. It is noted that data in Fig. 3.2 (A and B) were obtained by repeating part of the work described by Read *et al.* (1992), with the exception that the

Fig. 3.1. Kinetics of transmission of unmodified R751 to recipient cells specifying *EcoRI* or *EcoKI*.

(A) R751 transconjugants from conjugations of GI65R(R751) donor cells and either recipients GI65N (O, control) or *EcoRI* restricting recipients GI65N(NTP14) (●). (B) R751 transconjugants from conjugations of NM654R(R751) donor cells and either recipient cells NM816N (□, control) or *EcoKI* restricting recipients GI65N (■). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean values from three experiments.

Fig. 3.1

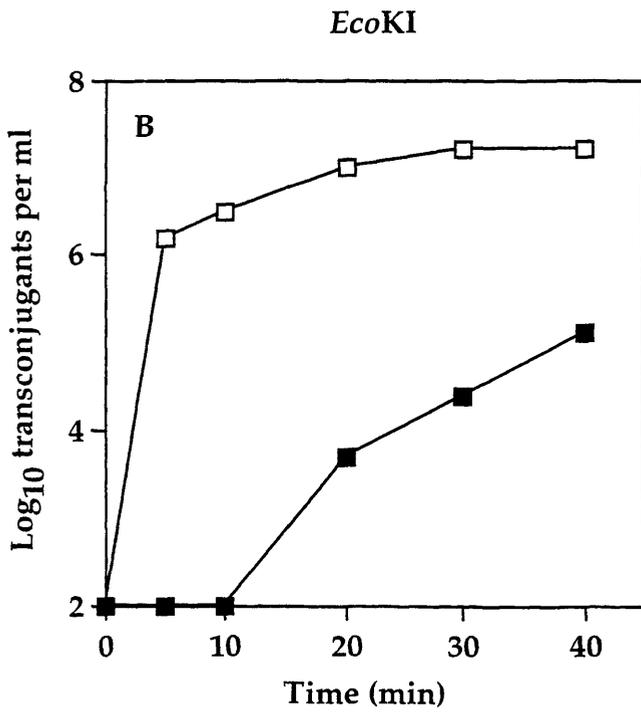
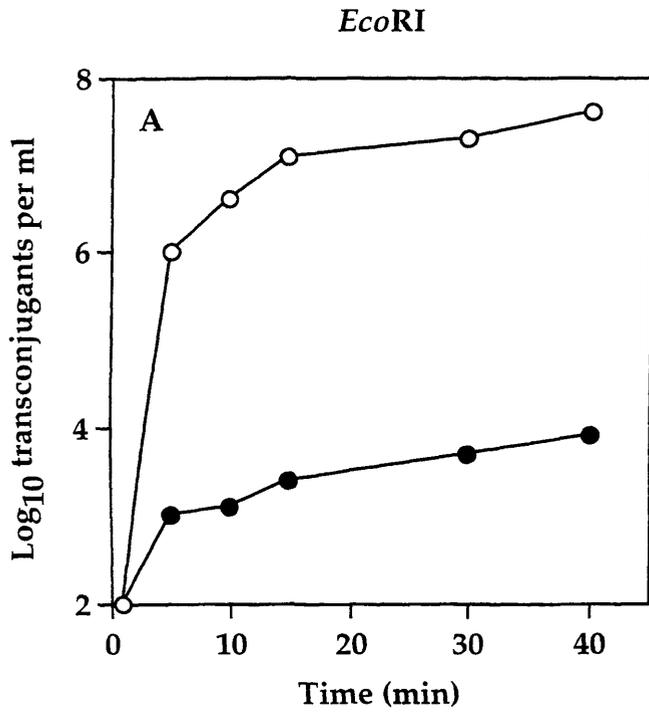
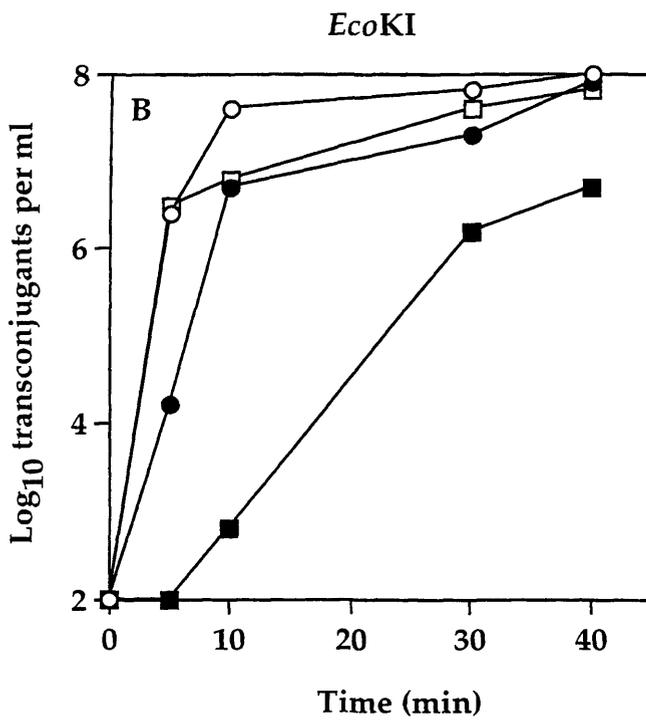
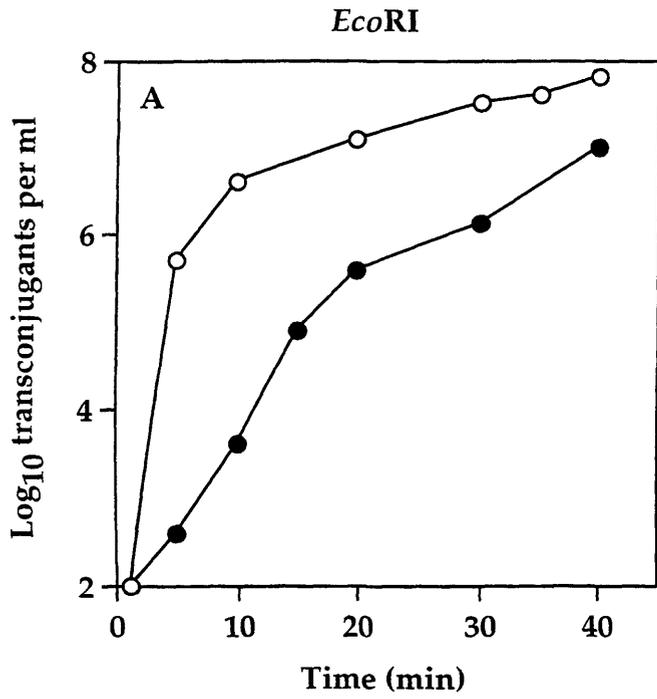


Fig. 3.2. Kinetics of transmission of unmodified pLG221 (Ard⁺) and pLG292 (Ard⁻) to recipient cells specifying *EcoRI* or *EcoKI*.

(A) pLG221 transconjugants from conjugations of GI65R(pLG221) donor cells and either recipient cells GI65N (○, control) or *EcoRI* restricting recipient GI65N(NTP14) (●). (B) pLG221 transconjugants from NM654R(pLG221) donor cells and either recipient cells NM816N (○, control) or *EcoKI* restricting recipient GI65N (●). pLG292 transconjugants from NM654R(pLG292) donor cells and either recipient cells NM816N(□, control) or *EcoKI* restricting recipient GI65N (■). Transconjugants were selected on NA plates containing Nal and Km. Data are the mean values of three experiments

Fig. 3.2



conjugations were performed on the surface of filters rather than in liquid. In addition, Wilkins *et al.* (1996) has described the effects of type II restriction (*EcoRI*) on the transmission of R751. However, the data presented in Fig. 3.1 (A) are more comprehensive as they describe the effects of type II restriction on R751 transmission at intervals over a 40 minute period.

Transmission of R751 was considerably more sensitive to restriction than *Collbdrd*. After 40 minutes, transmission of R751 was reduced 10,000-fold by *EcoRI* restricting recipients and 100-fold by *EcoKI* restricting recipients (Fig. 3.1 A and B). In contrast, productive transfer of pLG221 to the *EcoRI* restricting recipient surged after an initial delay with only a 10-fold reduction in the number of pLG221 transconjugants formed by 40 minutes (Fig. 3.2 A). Interestingly, a similar result was achieved for the transmission of the *Collbdrd ardA* mutant (pLG292) to the *EcoKI* restricting recipient (Fig. 3.2 B). This phenomenon was originally reported by Read *et al.* (1992).

EcoKI restriction had very little effect on the transmission of pLG221 (Fig. 3.2 B), with transconjugant yields reaching the same value as those produced in the non-restricting recipients after 40 minutes. This *EcoKI* resistance of pLG221 can largely be attributed to the *ardA* gene present on the plasmid (Read *et al.*, 1992).

Results in Fig. 3.1 and Fig. 3.2 confirm the findings of Wilkins *et al.* (1996) and Read *et al.* (1992), respectively. Taking into consideration the number of restriction enzyme target sites that *Collbdrd* and R751 carry (section 3.1) and the effect that restriction has on each plasmids' transmission, it is clear that *Collbdrd* encodes an *ardA*-independent restriction evasion mechanism.

3.2.2 Effects of co-transferring R751 with *Collbdrd* to a restricting recipient

The events that lead to the alleviation of type I and type II restriction after multiple transfers of *Collbdrd* are unknown. The three hypotheses described in section 3.1 predict that restriction in the recipient will be alleviated *in trans*. This notion was investigated by testing whether *Collbdrd* could

Fig. 3.3. Co-transfer of plasmids R751 and pLG221 to recipient cells specifying *EcoRI*.

(A) Yield of Nal^{R} , Tc^{R} R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control) or *EcoRI* restricting recipients GI65N(NTP14) (●), GI65R(R751, pLG221) donors and recipients GI65N(NTP14) (■). (B) Yield of Nal^{R} , Km^{R} (pLG221) transconjugants from GI65R(R751, pLG221) donors and either recipients GI65N (Δ, control) or GI65N(NTP14) (▲). (C) Yield of Nal^{R} , Tc^{R} , Km^{R} (R751, pLG221) transconjugants from GI65R(R751, pLG221) donors and either recipients GI65N (∇, control) or GI65N(NTP14) (▼). Transconjugants were selected on NA plates containing Nal, Tc, Km. Data are the mean values of three experiments.

Fig. 3.3

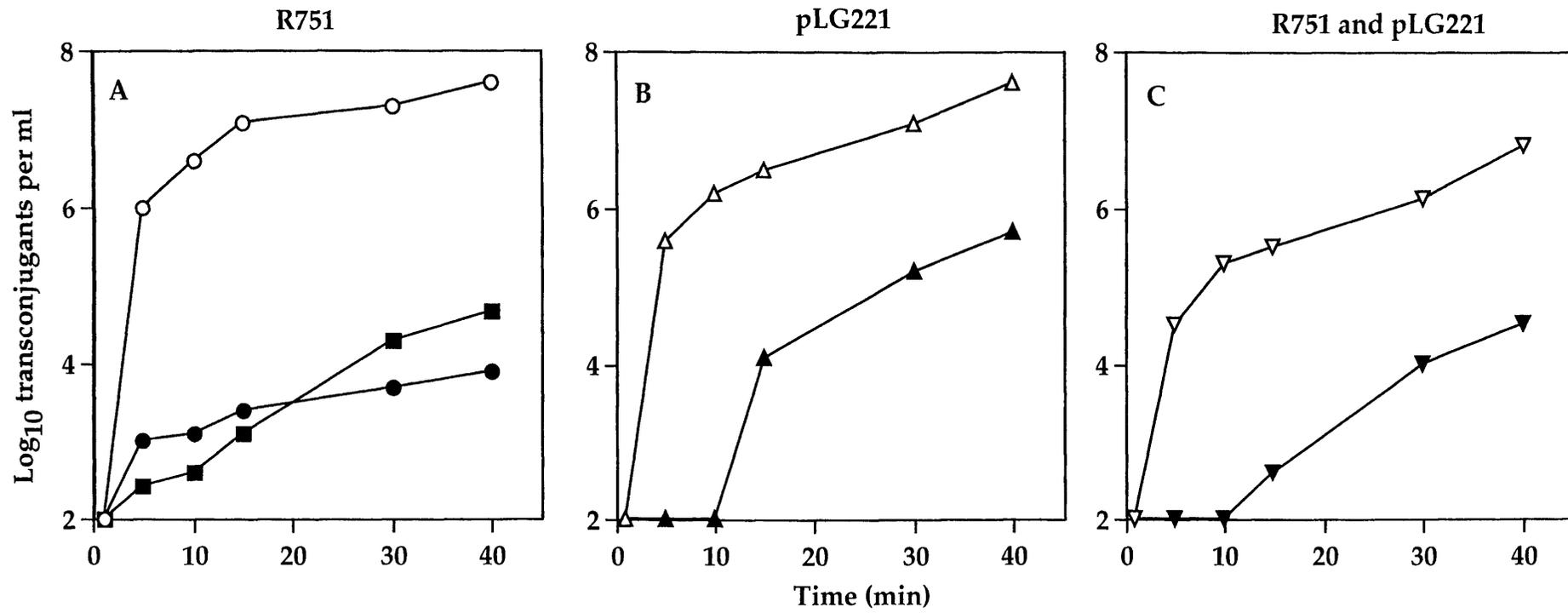


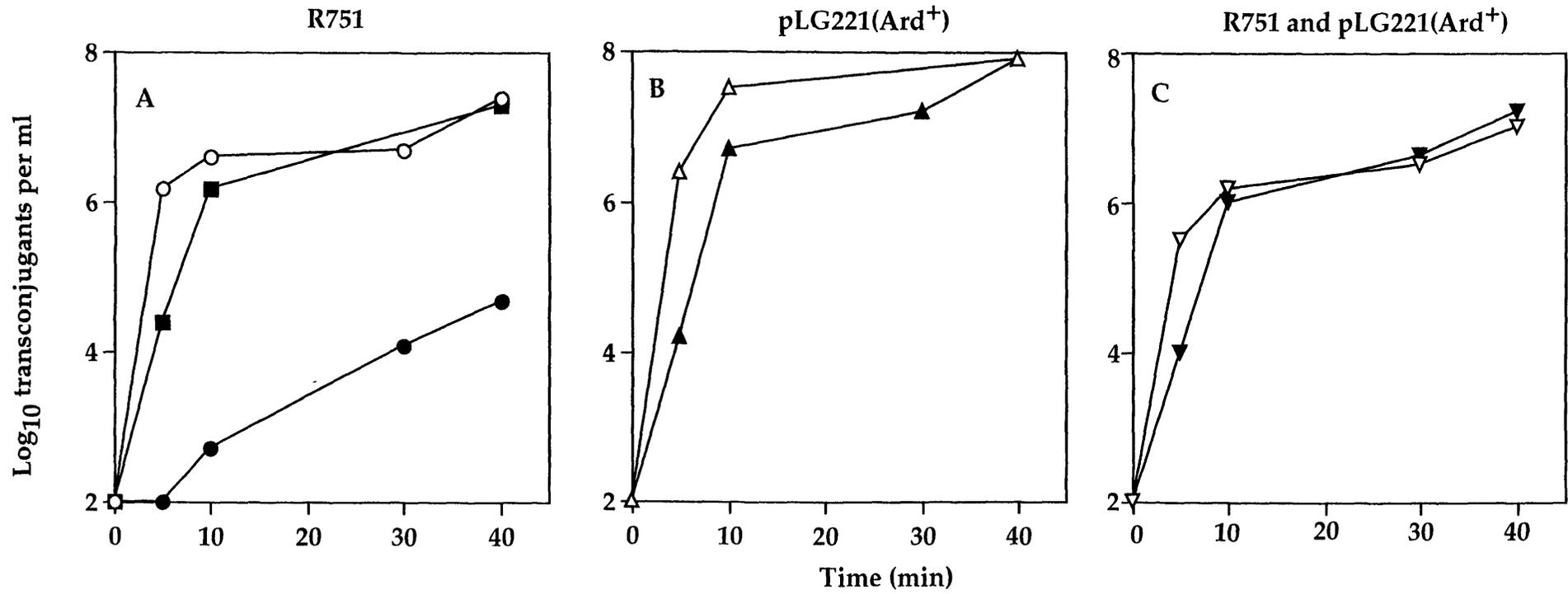
Fig. 3.4. Co-transfer of R751 and pLG292 (Ard^r) to a recipient specifying *Eco*KI.

(A) Yield of NaI^R, Tc^R (R751) transconjugants from conjugations of NM654R(R751) donors and either recipients NM816N (○, control) or *Eco*KI restricting recipients GI65N (●), NM654R(R751, pLG292) donors and recipients GI65N (■). (B) Yield of NaI^R, Km^R (pLG292) transconjugants from NM654R(R751, pLG292) donors and either recipients NM816N (Δ, control) or GI65N (▲). (C) Yield of NaI^R, Tc^R, Km^R (R751, pLG292) transconjugants from NM654R(R751, pLG292) donors and either recipients NM816N (∇, control) or GI65N (▼). Transconjugants were selected on NA plates containing NaI, Tc, Km. Data are the mean value of three experiments.

Fig. 3.5. Co-transfer of R751 and pLG221 (Ard⁺) to a recipient specifying *Eco*KI.

(A) Yield of Nal^R, Tc^R (R751) transconjugants from conjugations of NM654R(R751) donors and either recipients NM816N (○, control) or *Eco*KI restricting recipients GI65N (●), NM654R(R751, pLG221) donors and recipients GI65N (■). (B) Yield of Nal^R, Km^R (pLG221) transconjugants from NM654R(R751, pLG221) donors and either recipients NM816N (△, control) or GI65N (▲). (C) Yield of Nal^R, Tc^R, Km^R (R751, pLG221) transconjugants from NM654R(R751, pLG221) donors and either recipients NM816N (▽, control) or GI65N (▼). Transconjugants were selected on NA plates containing Nal, Tc, Km. Data are the mean values of three experiments.

Fig. 3.5



confer protection on restriction-sensitive plasmid R751 when the two plasmids are co-transferred together to a restricting recipient.

The standard experiment involved mating a donor strain harbouring both R751 and *CollIbdrd* (pLG221 or pLG292) with either an *EcoRI* or *EcoKI* restricting recipient. Restriction in the recipient strains was monitored throughout each mating by examining the number of R751 transconjugants formed. Any alleviation of restriction by transferring *CollIbdrd* would be indicated by an increase in the number of R751 transconjugants produced compared to those formed when R751 is transferred independently of *CollIbdrd*. The numbers of *CollIbdrd* transconjugants and R751/*CollIbdrd* transconjugants formed throughout the experiments were also measured. Such results were indicative of the efficiency of the mating and the activity of the *CollIbdrd*-mediated restriction evasion mechanism.

The results of these experiments are presented in Figures 3.3, 3.4 and 3.5. Figure 3.3 (A) shows that when R751 is co-transferred with pLG221 to an *EcoRI* restricting recipient the number of R751 transconjugants formed increases only marginally. This finding provides no compelling evidence of a *trans*-acting system operating to alleviate *EcoRI* restriction. However, Fig. 3.3 (A) does show a delay in the number of R751 transconjugants formed during the first 6-10 minutes of mating, which is characteristic of that seen in *CollIbdrd* transconjugant production (Fig 3.3 B), suggesting that the protection conferred on R751, although limited, is attributed to the *CollIb*-mediated restriction evasion process.

Figure 3.4 shows the data obtained when R751 is co-transferred with the *CollIbdrd ardA* mutant (pLG292) to an *EcoKI* restricting recipient. As was the case for *EcoRI* restriction (Fig. 3.3), transferring *CollIbdrd* (*ArdA*⁻) was only able to marginally alleviate destruction of transferring R751 from *EcoKI* restriction. This finding is indicated by the 10-fold increase in the number of R751 transconjugants formed in *EcoKI* restricting recipients when co-transferred with *CollIbdrd* (Fig. 3.4 A).

In contrast, Fig. 3.5 (A) shows that when R751 is co-transferred with the *ardA*⁺ *ColIbdrd* plasmid (pLG221) to an *EcoKI* restricting recipient, the yield of R751 transconjugants formed in the restricting recipient was equal to those formed in the non-restricting recipient. Comparisons with Fig. 3.4 (A) indicate that such protection can be attributed to the functional *ardA* gene present on pLG221. These results provide a good example of the effects of alleviating restriction in the recipient *in trans*.

Overall, the yield of all three types of transconjugant formed (A, B and C) in these matings involving donors carrying R751 and a *ColIbdrd* plasmid was reduced by ~5-fold, which may be a consequence of using a donor strain supporting the transfer of two conjugative plasmids. The reduced yields cannot be explained by killing of recipients, since the number of cells was monitored during each mating experiment and found to increase slightly, as expected for a growing culture (data not shown).

The data raise the important question of whether the relative time of entry of R751 and *ColIbdrd* into the restricting recipient cells is important. *ColIb*-mediated alleviation of restriction is delayed by 6-10 minutes and it may be possible that R751 transfer is already completed by this time. This question is addressed in the following sections.

3.2.3 Rescue of R751 from restriction by *ColIbdrd* resident in the restricting recipient

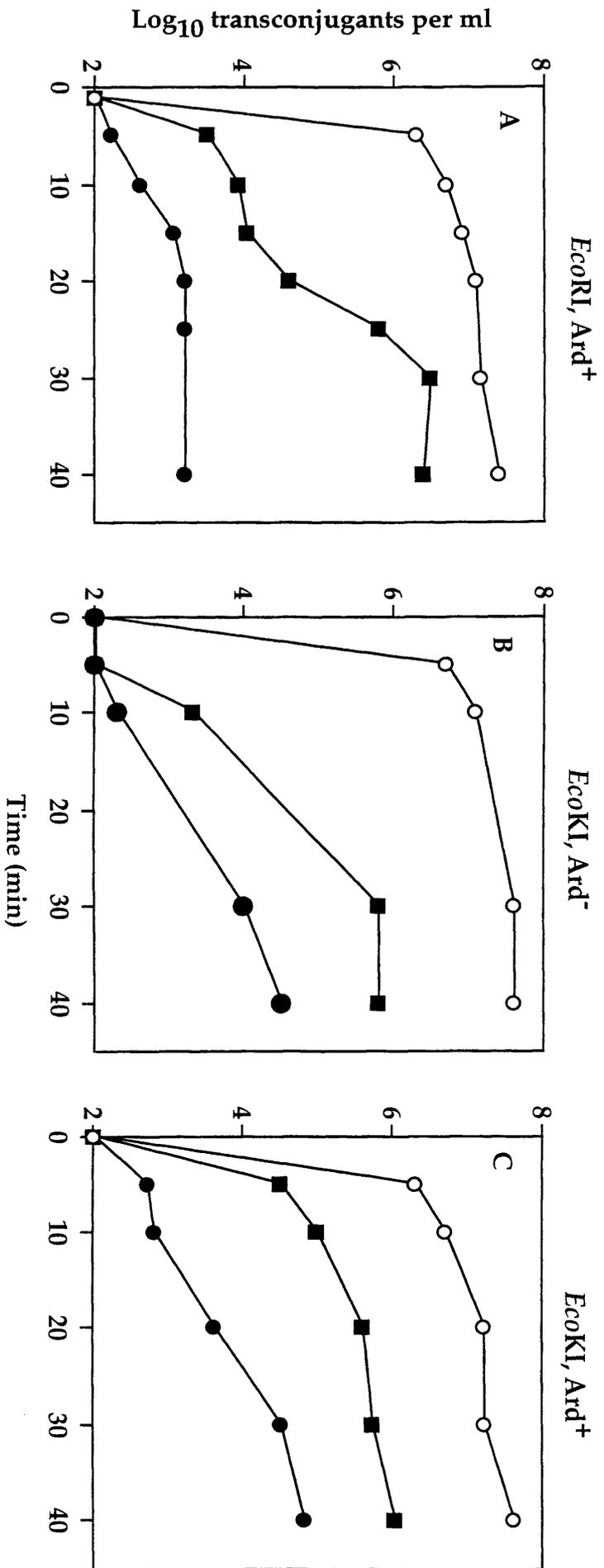
The previous section showed no significant protection of R751 from *EcoRI* or *EcoKI* restriction when the donor also transmitted the *ardA* mutant of *ColIbdrd*. This section further explores the *ardA*-independent restriction evasion mechanism by asking whether *ColIbdrd* plasmid resident in the recipient can protect transferring R751.

The experiments involved mating a donor strain harbouring R751 with either a *EcoKI* or *EcoRI* restricting recipient strain that also harboured *ColIbdrd* plasmids pLG221 (*ArdA*⁺) or pLG292 (*ArdA*⁻) (Fig. 3.6). Samples of the mating

Fig. 3.6. Transfer of plasmid R751 to a restricting recipient that harbours ColIb.

(A) R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control), GI65N(NTP14) (●, control) or GI65N pLG221, NTP14) (■). (B) R751 transconjugants from NM654R (R751) donors and recipients NM816N (○, control), GI65N (●, control) and GI65N (pLG292) (■). (C) R751 transconjugants from NM654R(R751) donors and recipients NM816N (○, control), GI65N (●, control) and GI65N(pLG221) (■). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean value of three experiment.

Fig. 3.6



cells were removed at intervals and examined for the number of R751 transconjugants.

Figure 3.6 (A) shows the results of transferring R751 to an *EcoRI* restricting recipient harbouring pLG221. In such a mating, the possibility of transferring R751 entering a restricting recipient without *ColIb* present was eliminated. Therefore, the issue of timing of plasmid transfer is not a factor. At 40 minutes, the yield of R751 transconjugants is 1,000-fold higher than the number produced when the mating system lacked pLG221. Such results suggest that *ColIbdrd* resident in the recipient does alleviate restriction by type II restriction enzymes *in trans*. The results are very different from those obtained when R751 co-transfers with pLG221, where no significant levels of *EcoRI* protection were observed (Fig. 3.3 A). In addition, these results rule out the hypothesis that the evasion mechanism operates by overwhelming the *EcoRI* restriction enzymes by substrate saturation, since such a phenomenon would require multiple rounds of transfers of *ColIbdrd* to the restricting recipient.

Figure 3.6 (B and C) also shows that both pLG292 and pLG221 conferred limited protection on incoming R751 from *EcoKI*. It is noted that pLG221 conferred more protection than pLG292, particularly during early stages. This presumably reflects the constitutive level of *ArdA* activity in an established *ColIbdrd* containing strain.

Figure 3.6 (B) shows that a resident pLG292 plasmid in the recipient conferred only modest protection on a transferring R751 plasmid. The 10-fold level of protection is similar to that observed when the two plasmids are co-transferred from the same donor strain. Therefore, expression of the *ardA*-independent restriction evasion mechanism appears to be unaffected by transfer of the plasmid. Such a finding suggests that the mechanism is expressed constitutively by *ColIbdrd* and is not dependent on transfer of the plasmid.

The modest protection conferred on R751 by pLG221 when resident in the *EcoKI* restricting recipient (Fig. 3.6 C) compared to the complete protection conferred when R751 and pLG221 are co-transferred (Fig. 3.5 A), suggests that the *ArdA* system operates optimally when specified by a transferring plasmid. Such findings are indicative that the *ardA* gene of *ColIb* is subject to zygotic induction (Chapter One, section 1.6.4).

3.2.4 Is conjugative transfer of *ColIbdrd* a prerequisite for alleviation of restriction?

A potential problem in the R751 protection experiments (section 3.2.3) is that *ColIbdrd* transfers from the *EcoRI* restricting recipients to the R751 donor. Experiments summarised in this section aimed to establish whether such transfers are a prerequisite for alleviation of type II restriction. Two strategies were adopted. The first involved a mating experiment where unmodified R751 was transferred to a restricting recipient harbouring wild type *ColIb* (pLG272). Wild type *ColIb* is naturally repressed for transfer functions and therefore transfer of the plasmid is a rare event. The transfer efficiency of pLG272 is 40,000-fold lower than that of pLG221. The second strategy involved the exploitation of the phenomenon of plasmid-encoded entry exclusion.

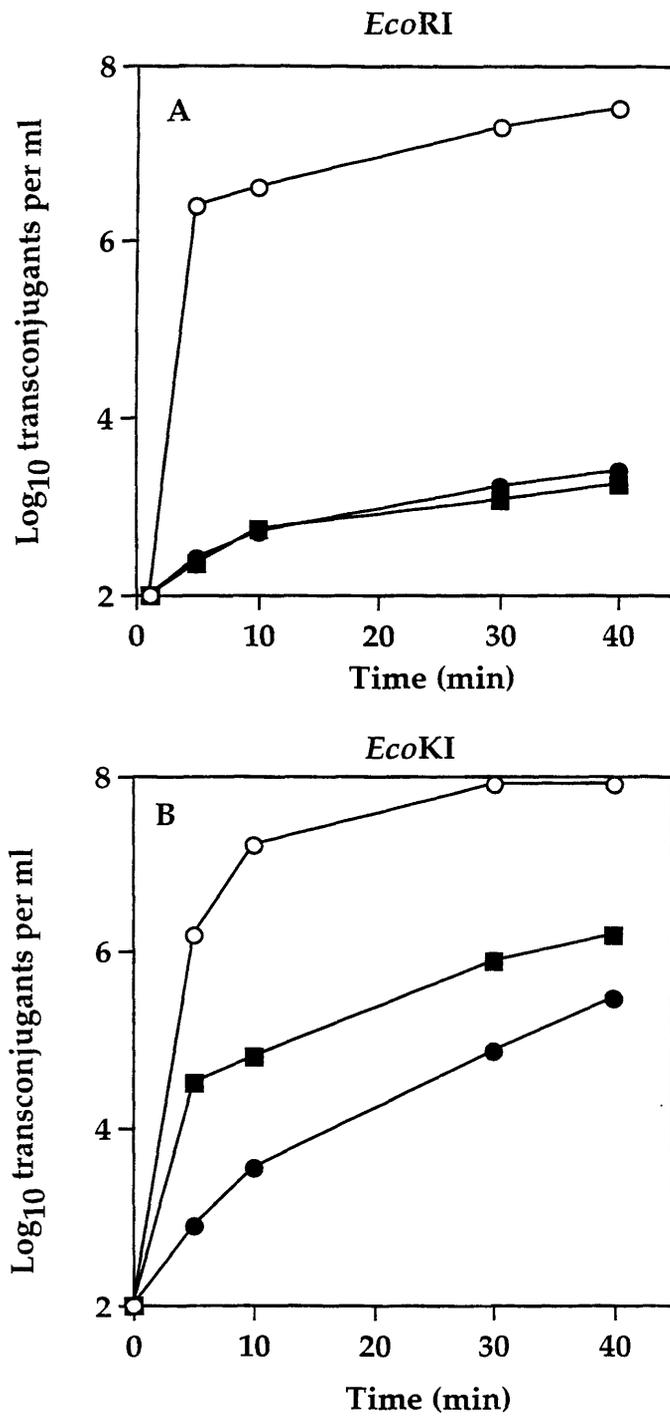
The matings described in section 3.2.3 were repeated substituting pLG221 for pLG272. Like pLG221, pLG272 also carries a Tn5 insertion that confers resistance to kanamycin although the plasmid does not carry a *drd* mutation (Howland *et al.*, 1989). Appropriate recipients were made by transferring pLG272 DNA into each strain by transformation. Strains were identified by their resistance to kanamycin. Filter mating experiments were then set up between a donor strain harbouring R751 and the restricting recipient strain harbouring pLG272. The numbers of R751 transconjugants were monitored throughout the experiment (Fig. 3.7).

Figure 3.7 (A) shows that when R751 is transferred to an *EcoRI* restricting recipient harbouring pLG272, transconjugant production was negligible,

Fig. 3.7. Transfer of R751 to a restricting recipient harbouring wild type ColIb.

(A) R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control), GI65N(NTP14) (●, control) or GI65N(NTP14, pLG272) (■). (B) R751 transconjugants from NM654R(R751) donors and recipients NM816N (○, control), GI65N (●, control) and GI65N(pLG272) (■). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean values of three experiments.

Fig. 3.7



amounting to 10^{-5} per cell at 40 minutes. Clearly, pLG272 does not confer protection on R751, demonstrating that the restriction-evasion mechanism active against type II enzymes is determined only by a *CollIbdrd* plasmid.

Figure 3.7 (B) shows the result of transferring R751 to an *EcoKI* restricting recipient harbouring pLG272. A small amount of protection was conferred on R751 by pLG272, since at 40 minutes the number of R751 transconjugants was 8-fold higher than those formed in restricting strains lacking pLG272. Interestingly, protection is initiated early, as found when pLG221 (*ArdA*⁺) was resident in the recipient cell (Fig. 3.6 C). These findings suggest that *ardA* is not under the complete control of the *CollIb* Fin system and is expressed at a low level. Read *et al.* (1992), also described similar findings where unmodified λ .0 plated with a greater efficiency on an *EcoKI* restricting strain harbouring pLG272 (e.o.p of 3.7×10^{-3}) compared to a restricting strain lacking a *CollIb* plasmid (e.o.p of 1.4×10^{-4}).

Although use of pLG272 confirmed that conjugative transfer of *CollIbdrd* is a prerequisite for alleviation by type II restriction, the data does not reveal the stage of *CollIb*'s conjugation cycle when alleviation occurs. Possibilities are mating pair formation; DNA transfer and expression of the plasmid's transfer genes. To explore this question further, the phenomenon of plasmid-encoded entry exclusion was exploited. The strategy here involved using the cloned *CollIb* exclusion genes (*eex*) in the donor to prevent transfer of pLG221 from the restricting recipient to the R751 donor. The exclusion genes of *CollIb* have been cloned giving recombinant plasmid pRR4 (Roscoe, 1996). The level of exclusion conferred by pRR4 on pLG221 is shown in Table 3.2.1. This displays the results of a 40-minute mating experiment where pLG221 was transferred to a recipient strain harbouring pRR4. The yield of pLG221 transconjugants was reduced 1,090-fold relative to the non-excluding control. However, quite surprisingly, when the recipient strain also harboured plasmid R751 as well as pRR4, the level of exclusion conferred on pLG221 was only 12-fold.

Table 3.2.1. Exclusion of pLG221 by the cloned ColIb exclusion (*eex*) genes in pRR4

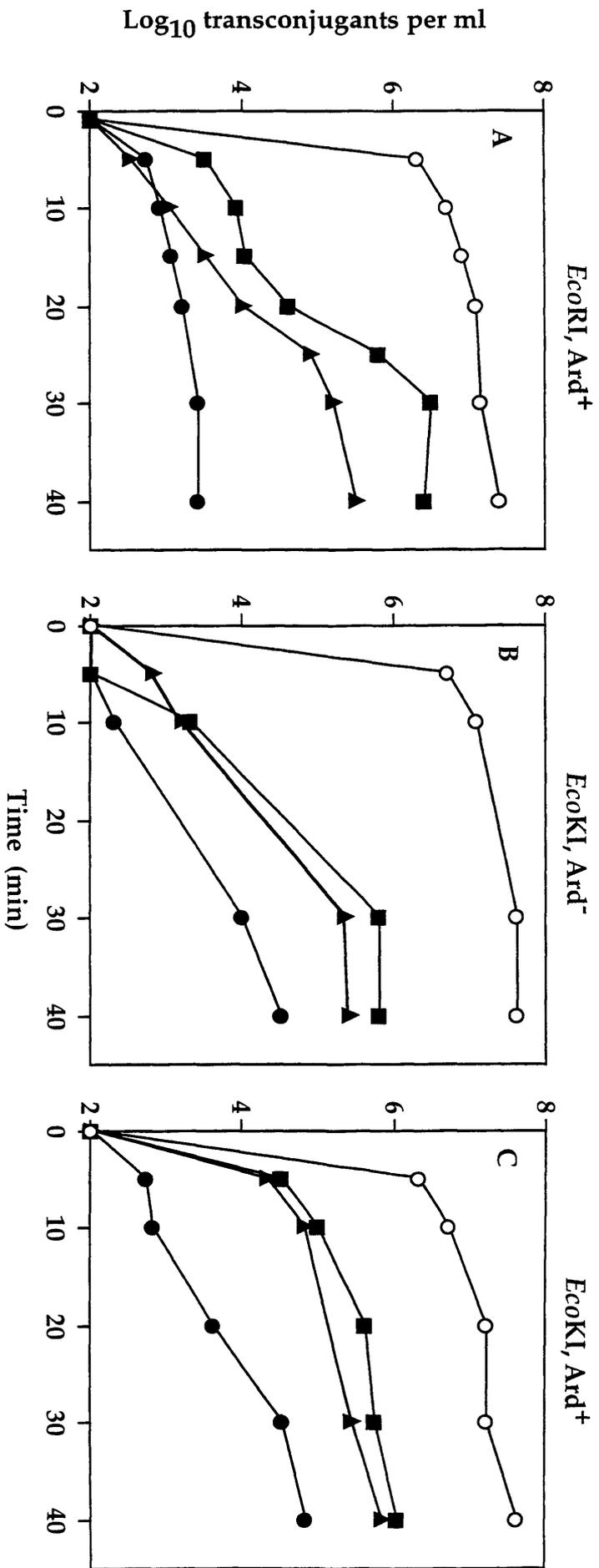
Donor	Recipient	pLG221 transconjugants/ml	*Level of exclusion
GI65N(pLG221)	GI65R	1.2×10^7	0
GI65N(pLG221)	GI65R(pRR4)	1.1×10^4	1090-fold
GI65N(pLG221)	GI65R(pRR4, R751)	1.0×10^6	12-fold

Equal numbers of donor and recipient cells were mixed together to give a concentration of 10^8 cells of each strain per ml. A 40 minute filter mating was performed at 37°C. The numbers of pLG221 transconjugants were determined by plating the mating mixture on NA plates containing Km ($25 \mu\text{g ml}^{-1}$) and Rif ($25 \mu\text{g ml}^{-1}$). Plasmid pRR4 is a recombinant plasmid carrying the cloned ColIb exclusion (*eex*) genes. *The level of exclusion conferred on transferring plasmid pLG221 by pRR4 was determined by comparing the number of pLG221 transconjugants produced in the recipient harbouring pRR4 to those produced in non-excluding strain GI65R.

Fig. 3.8. Transfer of R751 from a donor harbouring the cloned ColIb exclusion genes to a restricting recipient harbouring ColIb*drd*.

(A) R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control), GI65N(NTP14) (●, control), GI65N(NTP14, pLG221) (■), and GI65R(R751, pRR4) donors and GI65N(NTP14, pLG221) (▲) recipients. (B) R751 transconjugants from NM654R(R751) donors and recipients NM816N (○, control), GI65N (●, control) and GI65N(pLG292) (■) and NM654R(R751, pRR4) donors and recipient GI65N(pLG292) (▲). (C) R751 transconjugants from NM654R(R751) donors and recipients NM816N (○, control), GI65N (●, control) and GI65N(pLG221) (■), and donors NM654R(R751, pRR4) and recipient GI65N(pLG221) (▲). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean value of three experiment.

Fig. 3.8



This effect of pLG221 partially overcoming its own exclusion barrier is not due to plasmid R751 displacing pRR4 from the recipient cell. Throughout the mating, the fraction of recipient cells harbouring both R751 and pRR4 was measured by plating some of the mating mixture on NA plates containing Rif, Tc and Ap. The number increased slightly by 3-fold as expected for a growing culture. One possible explanation is that *Collbdrd* is able to gain entry to the defined recipient via the mating apparatus of R751. Another is that the presence of R751 in the pRR4 recipient interferes with the *Collb* exclusion system.

Figure 3.8 (B and C) shows the effects of transferring R751 from a donor strain expressing the *Collb* exclusion system to an *EcoKI* restricting recipient harbouring pLG292 (Drd^+ , $ArdA^-$) (B) or pLG221 (Drd^+ , $ArdA^+$) (C). In both cases, the exclusion system in the donor had no striking effect on the protection conferred on R751. Such findings indicate that conjugative transfer of *Collb* is not required for the *ardA*-independent function that alleviates type I restriction. The same conclusion was drawn by the data in Fig. 3.6 (B) as discussed previously.

In contrast, when R751 transferred from the donor to an *EcoRI* restricting recipient harbouring pLG221, the level of R751 protection is reduced by ~10-fold (Fig. 3.8 A). In the same experiment, pLG221 transconjugant production in the R751 donor was also reduced by 10-fold. Such results support the findings that alleviation of type II restriction requires conjugative transfer of *Collb*.

Whether the *Collb* exclusion system works by preventing mating-pair formation or DNA transfer is not known (Chapter One, section 1.3.3). Consequently, experiments involving the *Collb* exclusion genes could not be used to identify the stage of the conjugation cycle at which the plasmid-mediated restriction evasion mechanism is expressed. However, the experiments were useful in indicating that the anti-type II mechanism is not co-ordinately expressed with the transfer genes.

3.3 Summary

Four important conclusions can be drawn from the data in this chapter. First, the data are consistent with Read *et al.* (1992) in that transfer of multiple copies of *ColIbdrd* does alleviate restriction by type I and type II enzymes (Fig. 3.2). However, alleviation is apparently not due to the restriction enzymes being overwhelmed by substrate saturation (Read *et al.*, 1992), since, *ColIbdrd* alleviated restriction by both type I and type II enzymes when resident in the restricting recipient (Fig. 3.6).

Second, the *ardA*-independent anti-type I system is apparently different from that which alleviates type II restriction. This conclusion was drawn from the observation that type II restriction alleviation by *ColIbdrd* only occurs during conjugative transfer of the plasmid. In contrast, alleviation of type I restriction occurred independently of *ColIbdrd* transfer (Fig. 3.7).

Third, the *ColIb ardA* gene appears to be subject to zygotic induction. *ArdA*-mediated restriction alleviation was observed to be optimal when encoded by a transferring *ColIbdrd* plasmid rather than an equivalent plasmid resident in the restricting recipient (Fig. 3.5; Fig. 3.6 C). Such data suggest that *ardA* expression is enhanced in conjugation but do not distinguish the cell in which gene expression is amplified. This dilemma is addressed in Chapter Five.

Fourth, *ColIb* partially overcomes its own exclusion barrier when *IncP β* plasmid R751 is resident in the excluding recipient strain. One hypothesis that may explain this phenomenon is that *ColIb* utilises the mating apparatus of R751 to gain entry to the excluding recipient. Such a scenario would indicate that the *ColIb* exclusion mechanism operates by preventing mating-pair formation between two cells harbouring the same plasmid. However, such a hypothesis is inconsistent with the work of Hartskeerl and Hoekstra (1985). They found evidence to suggest that the *ColIb* exclusion system works by inhibiting DNA transfer between two donor cells (Chapter One, section 1.3.3).

Due to time constraints, this concept of interactions between conjugation systems was not pursued further. One way of testing the hypothesis would be to construct a specific mutant of R751 unable to express conjugative pili. Such a mutant would be expected to no longer provide the route of entry for the ColIb plasmid to the excluding recipient cell.

These summary points are followed up in subsequent chapters and general discussion.

Chapter Four

Construction of an *oriT* Mutant of *ColIbdrd* and its Conjugative Properties

4.1 Introduction

Data presented in the previous chapter suggests that *ColIbdrd* encodes a restriction-alleviation function, which is active against type II restriction enzymes and is expressed during conjugative transfer of the plasmid. There are three main stages of *ColIb*'s conjugation cycle when such a function could be expressed: during mating-pair formation, DNA transfer or throughout the cycle. To distinguish between these three stages, it was necessary to construct a mutant of *ColIb* defective in self-transfer by conjugation, yet proficient in forming mating pairs and in expressing the full complement of plasmid proteins. The purpose of this chapter is to describe the strategy employed in creating such a mutant.

As described in chapter one, the origin of transfer (*oriT*) of a plasmid is a *cis*-acting element required for DNA transfer by conjugation. Transfer is initiated at the *nic* site within the *oriT* by specific endonucleolytic cleavage of a single strand of the plasmid (Lanka and Wilkins, 1995). The main strategy for creating a mutant of *ColIbdrd*, defective in self-transfer, was to delete the *nic* region from the *oriT* of the plasmid. *ColIbdrd* is ~93 kb in size and specific alterations of sequences stretching only a few base pairs in length, internal of such a sizeable DNA genome, are not practical without using intermediate small recombinant plasmids. Therefore, before any alterations to the sequence of the *ColIbdrd oriT* could be made, the locus had to be cloned and its *nic* sequence determined. The latter was achieved by DNA sequencing.

The *nic* sequence was deleted from the cloned *oriT* by a procedure involving PCR-amplification. A selectable marker gene was then inserted into

the sequence and the disrupted *oriT* transferred back to the *CollIbdrd* plasmid by homologous recombination. The method used to replace the *oriT* of *CollIbdrd* with a selectable mutation was based on a gene replacement strategy described by Russell *et al.* (1989). The recombinant plasmid, carrying the disrupted *oriT* and selectable marker, was first linearised by restriction endonuclease digestion. This DNA was then introduced by transformation into an *E. coli recD* strain harbouring *CollIbdrd*. Under normal circumstances, linear DNA is subject to degradation by intracellular nucleases. However, the advantage of using an *E. coli recD* mutant is that the exonuclease activities of the RecBCD enzyme are abolished without affecting the enzymes ability to act as a DNA helicase in initiating RecA-mediated homologous recombination. The disrupted *oriT*, within the recombinant plasmid is integrated by a double crossover event spanning the selectable marker. The remainder of the recombinant plasmid is eventually destroyed. The minimum amount of homology that is required for efficient recombination is 44-90 nucleotides for RecBC and RecF dependent pathways (Shen and Huang, 1986).

Other experiments described in this chapter include examining the conjugative properties of the *CollIbdrd* Δ *nic* plasmids.

4.2 Results

4.2.1 Cloning and determination of the nucleotide sequence of the *CollIb oriT* locus

The *CollIb oriT* is located adjacent to the leading region of the plasmid on a 1.6 kb *PstI* fragment [Chapter One, Fig 1.1] (Rees *et al.*, 1987). This fragment has already been cloned in this laboratory and inserted into cloning vector pUC19 (Roscoe, 1996). The resulting recombinant plasmid (pCL2) was used initially as a template for determining the nucleotide sequence of *CollIb oriT*. Sequence reactions were performed as described in Chapter Two using M13/pUC universal forward and reverse primers. However, the initial sequence data obtained were very ambiguous. Consequently, the integrity of

the pCL2 template was re-examined, firstly by restriction enzyme analysis and secondly by PCR-amplification.

A series of 20-cycle PCR reactions was set up using universal forward and reverse primers with template pCL2. The temperature at which the primers were annealed to the pCL2 template was varied from 49-60°C. The resulting extension products produced in each reaction were examined by gel electrophoresis and in all cases unpredictable DNA bands of varying sizes were observed (data not shown). Such results are indicative of either non-specific annealing of primers or of a mixed population of template. However, restriction data for pCL2 were not indicative of the latter possibility.

To overcome the problem, the 1.6 kb *Pst*I insert was removed from pCL2 by restriction enzyme digestion and the fragment isolated from a 0.8% agarose gel after separation from pUC19 by electrophoresis. The fragment was purified from the agarose using the QIAEX II gel extraction kit (Qiagen) and ligated back into the *Pst*I site of a separate batch of pUC19 molecules forming recombinant pNA1. Isolation of pNA1 was achieved by transferring the ligated DNA by transformation into the *E. coli* Lac⁻ strain, DH5 α . Transformant colonies containing pNA1 were distinguished from those that had gained re-ligated pUC19 vector by their white appearance on LA plates containing X-gal, IPTG and Ap (Sambrook *et al*, 1989; see Chapter Two).

In order to establish the orientation of the 1.6 kb *Pst*I insert of pNA1 within the pUC19 vector, the recombinant plasmids were examined by restriction enzyme digestion. All four of the recombinant plasmids isolated were digested with *Cla*I and *Eco*RI, where the *Cla*I site is internal to the 1.6 kb insert and the *Eco*RI site adjacent to one of the *Pst*I sites. Three of the four recombinants tested had the orientation shown in Fig. 4.1 A, and one the opposite orientation. The latter plasmid was subsequently designated pNA1a.

Recombinant plasmids pNA1 and pNA1a were both tested for their ability to be mobilised by ColIb*drd* plasmid pLG221 from a *recA* donor. A *recA* donor strain was used in all mobilisation experiments summarised in Table

Fig. 4.1. Schematic diagram of the construction of pNA2

(A) Cleavage of pNA1 with *Pst*I and *Cla*I to remove the 1.15 kb of DNA containing ColIb *oriT*. The remaining 3.1 kb *Pst*I-*Cla*I fragment is discarded.

(B) Cleavage of cloning vector pIC19H at its *Pst*I and *Cla*I sites to obtain the 2.6 kb linearised vector. The small *Pst*I-*Cla*I fragment from the polylinker is lost through purification of the larger fragment.

(C) Insertion of the 1.15 kb *Pst*I-*Cla*I fragment from pNA1 into 2.6 kb pIC19H vector to give pNA2. (➡) represents the orientation (5' - 3') and annealing site for M13/pUC universal forward (UF) and reverse (UR) primers.

Fig. 4.1

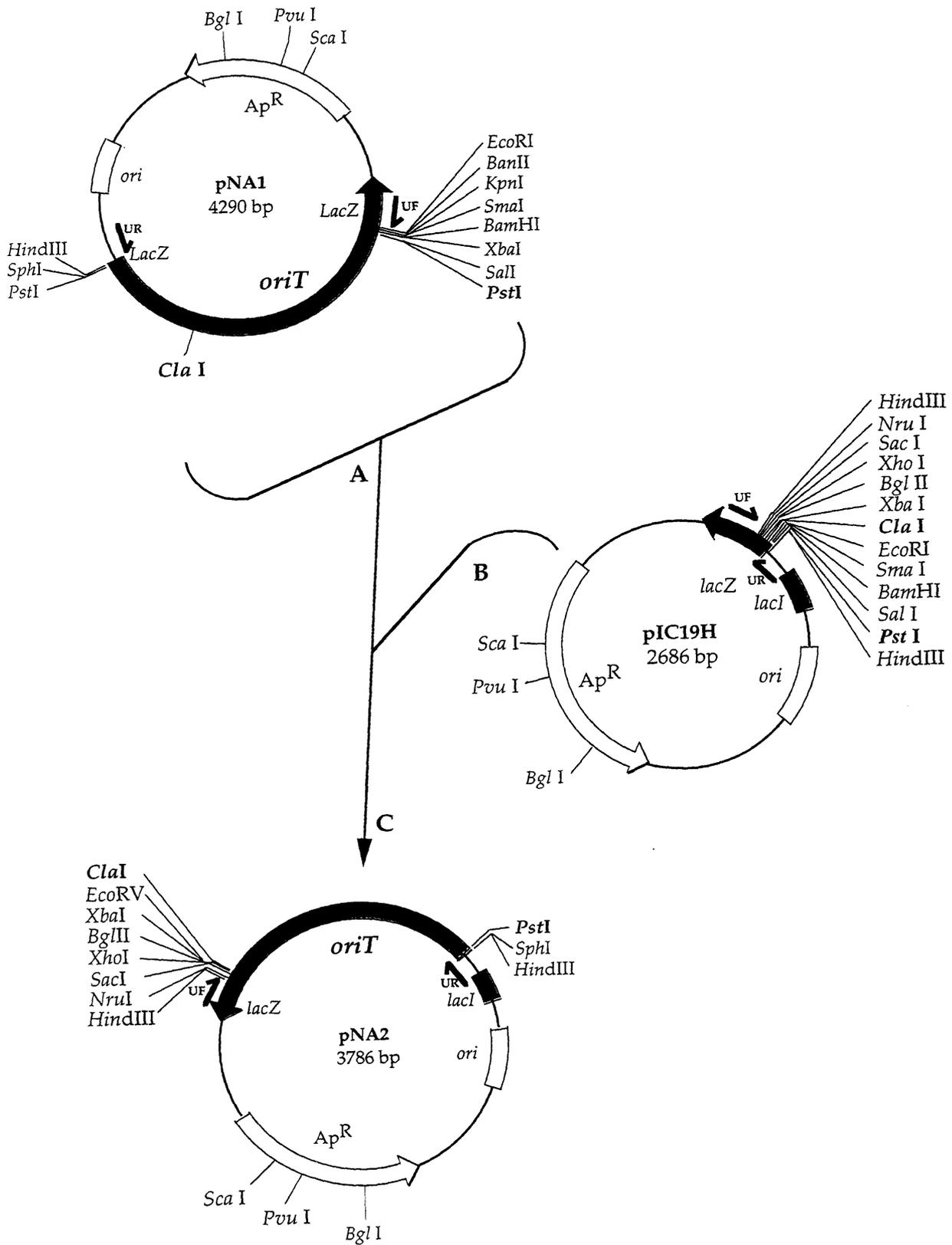


Table 4.2.1. Mobilisation of recombinant plasmids carrying the ColIb *oriT* locus

Donor Stain	Transconjugants / ml		*Mobilisation Frequency of Recombinant
	Recombinant	Helper Plasmid	
JO8(pLG221, pNA1)	2.0×10^6	3.0×10^6	0.66
JO8(pLG221, pNA1a)	1.1×10^6	2.0×10^6	0.55
JO8(pLG221, pNA2)	8.0×10^6	1.1×10^7	0.73
JO8(pLG221, pNA3)	4.0×10^6	1.4×10^7	0.28
JO8(pLG221, pNA7)	< 10	7.0×10^7	< 10^{-4}
JO8(pLG221, pNA8)	< 10	1.3×10^7	< 10^{-4}
JO8(pNA11, pNA2)	1.9×10^7	< 10	1.0
JO8(pNA12, pNA2)	1.1×10^7	< 10	1.0

The frequency of transfer of each plasmid was determined after a 40 minute filter mating at 37°C, using JO8 as the donor host strain and GI65N as the recipient. Recombinant plasmids pNA1, pNA1a, pNA2, pNA3, pNA7 and pNA8 all confer resistance to ampicillin. Helper plasmids pLG221, pNA11 and pNA12 all confer resistance to kanamycin. Plasmids pNA11 and pNA12 both carry a *cis*-acting lesion within the *oriT* which prevents their self-transfer. However, both plasmids still specify all other transfer functions. Transfer of recombinant plasmids were measured by plating the mating mixture on NA plates containing Ap (100 µg ml⁻¹) and Nal (25 µg ml⁻¹) and transfer of the helper plasmid measured by plating on NA plates containing Km (25 µg ml⁻¹) and Nal (25 µg ml⁻¹). *Mobilisation frequency is the number of mobilisable plasmids transferred per conjugative plasmid.

4.2.1, to prevent co-integrate formation between ColIb plasmids and recombinants carrying the ColIb *oriT* sequence. Test strains were made by transferring the recombinant plasmids by transformation to JO8(pLG221) competent cells. Donor strains JO8(pLG221, pNA1) and JO8(pLG221, pNA1a) were mated with Nal^{R} recipient strain, GI65N for 40 minutes on filters. The results of these transfers are presented in Table 4.2.1. Both pNA1 and pNA1a were found to be mobilised with an efficiency of 0.66 and 0.55, respectively, relative to the transfer of pLG221. Such results confirm that both recombinant plasmids contain the ColIb *oriT* and that its activity as a *cis*-acting site for the initiation of DNA transfer is unaffected by orientation.

To reduce the number of reactions required to determine the complete nucleotide sequence of the 1.6 kb *Pst*I insert within pNA1, the ColIb *oriT* was subcloned further. The 1.15 kb *Pst*I-*Cla*I fragment found within pNA1 was isolated and ligated into cloning vector pIC19H, forming pNA2 (Fig. 4.1). Cloning vector pIC19H is identical to pUC19 except it has a different set of cloning sites available within the polylinker (Marsh *et al.*, 1984).

Mobilisation of pNA2 from a *recA* host (JO8) with a frequency of 0.73 relative to the transfer of pLG221 confirmed that the plasmid was *oriT*⁺ (Table 4.2.1). The suitability of pNA2 as a template for DNA sequencing was further confirmed by PCR-amplification and restriction enzyme analysis.

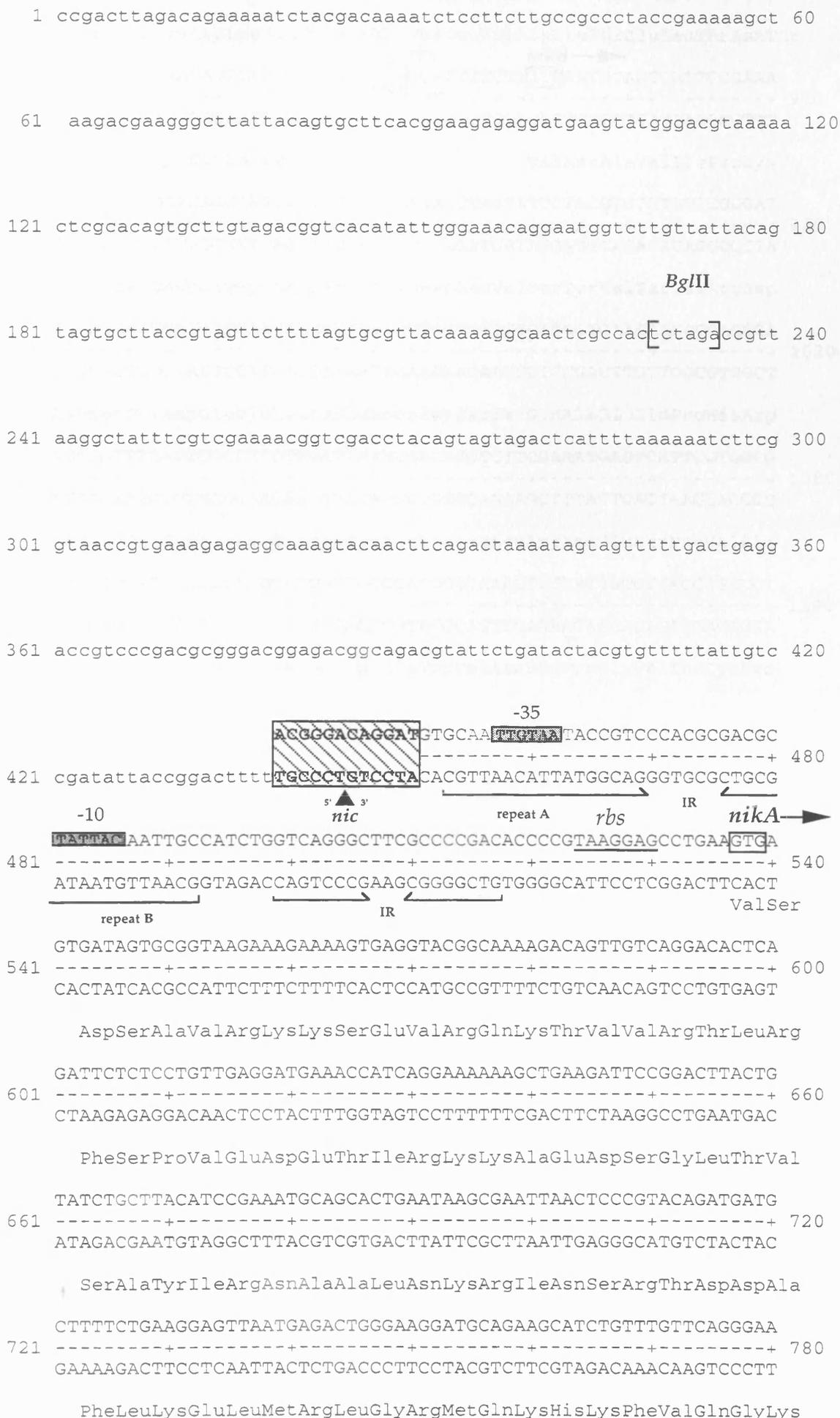
Initial sequence reactions were set up using universal forward and reverse primers. However, to obtain sequence data for both DNA strands, two new primers were designed (primer B and primer C; see Chapter Two). The sequence data obtained from pNA2 using universal forward and reverse primer and primer B and primer C, are presented in Fig. 4.2.

The location of the *nic* site in IncI1 plasmid R64 is known, as is the nucleotide sequence of the surrounding region (Furuya and Komano, 1991). Alignment of nucleotide sequences of a number of plasmids shows that R64 *nic* region belongs to a family with a consensus sequence of 5'-YATCCTG▼Y-3', where ▼ indicates the specific cleavage site (Lanka and Wilkins, 1995). The

Fig. 4.2. 1,151-bp DNA sequence containing the ColIb *oriT* and part of the *oriT* operon

Features shown include the *nic* region , where ▲ represents the site of cleavage by the relaxase enzyme; the two  start codons of *nikA* and *nikB* and their associated ribosome binding sites (*rbs*) which are underlined; and the  stop codon of *nikA*. Amino acid sequences deduced from the DNA sequence of the *nikA* and *nikB* reading frames are presented, where the start of each translated amino acid is in uppercase. Relevant restriction sites are indicated. Inverted repeat sequences are shown ( ). Inverted repeats A and B represent the site at which the *nikA* polypeptide binds to the DNA. Putative promoter sequences (-35 and -10 regions) are boxed . Nucleotides 1-437, presented in lower case, represent the leading region strand that is transferred to the recipient cell first during conjugation. All features shown, except the *nic* sequence, were identified by sequence comparisons to the published *oriT* sequence of the IncI1 plasmid R64 (Furuya and Komano, 1991; EMBL database accession number D90273). The *nic* sequence was determined by comparing the sequence data to the consensus 5'-YATCCTG▼Y-3' (Lanka and Wilkins, 1995; Fig. 1.2).

Fig. 4.2



```

781 AAAGAACCGGTGACAAGGAATACGCAGAAGTGCTTGTGGCCATAACCGAACTTACAAATA
-----+-----+-----+-----+-----+-----+-----+ 840
TTTCTTGGCCACTGTTCCCTTATGCGTCTTCACGAACAACGGTATTGGCTTGAATGTTTAT

ArgThrGlyAspLysGluTyrAlaGluValLeuValAlaIleThrGluLeuThrAsnThr
      rbs                                nikB →
841 CATTGCGTAAACAGTTAATGGAAGGTTGAGGATTCTCCTGTGAATGCAGTCATTCCGAAA
-----+-----+-----+-----+-----+-----+-----+ 900
GTAACGCATTTGTCAATTACCTTCCAACCTCCTAAGAGGACACTTACGTCAGTAAGGCTTT

LeuArgLysGlnLeuMetGluGly*                               ValAsnAlaValIleProLys

901 AAGAGAAGGGACGGCAAGTCTTCGTTTCGAAGACCTAGTATCCTACGTCTCTGTCCGGGAT
-----+-----+-----+-----+-----+-----+-----+ 960
TTCTCTTCCCTGCCGTTCAGAAGCAAGCTTCTGGATCATAGGATGCAGAGACAGGCCCTA

LysArgArgAspGlyLysSerSerPheGluAspLeuValSerTyrValSerValArgAsp

961 GACATGACCGATGAGGAACTGGATTTATCTTCTTCTTCGCAGGCTGAACAACCGCACCGA
-----+-----+-----+-----+-----+-----+-----+ 1020
CTGTACTGGCTACTCCTTGACCTAAATAGAAGAAGAAGCGTCCGACTTGTGGCGTGGCT

AspMetThrAspGluGluLeuAspLeuSerSerSerSerGlnAlaGluGlnProHisArg

1021 AGCCGTTTCAGTCGCCTTGTTGATTATGCGACACGTCTTCGAAATGAGTCATTTCGTGGCG
-----+-----+-----+-----+-----+-----+-----+ 1080
TCGGCAAAGTCAGCGGAACAATAACGCTGTGCAGAAGCTTACTCAGTAAGCACCGC

SerArgPheSerArgLeuValAspTyrAlaThrArgLeuArgAsnGluSerPheValAla

1081 CCGGTTGATGTCATGAAGGACGGCTGCGCATGGGTCAACTTTTATGGTGTACCTGCCCT
-----+-----+-----+-----+-----+-----+-----+ 1140
GGCCAACTACGTACTIONTTCCTGCCGACGCGTACCCAGTTGAAAATACCACAGTGGACGGGA

ProValAspValMetLysAspGlyCysAlaTrpValAsnPheTyrGlyValThrCysPro

1141 CATAACTGTAT
-----+----- 1151
GTATTGACATA

HisAsnCys

```

nucleotide sequence determined for the 1.15 kb *PstI-ClaI* fragment of pNA2 was found to have 97% identity over about 924 bp with the corresponding region surrounding *oriT* of R64 (EMBL database accession number D90273). The *nic* region of ColIb was identified by alignment of the sequence with the consensus sequence described by Lanka and Wilkins (1995). In fact, the 12-bp *nic* region identified for ColIb is identical to the *nic* region of plasmid R64. Other features identified within the sequence of the 1.15 kb *PstI-ClaI* fragment of pNA2 were also identified within the published sequence of R64 (Fig. 4.2).

4.2.2 Strategy employed to delete the *nic* sequence from the cloned ColIb *oriT*

Once the sequence of the ColIb *nic* site had been determined, a strategy involving PCR-amplification was employed to delete it from the cloned *oriT* (pNA2) as summarised in Fig. 4.3.

First, two inverse oligonucleotide primers were designed which were homologous to the regions of DNA flanking the *nic* region (primer 5 and primer 6, Fig. 4.3, **A**). A unique *EcoRI* restriction enzyme target sequence was added onto the 5' terminus of each primer. Hence, during amplification of the template by an inverse PCR reaction, 14 bp overlapping the *nic* region are lost and the target sequence for the *EcoRI* restriction enzyme is incorporated. In addition, these primers were also designed to avoid disruption of the *oriT* operon promoter found 12-bp upstream of the *nic* site (Fig. 4.3, **A**).

Before any amplification of the pNA2 template was carried out the PCR protocol was optimised as described in Chapter Two, section 2.6.1. The results of this optimisation process are presented in Fig. 4.3 (**B**). Specific comments were primer 5 and primer 6 have an optimal annealing temperature of 53°C (lane 3 and 4) and that 4 minutes is required for the whole of the pNA2 template to be amplified by *Taq* polymerase (lane 5).

The final PCR-amplification of pNA2 and the subsequent removal of the ColIb *nic* region from the plasmid was achieved using 50 ng of pNA2

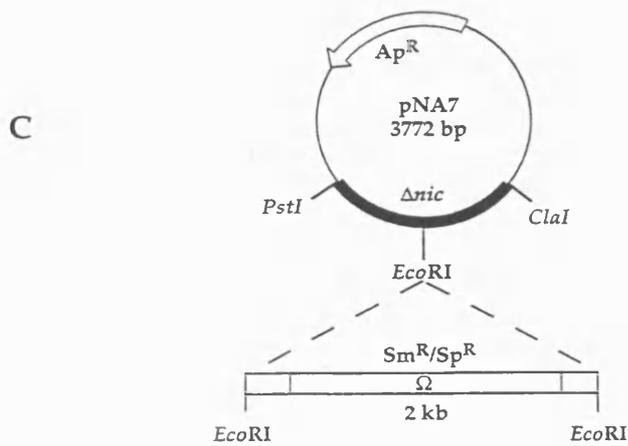
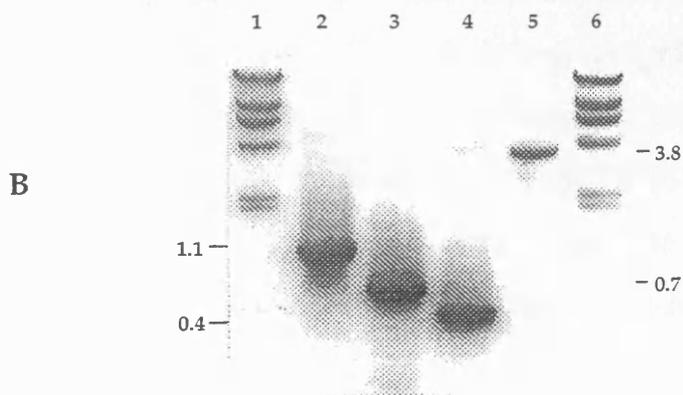
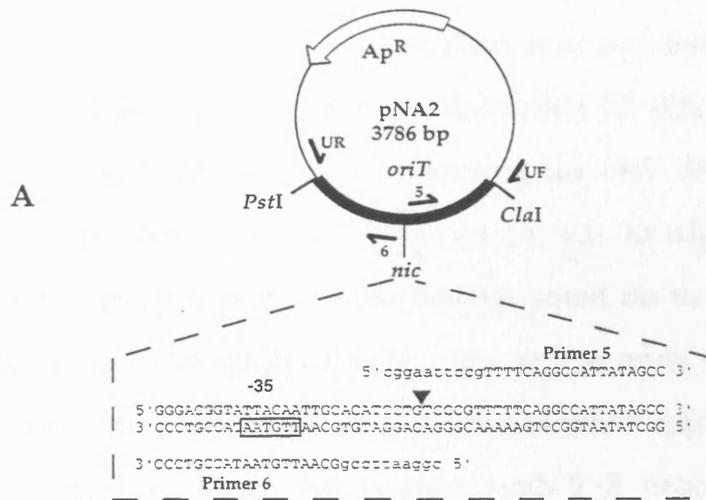
Fig. 4.3. Schematic diagram of the key stages in the construction of a ColIb *oriT* sequence deleted of the *nic* region

(A) Relationship between primers used for inverse PCR-amplification of pNA2 and the *nic* region. ▼ indicates the *nic* site; the boxed sequence is the -35 region of the putative promoter of the relaxase operon. Primer 5 extends into the leading region. Primer nucleotides shown in lower case are non-homologous to the ColIb region and contain the *EcoRI* target hexanucleotide (5'-gaattc-3'). (→) represents the direction (5' - 3') and annealing sites of primers 5 and 6 and the M13/pUC universal forward (UF) and reverse (UR) primers.

(B) Gel electrophoresis of PCR-products of pNA2. Molecular weight markers (λ xHind III) are in lanes 1 and 6. Primers used (2) UF and UR; (3) 6 and UR; (4) 5 and UF; (5) 5 and 6. The size of each extension product is indicated in kb.

(C) Cleavage of the linear PCR product with *EcoRI* and ligation gave pNA7 deleted of the *nic* region. The omega (Ω) intersposon was ligated into the unique *EcoRI* site, forming pNA8.

Fig 4.3



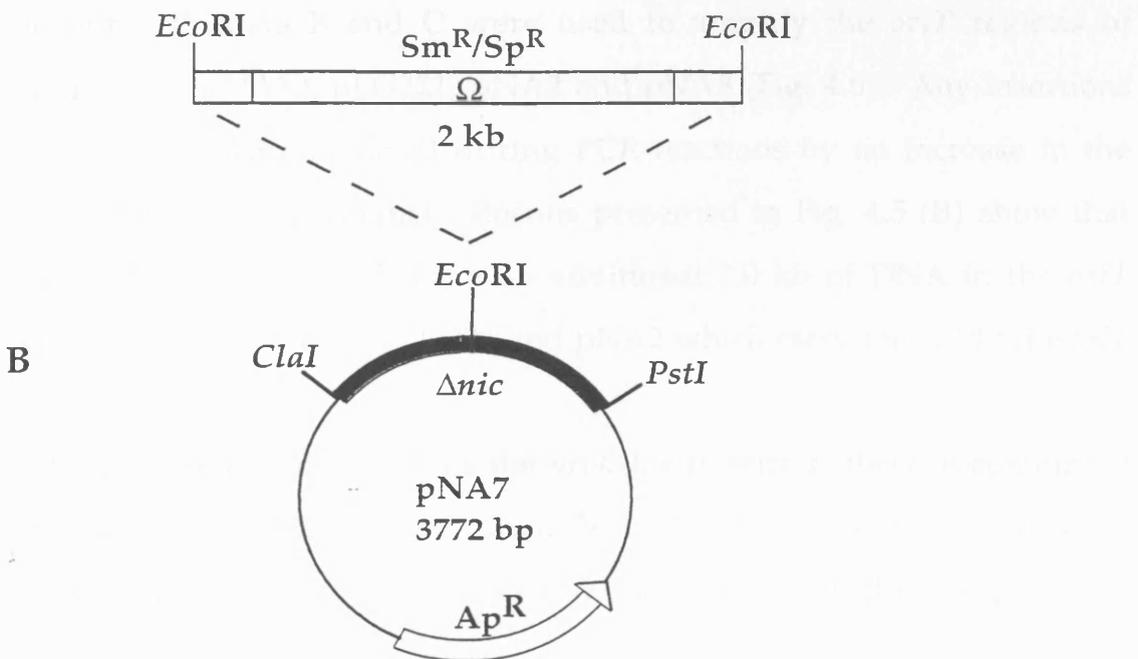
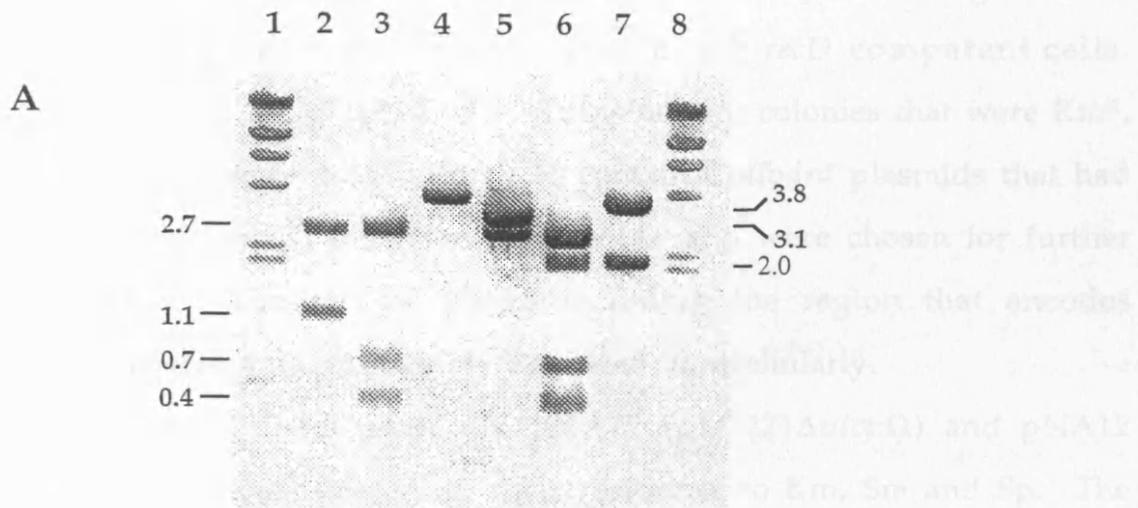
template; 20 pmol of both primer 5 and 6; 1 unit of proof-reading *Taq* polymerase (BIO-X-ACT, Bioline UK Ltd); 5 µl of 11.1 × PCR buffer; 40 µl of H₂O and the following PCR program: 95°C - 3 minutes (1 cycle); 95°C - 45 seconds, 53°C - 30 seconds, 72°C - 4 minutes (35 cycles) and 72°C - 4 minutes (1 cycle).

The resulting 3.8 kb extension product was isolated and purified from a 0.8% agarose gel using the QIAEX II gel extraction kit (Qiagen). The next stage in isolating a recombinant plasmid containing the *oriT* deletion mutant was to re-ligate the linear PCR product. One option was to digest and re-ligate the *EcoRI* restriction target sites that were incorporated on to the ends of the PCR-product. However, restriction target sites at the ends of PCR-products are difficult to digest. To avoid this complication another approach was used. One key feature of using a polymerase enzyme with 5'-3' proofreading exonuclease activity is that a large population of the linear product have flush ends. By exploiting this property, the PCR-product was re-circularised by blunt-end ligation. The DNA was then transferred by transformation to DH5α. Colonies that had gained a circularised plasmid were detected by their resistance to ampicillin. The deletion of the *nic* sequence from the resulting recombinant plasmid, pNA7, was confirmed by the inability of the plasmid to be mobilised by plasmid pLG221 (Table 4.2.1).

4.2.3 Construction of pNA8 containing the omega (Ω) intersposon

The next step involved the insertion of a selectable antibiotic resistance gene carried by the omega (Ω) intersposon into pNA7. The 2.0 kb fragment carries a streptomycin-spectomycin resistance gene flanked by inverted repeats with transcription and translation termination signals (Prentki and Krisch, 1984). The Ω fragment was released from pHP45Ω by *EcoRI* digestion and ligated into the *EcoRI* site of plasmid pNA7 to give pNA8 (Fig. 4.3, C). Like pNA7, pNA8 is also non-mobilisable by pLG221 (Table 4.2.1). Finally, further to confirm the structure of the mutant *oriT* locus, pNA7 and pNA8 were examined by restriction enzyme analysis (Fig. 4.4).

Fig. 4.4. Gel electrophoresis of restriction fragments of pNA7 and pNA8.



(A) Samples in lanes are (1) λ x *HindIII*; (2) pNA7 x *PstI/ClaI*; (3) pNA7 x *PstI/ClaI/EcoRI*; (4) pNA7 x *EcoRI*; (5) pNA8 x *PstI/ClaI*; (6) pNA8 x *PstI/ClaI/EcoRI*; (7) pNA8 x *EcoRI*; (8) λ x *HindIII*. Fragments sizes are presented in kb.

(B) Physical map of pNA7 and the 2.0 kb Ω fragment, which on ligation into the *EcoRI* site forms pNA8.

4.2.4 Construction of a *CollIbdrd* mutant lacking the *nic* region

The *oriTΔnic::Ω* mutation was transferred by recombination from pNA8 to the *CollIbdrd* plasmids pLG221 (ArdA⁺) and pLG292 (ArdA⁻). The 3.1 kb *Pst*I-*Cla*I fragment was released from pNA8 by restriction enzyme cleavage and the DNA transferred by transformation into *E. coli recD* competent cells, DL307(pLG221) and DL307(pLG292). Transformant colonies that were Km^R, Sm^R, Sp^R and Ap^S were inferred only to contain *CollIbdrd* plasmids that had acquired the *oriTΔnic::Ω* mutation from pNA8 and were chosen for further characterisation. The rest of pNA8 including the region that encodes ampicillin resistance was presumably degraded intracellularly.

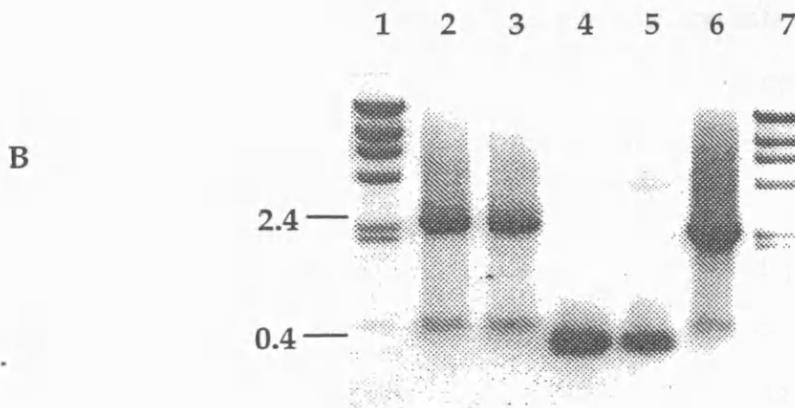
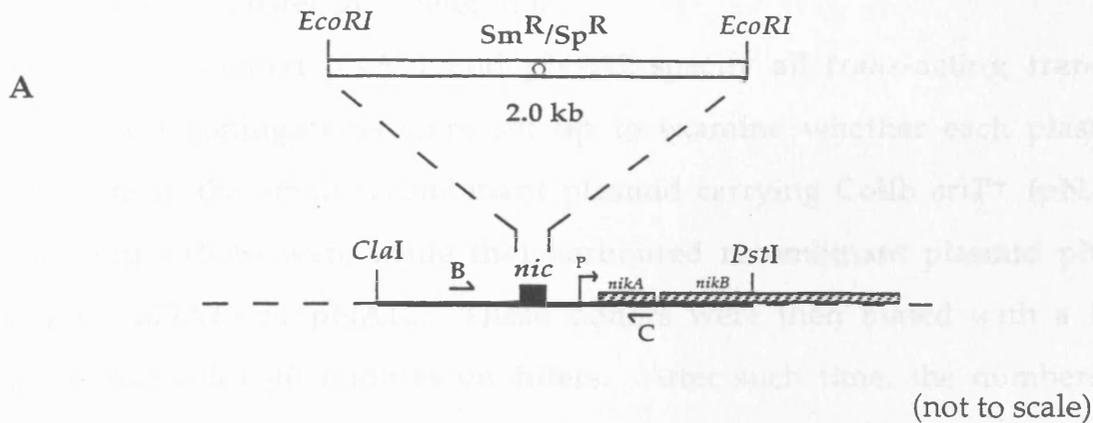
Recombinant *CollIb* plasmids pNA11 (pLG221Δ*nic::Ω*) and pNA12 (pLG292Δ*nic::Ω*) were isolated by their resistance to Km, Sm and Sp. The nature of the *oriT* within each plasmid was characterised by PCR-amplification. Primers B and C were used to amplify the *oriT* regions of plasmids pNA11, pNA12, pLG221, pNA2 and pNA8 (Fig. 4.5). Any insertions within this region were detected during PCR-reactions by an increase in the size of the amplification product. Results presented in Fig. 4.5 (B) show that pNA11, pNA12 and pNA8 all have an additional 2.0 kb of DNA in the *oriT* locus, which is not present in pLG221 and pNA2 which carry the wild type *oriT* locus.

Alternatively, the nature of the *oriT* locus within these recombinant plasmids could have been confirmed by a Southern blot hybridisation experiment. However, due to the time constraints of this project, this additional work was not undertaken.

4.2.5 Conjugative properties of pNA11 and pNA12

Recombinant plasmids pNA11 and pNA12 were tested for self-transfer by conjugation. Test strains were made by transferring pNA11 and pNA12 by transformation into GI65N competent cells and selecting for Km^R, Sm^R and

Fig. 4.5. Analysis of the nature of the *oriT* within pNA11 and pNA12 by PCR-amplification



(A) Relationship between primers B and C used for PCR-amplification of the ColIb *oriT* locus. (→) represents the direction (5'-3') and annealing sites of primer B and C. Plasmid pNA2 and pLG221 carry an undisrupted *oriT*. Plasmids pNA11, pNA12 and pNA8 are deleted of the *nic* region and contain the 2.0 kb Ω insertion. Also shown is the *oriT* operon promoter (P) and *nikA* and *nikB*. The region between the annealing sites of primer B and C is 0.4 kb.

(B) PCR reactions were performed using primer B and primer C with plasmids pNA11, pNA12, pLG221, pNA2 and pNA8. Gel electrophoresis of the resulting extension products and their sizes in kb are presented. Lanes 1 and 7 contain λ \times *Hind*III molecular weight markers. Templates used were (2) pNA11; (3) pNA12; (4) pLG221; (5) pNA2; (6) pNA8. Faint bands at the ends of lane 2, 3 and 6 are unused primer.

Sp^R. Conjugation experiments were then set up on filters using a Rif^R recipient strain (GI65R). After 40 minutes, the numbers of recipients that had acquired pNA11 or pNA12 were scored by plating on NA plates containing Rif and Km. No transconjugants were detected ($< 2 \times 10^{-6}$ relative to the transfer efficiency of pLG221). Hence, both pNA11 and pNA12 were found to be completely defective in self-transfer by conjugation.

To confirm that pNA11 and pNA12 specify all *trans*-acting transfer functions test conjugations were set up to examine whether each plasmid could mobilise the small recombinant plasmid carrying ColIb *oriT*⁺ (pNA2). Donor strains (JO8) were made that harboured recombinant plasmid pNA2 and either pNA11 or pNA12. These donors were then mated with a Rif^R recipient strain for 40 minutes on filters. After such time, the numbers of pNA2 transconjugants were examined by plating on NA plates containing Rif and Ap. Both pNA11 and pNA12 were able to mobilise pNA2 with an efficiency of 1.0 from a *recA* host relative to pLG221 directed mobilisation (Table 4.2.1). In summary pNA11 and pNA12 are inferred to specify all of the *trans*-acting proteins required for conjugation but to contain a *cis*-acting lesion preventing self-transfer. These results also confirm that each plasmid has retained its Drd⁺ phenotype.

Finally, the ArdA phenotypes of pNA11 and pNA12 were examined. The test involved infecting an *EcoKI* restricting *E. coli* strain (GI65N) with a λ , *vir* lacking *EcoKI* modification (λ *vir.0*) and examining the number of plaques formed. In strains harbouring pNA11, the number of plaque forming units produced by λ *vir.0* was 3.8×10^8 ml⁻¹ compared to 1.7×10^6 ml⁻¹ formed on GI65N alone. In contrast, the number of plaque forming units produced on strains harbouring pNA12 was 8.0×10^7 ml⁻¹ (Appendix I). These results confirm that pNA11 has an ArdA⁺ and pNA12 an ArdA⁻ phenotype.

4.3 Discussion

Results in this chapter detail the successful construction of a specific mutation within the *oriT* of ColIb*drd*. The resulting mutants, pNA11 (Drd⁺, ArdA⁺) and pNA12 (Drd⁺, ArdA⁻), were completely defective in self-transfer by conjugation. However, the ability of pNA11 and pNA12 to specify all other transfer functions was confirmed by the mobilisation of an *oriT*⁺ recombinant plasmid with a maximal efficiency (Table 4.2.1). In summary, pNA11 and pNA12 specify all the *trans*-acting proteins required for conjugation, yet, contain a *cis*-acting lesion within the *oriT* preventing self-transfer.

Several notable observations were made during the construction of these mutants and they are described below.

Nucleotide sequence similarities are only found between the *oriT* structures of closely related plasmids (Furuya and Komano, 1991). This is confirmed as the nucleotide sequence of the *oriT* of ColIb only shares significant sequence similarities with the published *oriT* sequence of closely related IncI1 plasmid R64 (section 4.2.2). Significant sequence similarities exist between the *nic* regions of different conjugative plasmids (Lanka and Wilkins, 1995). These similarities have been organised into three groups presented in Chapter One, Fig. 1.2. These findings are further supported here, as the *nic* region of ColIb bears 100% identity with the corresponding region of R64, which in turn belongs to a family of *nic* regions harbouring the consensus 5'-YATCCTG▼Y-3'. Other members of this group include plasmids from the incompatibility group P and sequences from the Ti plasmid of *Agrobacterium tumefaciens* (Lanka and Wilkins, 1995).

The location and sequence of the ColIb *oriT nic* site was confirmed by an approach alternative to more conventional methods described by Furuya and Komano (1991) (see Chapter One, section 1.3.2). First, the location of the putative ColIb *nic* region was identified by alignment of the *oriT* sequence data with the consensus sequence 5'-YATCCTG▼Y-3'. Secondly, the legitimacy of the predicted *nic* sequence was confirmed by the inability of recombinant

plasmids, which had been relieved of their *nic* sequence, to be mobilised by *ColIbdrd* (Table 4.2.1). This whole approach was enabled by the extensive database on *oriT* regions.

It has been proposed that the IncI1 conjugation system may have evolved following fusion of two distinct types of conjugative plasmid (Rees *et al.*, 1987). Such a concept arose from the findings that IncI1 plasmids and plasmid F carry genes which have almost identical functions. These genes include *psiB* and *ssb*, which are found within the leading regions of both plasmid F and *ColIb* (Chapter One, sections 1.6.1 and 1.6.2). In addition, the nucleotide sequences of these *ColIb* genes share significant similarity to those of plasmid F, which indicates a common ancestry. In contrast, the *oriT* of IncI1 plasmids *ColIb* and R64 share similar properties to the *oriT* of IncP plasmids RP4 and R751 (Lanka and Wilkins, 1995). The *oriT* of both *ColIb* and R64 can be found situated adjacent to the leading region of each plasmid. If the hybrid hypothesis is correct then there must be a point within the leading region of these two IncI1 plasmids where sequence similarities with plasmid F diverge. Unfortunately, the lack of sequence data available for the leading region of plasmid F has prevented this point from being determined. However, work is currently underway to determine the complete nucleotide sequence of F.

Plasmid *ColIbdrd* only specifies one particular species of *oriT*. Removal of 14 bp from around the *nic* site eliminated the plasmid's ability to self-transfer by conjugation. These results rule out the possibility of another *oriT* structure being present on *ColIb* other than that determined by Rees *et al.* (1987). If a second *oriT* were present on *ColIbdrd*, then transfer of the plasmid by conjugation would not be eliminated by a lesion in the orthodox *oriT* locus, unless the second *oriT* operates under alternative conditions to those induced during intraspecific conjugations between *E. coli*.

Chapter Five

Analysis of the Alleviation of Type I and Type II Restriction Using Non-Transferable Mutants pNA11 and pNA12

5.1 Introduction

Conjugative transfer of *ColIbdrd* is an essential requirement for expression of the plasmid's anti-type II restriction alleviation function. The purpose of this chapter is to determine the stage during conjugation when restriction-alleviation occurs; is it mating-pair formation, DNA transfer or throughout the cycle? To test this, the *ColIbdrd oriT* mutant described in Chapter Four was examined for its ability to confer protection on transferring R751.

Other experiments described include examining whether the *ColIbdrd oriT* mutants immobilised in donor strains could protect transferring R751 from *EcoKI* and *EcoRI* restriction.

5.2 Results

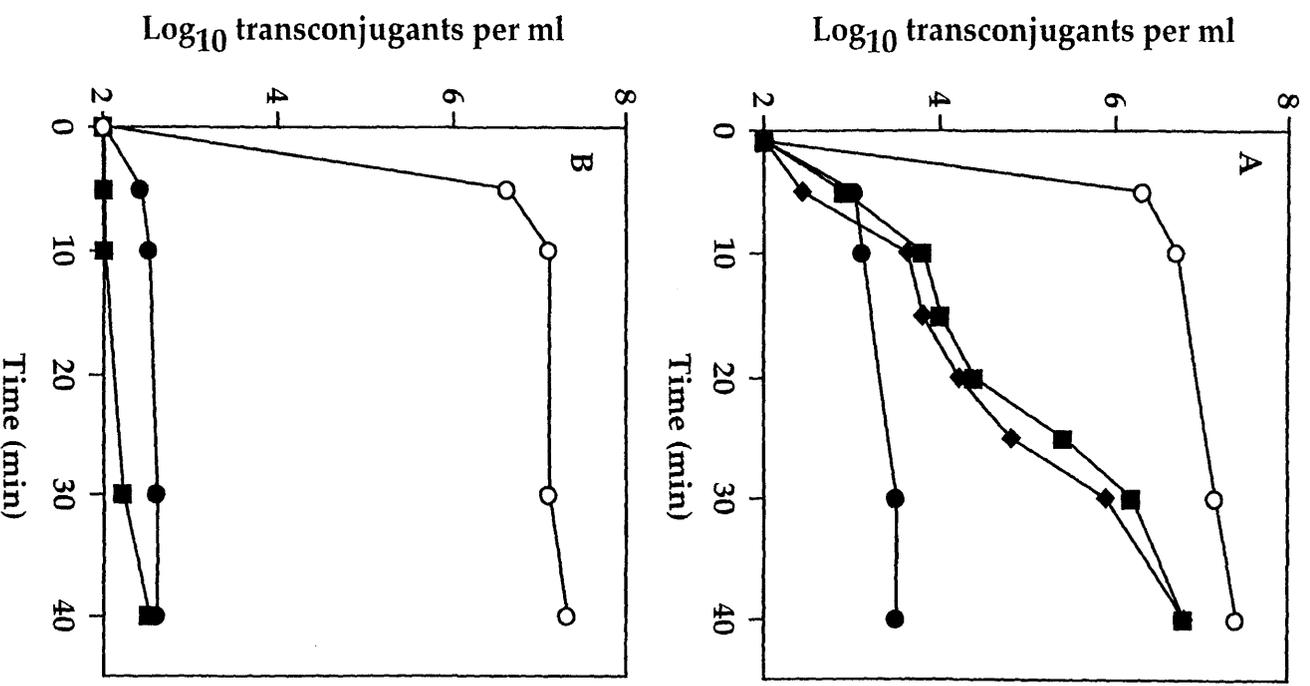
5.2.1 Rescue of R751 from *EcoRI* restriction by *ColIbdrd oriT* mutant pNA11

This section is aimed at determining the stage in the *ColIb* conjugation cycle when *EcoRI* restriction is alleviated. An interrupted mating experiment was performed on filters using a donor strain harbouring R751 and an *EcoRI* restricting recipient harbouring pNA11. The numbers of R751 transconjugants formed in the restricting recipients were measured throughout the experiment and the results are presented in Fig. 5.1 (A). Over a 40-minute period the yield of R751 transconjugants produced in recipients harbouring pNA11 was very similar to that produced in a recipients harbouring pLG221. Such results rule out the possibility that the alleviation function is expressed during the act of DNA transfer.

Fig. 5.1. Transfer of R751 to an *Eco*RI restricting recipient harbouring pNA11.

(A) R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control), GI65N(NTP14) (●, control), GI65N(NTP14, pLG221) (■, control) or GI65N(NTP14, pNA11) (◆). (B) R751 transconjugants from conjugations of GI65R(R751) donors and recipients GI65N (○, control) or GI65N(NTP14) (●, control) and GI65R(R751, pNA11) donors and recipient GI65N(NTP14) (■). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean values of three experiments.

Fig. 5.1



To explore the potential role of mating-pair formation in the alleviation of *EcoRI* restriction, it was examined whether pNA11 could confer the same degree of protection on transferring R751 when present in the donor rather than the recipient. The prediction is that R751 should be protected if restriction is alleviated by the act of forming mating pairs. To this end, a donor strain harbouring R751 and pNA11 was mated with an *EcoRI* restricting recipient. Plasmid pNA11 is transfer defective and will not be able to transfer itself. However, the plasmid encodes all other transfer functions including the ability to form mating pairs. The yields of R751 transconjugants formed during such experiments are presented in Fig. 5.1 (B). These results show quite clearly that no protection is conferred on R751 by pNA11 in the donor, suggesting that restriction-alleviation is not a product of the act of mating-pair formation. In addition, this result shows that alleviation of restriction requires the presence of the *ColIbdrd* plasmid in the restricting strain. Therefore, the potential for the genetic pathways involved in mating-pair formation being implicated in restriction-alleviation cannot be ruled out.

One important discovery, which was made during these experiments, was that the *EcoRI* producing plasmid NTP14 was mobilised to the R751 donor strain by the *ColIbdrd* plasmid in the restricting recipient. This issue is examined further in the following section.

5.2.2 Mobilisation of the *EcoRI* producing plasmid NTP14 by *ColIbdrd*

It was observed during experiments described in section 5.2.1 that plasmid NTP14, which encodes the type II R-M system *EcoRI*, was mobilised from a strain also harbouring a *ColIbdrd* plasmid. To test this further, an interrupted mating experiment was set up on filters between a donor strain harbouring both NTP14 and *ColIbdrd* and a *Rif^R* recipient strain. The numbers of NTP14 transconjugants produced were measured by selecting for *Rif^R* and *Ap^R* cells (Fig. 5.2). It is clear from these data that *ColIbdrd* plasmids pLG221 and pNA11 are able to mobilise plasmid NTP14 efficiently. Although plasmid

Fig. 5.2. Mobilisation of NTP14 by Collb.

NTP14 transconjugants from conjugations of GI65N(NTP14, pLG221) (□) or GI65N(NTP14, pNA11) (○) and GI65N(NTP14, pLG272) (Δ) donors and recipient GI65R. Transconjugants were selected on NA plates containing Rif and Ap. Data are the mean values of two experiments.

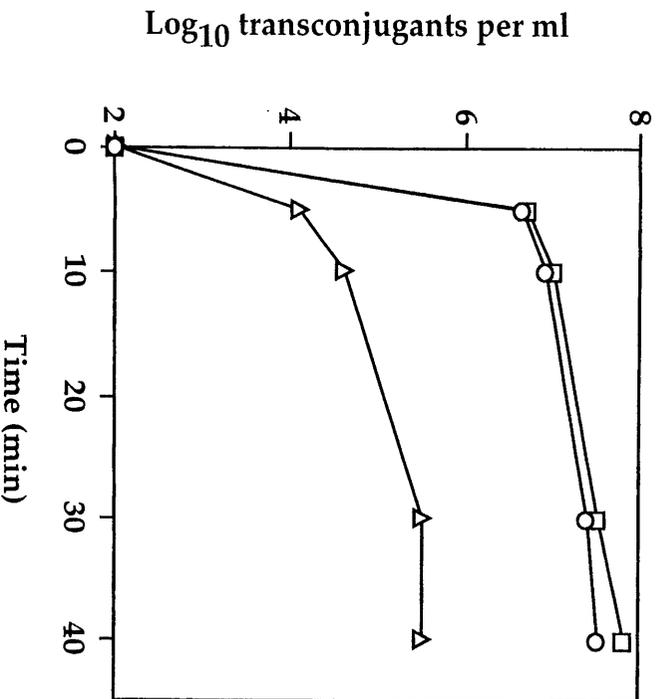


Fig. 5.2

pNA11 cannot transfer itself, it still specifies all functions required *in trans* for transfer. At 40 minutes, pNA11 mobilised NTP14 with an efficiency of 1.0 relative to pLG221 directed mobilisation (Fig. 5.2). In contrast, the number of NTP14 transconjugants was reduced by 141-fold when the mobilising plasmid was ColIb wild type. This reduction can be attributed to the repression of the ColIb conjugation system.

The potential implications of these findings is that during experiments described in section 5.2.1 ColIb*drd* directed mobilisation of NTP14 to R751 donors may result in R751 being modified at its *EcoRI* target sites before entry to the *EcoRI* restricting recipients. Therefore, the R751 plasmid would acquire *EcoRI* protection before entry. This notion is examined further in Chapter Six.

5.2.3 A non-transferable ColIb*drd* confers no *EcoKI* protection on R751

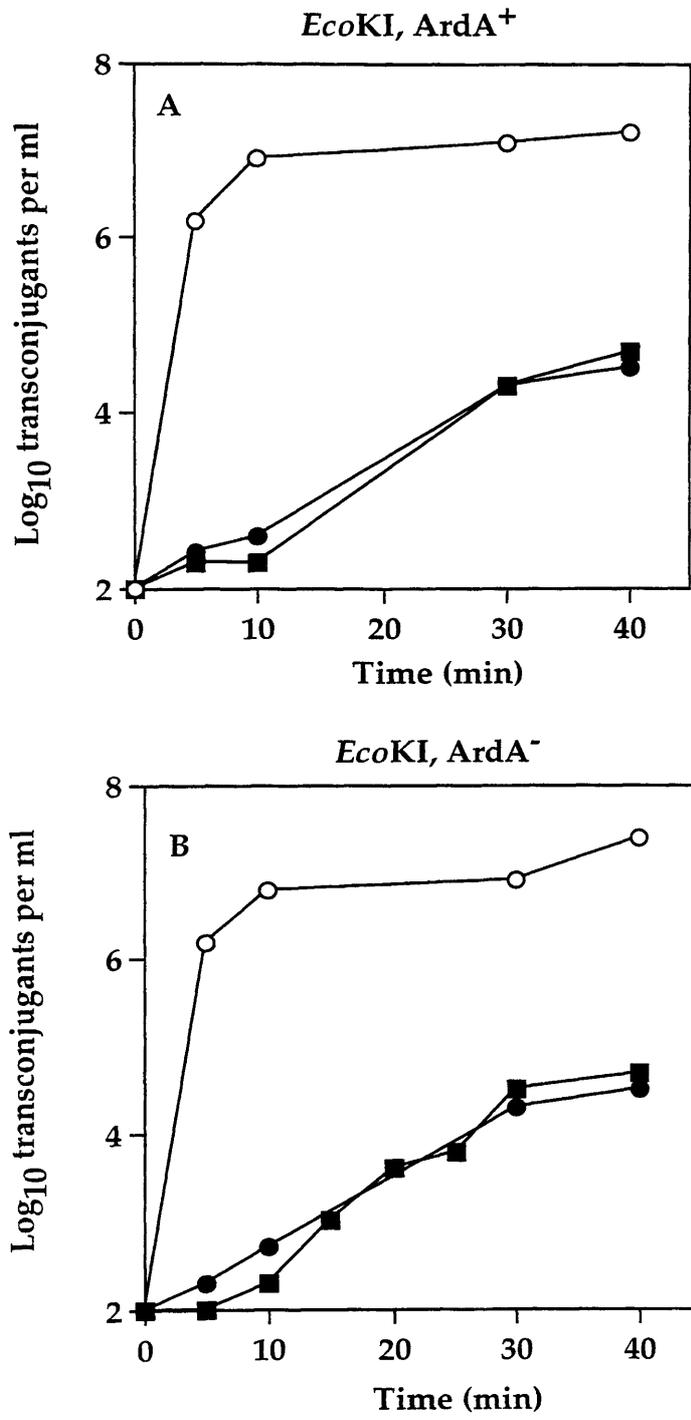
In Chapter Three, the genetic procedure involving plasmid R751 was used to show that the *ArdA* system of ColIb operates optimally when specified by a transferring plasmid (Fig. 3.5 A and Fig. 3.6 C). However, from such results it was impossible to determine in which strain enhanced production of the *ArdA* protein occurred. It might occur in the donor strain if the *ArdA* protein is transferred to the recipient during conjugation, as is the ColIb Sog protein (Rees and Wilkins, 1989). To test this idea, an interrupted mating experiment was performed on filters using a donor strain harbouring both R751 and pNA11(*ArdA*⁺) with an *EcoKI* restricting recipient. Samples of the mating mix were removed and examined for the number of R751 transconjugants (Fig. 5.3 A). Results clearly show that pNA11 is unable to confer any protection on transferring R751 DNA from destruction by *EcoKI*. Such a finding rules out the possibility that the *ArdA* protein is transmissible by conjugation.

A similar mating experiment was carried out using an R751 donor strain harbouring pNA12(*ArdA*⁻) (Fig. 5.3 B). Again, no significant increase in the number of R751 transconjugants was observed. Like *ArdA*, the second

Fig. 5.3. Transfer of R751 from a donor harbouring pNA11(ArdA⁺) or pNA12(ArdA⁻) to an *Eco*KI restricting recipient.

(A) R751 transconjugants from conjugations of NM654R(R751) donors and either recipients NM816N (○, control) or GI65N (●, control) and NM654R(R751, pNA11) donors and recipient GI65N (■). (B) R751 transconjugants from conjugations of NM654R(R751) donors and either recipients NM816N (○, control) or GI65N (●, control) and NM654R(R751, pNA12) donors and recipient GI65N (■). Transconjugants were selected on NA plates containing Nal and Tc. Data are the mean values of three experiments.

Fig. 5.3



*Eco*KI alleviation function encoded by *ColIb* is not transmissible through the donor. Instead, the alleviation process requires the presence of *ColIb* in the restricting strain.

5.2.4 *ColIbdrd* transfer does not mediate retrotransfer of NTP14

Data presented in section 5.2.2 clearly shows that *ColIbdrd* is able to efficiently mobilise plasmid NTP14. It is proposed that *ColIbdrd* might itself evade destruction by *Eco*RI during conjugation by mediating retrotransfer of NTP14 from the recipient cell to the donor cell; such that transmission of NTP14 to the donor cell results in the immediate expression of the gene encoding the *Eco*RI methylase enzyme which in turn modifies all 20 *Eco*RI target sites present on the backbone of *ColIbdrd* before the plasmid enters the restricting recipient by conjugation. The phenomenon of retrotransfer has been described as the transmission of chromosomal or plasmid DNA from the recipient cell to the donor cells supporting conjugation (Sia *et al.*, 1996). If this is how *ColIb* evades type II restriction then it might explain the 6-10 minute delay in *ColIbdrd* transconjugant formation in an *Eco*RI restricting recipient harbouring NTP14 (Fig. 3.2 A). The delay being attributed to the time taken for transmission of NTP14 to the donor and methylation of the 20 *Eco*RI target sites present on *ColIb*. Also, the requirement for multiple rounds of transfer of *ColIbdrd* to the restricting recipient could be explained in that this is required to initiate retrotransfer of NTP14.

To test this idea an interrupted filter mating experiment was performed between a donor strain harbouring *ColIbdrd* plasmid pLG221 and a recipient strain harbouring NTP14. The number of pLG221 transconjugants produced and the number of donor cells that acquired NTP14 were measured throughout the mating (Table 5.2.1). No donor cells that acquired plasmid NTP14 were detected throughout the 40-minute duration of the mating experiment despite successful transmission of pLG221 to the *Eco*RI restricting recipient. These results suggest that transfer of *ColIbdrd* cannot potentiate the

Table 5.2.1. Analysis of ColIb-mediated retrotransfer of mobilisable plasmid NTP14

Donor	Recipient	pLG221 Transconjugants ml ⁻¹	NTP14 Transconjugants ml ⁻¹
GI65R(pLG221)	GI65N	2.6 x 10 ⁷	—
GI65R(pLG221)	GI65N(NTP14)	1.0 x 10 ⁷	< 10
GI65N(pLG221, NTP14)	GI65R	2.0 x 10 ⁷	1.1 x 10 ⁷

Equal numbers of donors and recipients cells were mixed together to give a concentration of 10⁸ cells of each strain per ml. A 40 minute filter mating was performed at 37°C. The numbers of pLG221 (Km^R) and NTP14 (Ap^R) transconjugants were determined by plating the mating mix on NA plates containing the appropriate antibiotic selection. Plasmid NTP14 is mobilised with a frequency of 0.55 relative to the transfer of pLG221, where mobilisation frequency is the number of mobilised plasmids transferred per conjugative plasmid.

retrotransfer of NTP14. This is therefore not how *ColIbdrd* evades destruction by type II restriction enzyme, *EcoRI*.

5.3 Discussion

Three important conclusions can be drawn from data in this chapter. First, the *ColIb* *ArdA* protein is not transmissible from the donor. Evidence is that a *ColIbdrd* (pNA11) plasmid immobilised in a donor strain could confer no protection on transferring R751 from *EcoKI* restriction in the recipient (Fig. 5.3 A). In contrast, *ColIbdrd* (pLG221) itself confers complete protection of R751 (Chapter 3, Fig. 3.5 A). These experiments indicate that enhanced expression of the *ardA* gene occurs in the recipient cell shortly after entry of the *ColIbdrd* plasmid.

Second, like *ArdA*, the components making up the unidentified *ardA*-independent anti-type I system are not transmissible by conjugation (Fig. 5.3 B). In summary, alleviation of *EcoKI* restriction by *ColIb(ArdA⁻)* can only occur within the cell that contains the plasmid. The mechanism by which restriction is alleviated is expressed constitutively by a *drd⁺* and *drd* *ColIb* plasmid (see Chapter Three).

Third, *ColIbdrd*-mediated alleviation of *EcoRI* restriction appeared to occur during the formation of mating-pairs (Fig. 5.1 A). Further experiments indicated that the mechanism has to be expressed in the strain encoding the restriction system (Fig. 5.1 B). These experiments led to the finding that *ColIbdrd* is able to mobilise efficiently the *EcoRI* producing plasmid NTP14 (Fig. 5.2). It may be possible that mobilisation of NTP14 to the R751 donor results in the modification of R751 *EcoRI* target sites before the plasmid is transferred to the restricting recipient. If such a hypothesis were correct, then it would explain a number of observations made during experiments described in Chapter Three. These include the initial lag in R751 transconjugant production in restricting recipients harbouring *ColIbdrd* (Fig. 3.6 A). The lag reflects the time taken for NTP14 to be mobilised to R751 donors and for

complete methylation of the R751 replicon. This issue will be addressed further in Chapter Six.

Another outcome of this finding was that it provided a possible explanation as to how *ColIbdrd* itself might avoid destruction by *EcoRI* during conjugation. The idea being that *ColIbdrd* mediates retrotransfer of NTP14 from the recipient to the *ColIbdrd* donor resulting in modification of the *EcoRI* target sites of the plasmid. However, results in Table 5.2.1 clearly show this not to be the case.

The possibility that *ardA*-independent alleviation of *EcoKI* by *ColIb* is mediated by mobilisation of the *EcoKI* R-M genes can be ruled out by the fact that these genes are located on the *E. coli* chromosome.

One other dilemma that needs to be addressed is that mobilisation of the *EcoRI* R-M genes should result in some killing of recipient cells. All these issues involving the mobilisation of NTP14 are addressed in Chapter Six.

Chapter Six

Construction of a *mob* Mutant of Plasmid NTP14

6.1 Introduction

Data in Chapter Five clearly showed that *ColIbdrd* is able to mobilise efficiently the *EcoRI* producing plasmid NTP14 when resident in the same strain (Fig. 5.2). The aim of this chapter is to establish whether the protection conferred on transferring R751 (Fig. 3.6 A) from *EcoRI* restriction is *ColIbdrd*-mediated or the consequence of NTP14 mobilised to the R751 donor. To address this dilemma it was necessary to construct a mutant of NTP14, which was unable to be mobilised by a conjugative plasmid.

Details concerning the nature and organisation of NTP14 are limited. The only published data available is in Smith *et al.* (1976), who reported that NTP14 was isolated as a recombinant plasmid present in a strain of *E. coli* after an unidentified *fi*⁺ R-factor encoding the *EcoRI* R-M system and a number of antibiotic resistance determinants was mobilised by F-like plasmid R1-19K⁻. NTP14 is 17 kb in size and has a copy number of ~14 per chromosome. The plasmid encodes the type II R-M system *EcoRI*, resistance to ampicillin and colicin E1. From DNA-DNA re-association experiments, the composition of NTP14 was determined to be made up of 33% NTP1, 33% ColE1 and 34% R1-19k⁻ (Smith *et al.*, 1976).

The experimental design for creating and isolating a *mob* mutant of NTP14 was made difficult due to the limited data available for the plasmid. Consequently, several attempts were made to isolate a mutant. The first was based on the fact that NTP14 shares 33% of its genome with ColE1. The *mob* region of ColE1 consists of a *nic* site also known as *bom* (basis of mobility) and four related *mob* genes, *mbeA*, *mbeB*, *mbeC* and *mbeD* (Boyd *et al.*, 1989). Like NTP14, ColE1 also possess a unique *ClaI* cleavage site, which in ColE1 is located in the largest *mob* gene, *mbeA* (Boyd *et al.*, 1989). Based on the fact that each

plasmid contains a unique *Cla*I cleavage site, it was assumed that NTP14 had inherited its *mob* region from ColE1. The *Cla*I site of NTP14 was utilised for the insertion of an antibiotic resistance marker gene with the aim of disrupting the putative *mob* region. However, none of the recombinant plasmids isolated, typified by pNA14, demonstrated a Mob^- phenotype.

The second attempt to isolate a Mob^- mutant of NTP14 involved random transposon mutagenesis of the plasmid using Tn5. Several NTP14::Tn5 insertions were isolated; yet, again none determined a Mob^- phenotype. The failure of these two attempts may be the consequence of the plasmid's copy number such that the recombinant plasmids may be part of a mixed population where the original NTP14 plasmids are able to provide the missing function. This prevents the Mob^- phenotype from being expressed.

The final attempt to isolate a mutant involved cloning restriction fragments of NTP14 into a non-mobilisable cloning vector with the aim of isolating a recombinant plasmid with an R_{RI}^+ , Mob^- phenotype. The choice of cloning vector for these experiments was critical. It was important to choose a vector that could not be mobilised by *ColIbdrd* yet had a similar copy number to that of NTP14. Cloning vector pBR322 has a copy number of ~22 per chromosome but was unsuitable for these experiments as it carries the ColE1 *bom* region and can therefore be mobilised by a number of conjugative plasmids (Finnegan and Sherratt, 1982). Instead, cloning vector pBR328 was chosen, which is a derivative of pBR322 but does not carry the ColE1 *bom* sequence (Soberon *et al.*, 1980; Sambrook *et al.*, 1989).

To this end, experiments described in this chapter include; first, restriction enzyme analyses of plasmid NTP14; Second, the details of the cloning of restriction fragments of NTP14 into pBR328 as well as the tests used to determine which recombinants were R_{RI}^+ , Mob^- ; Third, the transmission kinetics of R751 and *ColIbdrd* to a recipient cell harbouring the R_{RI}^+ , Mob^- recombinant are described to confirm the restriction phenotype, and finally the conjugation experiments are described, which determine whether the

protection conferred on transferring R751 from *EcoRI* restriction is mediated by *ColIbdrd* in the recipient or the consequence of NTP14 being mobilised to the R751 donor.

Additional experiments described in this chapter include examining the viability of unmodified recipient cells that acquired the transmitted *EcoRI* R-M genes by conjugation.

6.2 Results

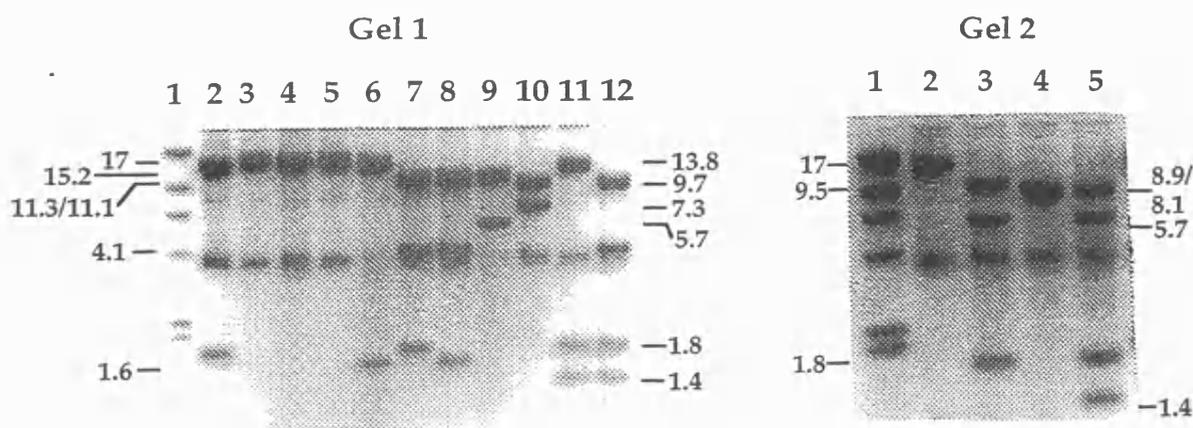
6.2.1 Restriction analysis of plasmid NTP14

As a prerequisite to the cloning of restriction fragments of NTP14 a physical map of the plasmid had to be generated. Construction of the map involved restriction enzyme analysis of NTP14 using enzymes that cleave the plasmid no more than twice. These enzymes included *HindIII*, *ClaI*, *SalI*, *BglII* and *BamHI*. To generate useful data for mapping the cleavage sites on NTP14, single, double and triple digests were performed. The combinations used and the resulting restriction data obtained are presented in Table 6.2.1. The restriction map of NTP14 is presented in Fig. 6.1 A.

The location of the *EcoRI* R-M genes on NTP14 is unknown. However, the location of these genes is known for *E. coli* plasmid R113. The *EcoRI* restriction endonuclease and methylase genes are found within close proximity of each other and span 2.2 kb of R113. The nucleotide sequence of these genes and surrounding regions has been determined and contains an internal 1.8-kb *HindIII* fragment 1.4-kb upstream of a *SalI* site (Greene *et al.*, 1981; EMBL: JO1675). Interestingly, NTP14 also has a 1.8 kb *HindIII* fragment located 1.4-kb from a unique *SalI* site (Fig. 6.1 A). From these similarities it is proposed that the *EcoRI* R-M genes of NTP14 lie in the vicinity of the 1.8 kb *HindIII* fragment. Consequently, the strategy for cloning fragments of NTP14 with the aim of isolating recombinants that are R_{RI}^+ was based around this observation.

Table 6.2.1 Restriction data for NTP14 (17 kb)

Gel No.	Lane No.	Digest	Fragment sizes (kb)
1	2	<i>Hind</i> III	15.2 1.8
	3	<i>Bgl</i> III	17
	4	<i>Cla</i> I	17
	5	<i>Sal</i> I	17
	6	<i>Hind</i> II/ <i>Bgl</i> II	15.2 1.6 0.2
	7	<i>Hind</i> III/ <i>Cla</i> I	11.1 4.1 1.8
	8	<i>Hind</i> III/ <i>Cla</i> I/ <i>Bgl</i> III	11.1 4.1 1.6 0.2
	9	<i>Cla</i> I/ <i>Bgl</i> III	11.3 5.7
	10	<i>Sal</i> I/ <i>Cla</i> I	9.7 7.3
	11	<i>Sal</i> I/ <i>Hind</i> III	13.8 1.8 1.4
	12	<i>Sal</i> I/ <i>Hind</i> III/ <i>Cla</i> I	9.7 4.1 1.8 1.4
	2	2	<i>Bam</i> HI
3		<i>Bam</i> HI/ <i>Hind</i> III	9.5 5.7 1.8
4		<i>Bam</i> HI/ <i>Sal</i> I	8.9 8.1
5		<i>Bam</i> HI/ <i>Hind</i> III/ <i>Sal</i> I	8.1 5.7 1.8 1.4



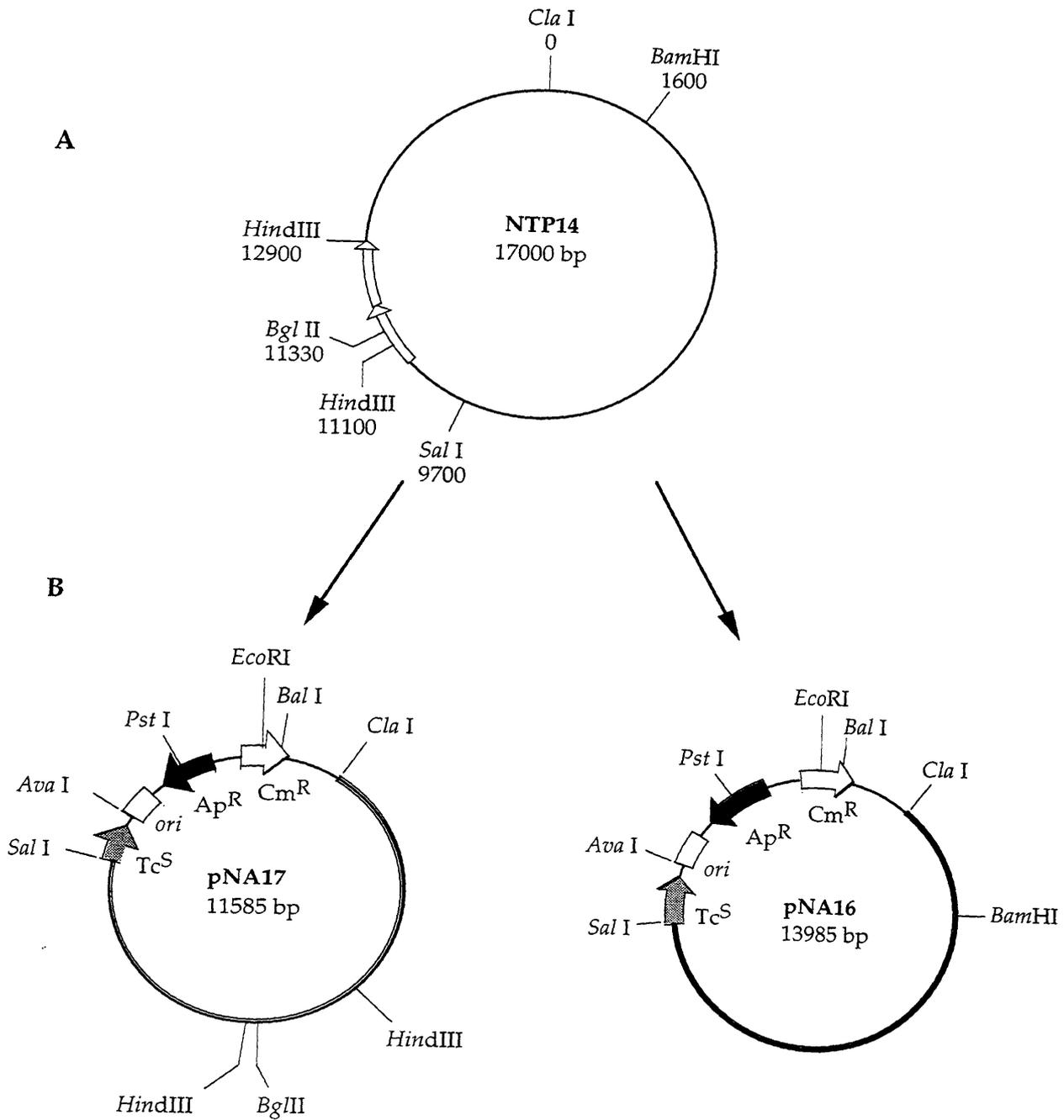
Restriction fragments of NTP14 were separated on 0.8% agarose gels by electrophoresis. λ x *Hind*III molecular weight markers were run in lane 1 of each gel. The size of each restriction fragment of NTP14 was calculated and is shown in kb next to each gel and in the above table. Faint bands running at 3.8 kb in gel 1, lanes 2-5 and 10-11 and gel 2, lanes 2-5 are uncut NTP14.

Fig. 6.1. Diagram depicting the construction of recombinant plasmids pNA16 and pNA17.

(A) Physical map of 17 kb plasmid NTP14 showing the target sites for restriction enzymes *Hind*III, *Bgl*II, *Cla*I, *Bam*HI and *Sal*I. Co-ordinates in bp run from the unique *Cla*I site. Also shown is the suspected location of the *Eco*RI R-M genes (\rightleftarrows) as found on plasmid R113 on a 2.2 kb fragment overlapping the 1.8 kb *Hind*III fragment (Greene *et al.*, 1981; EMBL database acc. no. J01675). Map was compiled using the restriction data in Table 6.2.1.

(B) NTP14 was digested with *Sal*I and *Cla*I yielding two fragments of 7.3 kb and 9.7 kb in size. These two fragments were ligated into the 4.2 kb *Sal*I-*Cla*I fragment of cloning vector pBR328 (thin line), forming pNA16 and pNA17, respectively. pNA17 carries the *Eco*RI system; pNA16 does not. Both recombinants are non-mobilisable.

Fig. 6.1



Not To Scale

6.2.2 Construction of pNA16 and pNA17

It is proposed that the *EcoRI* R-M genes of NTP14 overlap the 1.8 kb *HindIII* fragment of the plasmid. To isolate recombinant plasmids that carry these genes and not the *mob* region of NTP14, a strategy was employed, which involved cloning fragments of NTP14 avoiding disrupting the 1.8 kb *HindIII* fragment and surrounding region. To this end, the 7.3- and 9.7-kb *ClaI-SalI* fragments of NTP14 were removed by restriction enzyme digestion and isolated from a 0.8% agarose gel after separation by electrophoresis. Both the 7.3- and 9.7-kb fragments were purified from the agarose using the QIAEX II gel extraction kit (Qiagen) and ligated into the 4.2 kb *ClaI-SalI* portion of pBR328 forming recombinants pNA17 and pNA16, respectively (Fig. 6.1 B). Isolation of these recombinant plasmids was achieved by transferring the ligated DNA by transformation to DH5 α competent cells. Transformant colonies that were Ap^R, Cm^R and Tc^S were inferred to carry the vector and a cloned insert at the Tc gene. Constructs were confirmed initially by restriction enzyme analysis.

6.2.3 Characterisation of pNA16 and pNA17

Characterisation of recombinants pNA16 and pNA17 involved examining the mobilisation and restriction phenotypes of each plasmid. The R_{RI} phenotypes of pNA16 and pNA17 were examined by infecting their *E. coli* host strain (GI65N) with a λ *vir* lacking *EcoRI* modification and examining the number of plaques formed. The e.o.p of λ *vir*.K was 1.0 in GI65N(pNA16) relative to its plasmid free host and 0.001 in GI65N(NTP14). In contrast, the e.o.p of λ *vir*.K on strains harbouring pNA17 was 0.001 (Appendix I). These results confirm that pNA17 has a R_{RI}⁺ phenotype and pNA16 an R_{RI}⁻ phenotype, as predicted from restriction enzyme cleavage comparisons of NTP14 with R113 (Fig. 6.1 A). It was assumed that pNA17 must also carry the *EcoRI* methylase gene to prevent the host strain from attack by the *EcoRI* endonuclease component.

Table 6.2.2. Mobilisation of recombinant plasmids pNA16 and pNA17 by ColIb*drd*

Donor	Recipient	Transconjugants per ml		*Mobilisation frequency
		pLG221	Recombinant plasmid	
GI65N(pLG221, NTP14)	GI65R	2.7×10^7	2.4×10^7	0.88
GI65N(pLG221, pBR328)	"	4.3×10^7	< 10	$< 2.3 \times 10^{-7}$
GI65N(pLG221, pNA16)	"	3.1×10^7	< 10	$< 3.2 \times 10^{-7}$
GI65N(pLG221, pNA17)	"	1.5×10^7	< 10	$< 6.7 \times 10^{-7}$

The frequency of transfer of each plasmid was determined after a 40 minute filter mating at 37°C, using GI65N as the donor host strain and GI65R as the recipient. Plasmids pNA16, pNA17, pBR328 and NTP14 all confer resistance to ampicillin. Helper plasmid pLG221 confers resistance to kanamycin. Transfer of pNA16, pNA17, pBR328 and NTP14 was measured by plating the mating mixture on NA plates containing Ap ($100 \mu\text{g ml}^{-1}$) and Rif ($25 \mu\text{g ml}^{-1}$) and transfer of the helper plasmid measured by plating on NA plates containing Km ($25 \mu\text{g ml}^{-1}$) and Rif ($25 \mu\text{g ml}^{-1}$). *Mobilisation frequency is the number of mobilisable plasmids transferred per conjugative plasmid.

Recombinants pNA16 and pNA17 were then tested for their ability to be mobilised by *ColIbdrd*. Test strains were made by transferring pNA16 and pNA17 DNA by transformation to GI65N(pLG221) and selecting for Km^R, Ap^R and Cm^R. Conjugation experiments were set up on filters using donors described in Table 6.2.2 and a recipient strain (GI65R). At 40 minutes, the numbers of recipients that had acquired either pNA16 or pNA17 were scored by plating on NA plates containing Rif and Ap (Table 6.2.2). Despite efficient transfer of pLG221, neither pNA16 nor pNA17 was detected in recipient cells. These results confirm that the cloning strategy described in section 6.2.2 was successful in isolating a recombinant plasmid that was Mob⁻. In addition, these results also suggest that the *mob* region of NTP14 must overlap either the *ClaI* site or the *SalI* site where NTP14 was cleaved as a prelude to making constructs.

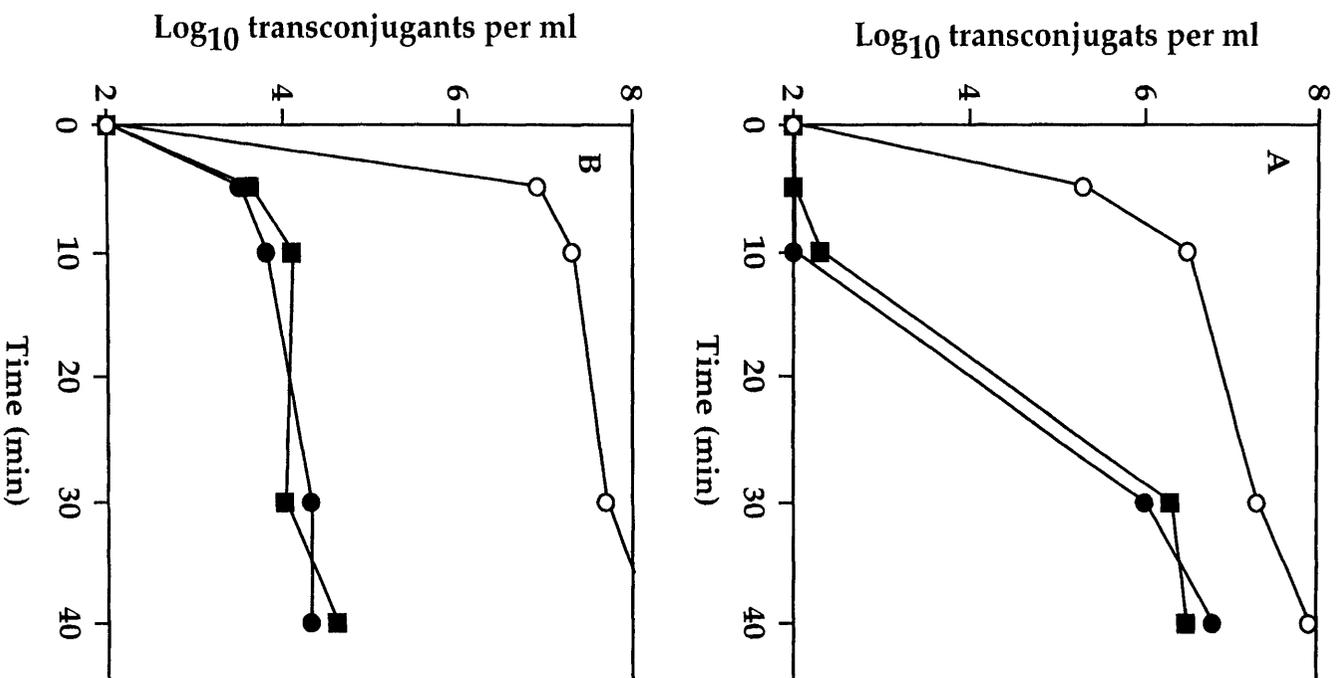
To examine further the restriction properties of pNA17, the transmission kinetics of pLG221 and R751 to a recipient harbouring the recombinant were measured. Two separate conjugation experiments were set up on filters using either GI65R(pLG221) or GI65R(R751) as donors and a Nal^R recipient harbouring pNA17. The yields of pLG221 and R751 transconjugants are shown in Fig. 6.2 A and B, respectively. *EcoRI* restriction still has the same effect on the transmission of pLG221 and R751 if it is encoded by NTP14 or pNA17. These results confirm that pNA17 is a suitable non-mobile source of the *EcoRI* R-M genes.

Finally, the colicin E1 producing phenotypes of pNA16 and pNA17 were examined. Experiments were performed as described in Chapter Two, section 2.3.1. Test strains GI65N, GI65N(pNA16), GI65N(pNA17) and GI65N(NTP14) were examined for colicin production using GI65N as an indicator strain. Only strain GI65N(NTP14) was found to be colicin producing (data not shown), suggesting that the colicin E1 gene, *cea*, of NTP14 overlaps either the *ClaI* or *SalI* cleavage site of the plasmid. Additionally, pNA16 and pNA17 were also examined for their ability to confer immunity against the lethal effects of

Fig. 6.2. Transfer of pLG221 and R751 to a recipient strain harbouring recombinant plasmid pNA17.

(A) pLG221 transconjugants from conjugations of GI65R(pLG221) donors and recipients GI65N (○, control), GI65N(NTP14) (●, control) and GI65N(pNA17) (■). (B) R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control); GI65N(NTP14) (●, control) or GI65N(pNA17) (■). Transconjugants were selected on NA plates containing either Nal/Km or Nal/Tc.

Fig. 6.2



colicin E1 on their host strain. GI65N(NTP14) was used as the test strain and GI65N, GI65N(pNA16), GI65N(pNA17) and GI65N(NTP14) as the indicator strains. Only pNA16 and NTP14 were able to confer any protection on their hosts strains indicating that the colicin E1 immunity gene, *imm*, of NTP14 can be found located somewhere within the 9.7 kb *ClaI-SalI* fragment of the plasmid (Fig. 6.1 A).

6.2.4 ColIb cannot alleviate *EcoRI* restriction *in trans*

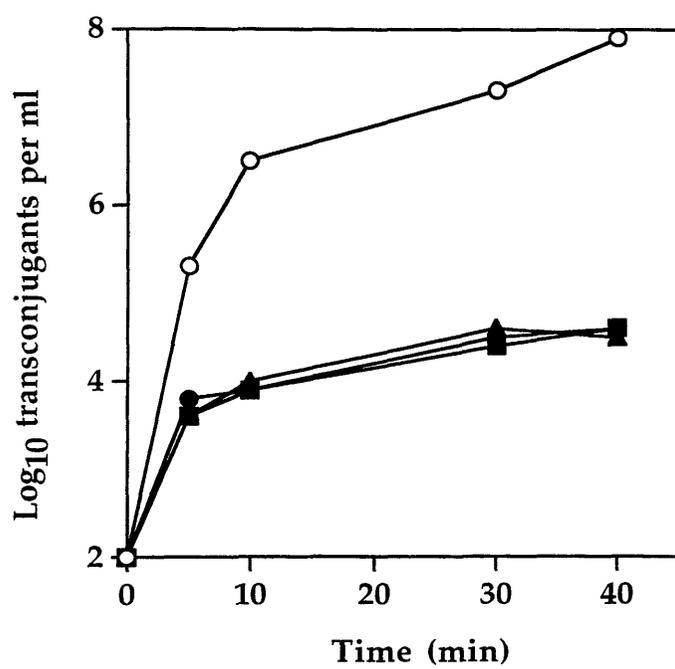
To determine whether the protection conferred on transferring R751 against *EcoRI* restriction (Fig. 3.6 A) is ColIb-mediated or due to mobilisation of NTP14 to the R751 donor, a conjugation experiment was carried out using pNA17. Mating experiments were set up on filters between a donor harbouring R751 and a recipient harbouring pLG221 and pNA17. Samples of the mating cells were removed at intervals and examined for the numbers of R751 transconjugants produced. Figure 6.3 clearly shows that no *EcoRI* protection is conferred on transferring R751 by ColIb*drd* when the R-M genes cannot be mobilised to the R751 donor. These results confirm that the protection originally conferred on R751 from *EcoRI* restriction, as shown in Fig. 3.6 A, is not ColIb-mediated. Such protection must be due to mobilisation of NTP14 to the R751 donor strain and subsequent methylation of R751 before transfer.

Therefore, it seems likely that any evasion of *EcoRI* restriction observed during transfer of ColIb*drd* (Fig. 3.2 A) must be due to a mechanism that can only operate on ColIb*drd* rather than a *trans*-acting mechanism as suggested in Chapter Three (section 3.2.3). Such findings explain why no significant *EcoRI* protection is conferred on R751 by ColIb*drd* when the two plasmids are co-transferred together (Fig. 3.3 A).

Fig. 6.3. Transfer of R751 to a recipient harbouring both *ColIbdrd* and recombinant plasmid pNA17 (R-M RI^+).

R751 transconjugants from conjugations of GI65R(R751) donors and either recipients GI65N (○, control); GI65N(pNA17) (●); GI65N(pLG221, pNA17) (■) or GI65N(pNA11, pNA17) (▲). Transconjugants were selected on NA plates containing Nal and Tc.

Fig. 6.3



6.2.5 Conjugative transfer of the *EcoRI* R-M genes causes some killing of recipient cells

Type II R-M genes are considered to be selfish as they tend to impose selective pressure for the maintenance of the plasmid that encodes them (Naito *et al.*, 1995; O'Neill *et al.*, 1997). However, this is not necessarily true for type I R-M systems. As described in Chapter One, section 1.5.3, the restriction activity of the *EcoKI* enzyme in a new *E. coli* host strain is modulated by the ClpXP protease enzyme leaving its modification activity in tact. The result of *EcoKI* restriction activity being modulated is very little cell death after transmission of the R-M genes to a new unmodified host (Makovets *et al.*, 1998).

It has been shown in experiments described in this thesis that plasmid NTP14, which encodes the type II R-M system *EcoRI*, is mobilisable by *ColIbdrd* (Fig. 5.2). This observation raises the question of how recipients of the plasmid survive the acquisition of the new R-M genes. This section is aimed at examining the consequences for an unmodified *E. coli* strain of receiving the *EcoRI* R-M genes by transfer of the plasmid NTP14.

The experiment involved mating donors harbouring pNA11 and NTP14 and an unmodified recipient strain. The conjugations were set up on filters and the numbers of NTP14 transconjugants produced were determined at intervals during the mating. The yield of NTP14 transconjugants were then compared to the input number of recipient cells in order to evaluate the viability of the recipients after the acquisition of the *EcoRI* R-M genes. As a useful control, a similar experiment was performed using a recipient harbouring pJH16. This is a recombinant plasmid that carries the *EcoRI* methylase gene (Heitman *et al.*, 1989). The results of these experiments are presented in Table 6.2.3. It is clear from the percentage of NTP14 transconjugants produced in an unmodified host compared to those produced in a host carrying pJH16 that conjugative transfer of the *EcoRI* R-M genes on a plasmid must have resulted in some recipient cell death, although the effect is

Table 6.2.3. Analysis of the viability of an unmodified *E. coli* strain during the acquisition of the *EcoRI* R-M genes by conjugation

Donor	Recipient	Time (min)	Recipients per ml	*% transconjugants	NTP14 transconjugants per ml
GI65N(pNA11, NTP14)	GI65R	0	7.5×10^7	0	ND
		5	7.3×10^7	5.3	4.0×10^6
		10	4.5×10^7	20	1.5×10^7
		30	3.4×10^7	20	1.5×10^7
		40	2.4×10^7	16	1.2×10^7
GI65N(pNA11, NTP14)	GI65R(pJH16)	0	6.5×10^6	0	< 10
		5	5.2×10^6	15	1.0×10^6
		10	4.6×10^6	30	2.0×10^6
		30	9.0×10^6	46	3.0×10^6
		40	8.0×10^6	81	5.3×10^6

Equal numbers of donors and recipient cells were mixed together and a filter mating was performed at 37°C for 40 minutes. Samples of the mating cells were removed at intervals and examined for the number of NTP14 transconjugants by plating on NA plates containing Rif (25 µg ml⁻¹) and Ap (100 µg ml⁻¹). Throughout the experiment the number of recipient cells were monitored by plating samples of the mating mix on Rif plates. Helper plasmid pNA11 carries a *cis*-acting lesion within the *oriT* which prevents its self-transfer. However, the plasmid still specifies all other transfer functions *in trans*. Plasmid pJH16 is a recombinant plasmid carrying the *EcoRI* methylase gene. * % transconjugants were calculated by comparing the input number of recipient cells/ml to the number of NTP14 transconjugants/ml produced. Numbers are the mean value of three experiments.

minimal. These results suggest that, like other type II plasmid-encoded R-M systems tested (Naito *et al.*, 1995), the *EcoRI* genes also display selfish behaviour. The fact that not every recipient cell is killed by *EcoRI* could be due to either the activity of repair systems employed by the *E. coli* host strain to mend DNA cleavages or production of the methylase prior to synthesis of the endonuclease.

6.3 Discussion

Results in this chapter detail the successful isolation of a recombinant plasmid with an R_{RI}^+ , Mob^- phenotype. The plasmid, pNA17, was isolated by cloning the 7.3 kb *ClaI-SalI* fragment of NTP14 into vector pBR328. In addition, pNA17 was inferred to specify a similar R_{RI}^+ phenotype to NTP14 as determined during experiments examining the transmission kinetics of both R751 and *ColIbdrd* to a recipient harbouring either NTP14 or pNA17 (Fig. 6.2).

During the construction of recombinant pNA17, several other notable observations were made which further characterise NTP14. These observations are described below.

The cloning of the two *ClaI-SalI* fragments of NTP14 resulted in the generation of two recombinant plasmids, pNA16 and pNA17, neither of which were mobilisable by *ColIbdrd* (Table 6.2.2). These findings indicate that the *mob* region of NTP14 must overlap either the *ClaI* or *SalI* site of the plasmid. If the *mob* region of NTP14 is homologous to that of plasmid ColE1, then comparisons between the restriction data for the two plasmids would suggest that disruption is due to cleavage by *ClaI* (Section 6.1).

Such cloning experiments were also useful in indicating that the colicin E1 gene (*cea*) of NTP14, which must have been inherited from ColE1, can be found overlapping either the *ClaI* or *SalI* cleavage sites of the plasmid (section 6.2.3). However, examination of the nucleotide sequence of ColE1 revealed that there are no *ClaI* or *SalI* cleavage sites present within *cea*. It seems more likely that the *cea* gene of NTP14 is not overlapping one of these two

restriction sites and that some other explanation is necessary. Furthermore, it was shown by Smith *et al.* (1976) that about 1/3 portion of NTP14 (~ 5.6 kb) consists of ColE1 DNA. For both the *mob* region and *cea* gene of NTP14 to be inherited from ColE1 and be disrupted by *Cla*I and *Sal*I is impossible as these two sites are 7.3- and 9.7 kb apart (Fig. 6.2). One possibility is that the ColE1 derived DNA in NTP14 carries an insertion or is interspersed throughout the NTP14 genome. Despite these complications, the colicin E1 immunity gene *imm* was found located within the 9.7 kb *Cla*I-*Sal*I fragment of NTP14 (Fig. 6.1). Further work is required to locate the exact locations of these three loci on NTP14.

Despite the interesting observations that were made during the course of the work described in this chapter, only one important conclusion could be drawn, which is relevant to the main body of this thesis. That is, that *ColIbdrd* does not alleviate *Eco*RI restriction *in trans* during its conjugative transfer (section 6.2.4, Fig. 6.3). It appears that *ColIb* itself evades *Eco*RI restriction by a *cis*-acting anti-restriction function rather than a *trans*-acting one as previously suggested (Chapter Three, section 3.2.3). Such findings explain why no significant protection is conferred on R751 by *ColIbdrd* against *Eco*RI restriction when the two plasmids are co-transferred together (Fig. 3.3 A). In addition, the hypotheses that conjugative transfer of *ColIbdrd* alleviates *Eco*RI restriction either by altering the physiological state of the recipient cell or by substrate saturation of the restriction enzymes can now be ruled out as both would have a *trans*-acting effect. However, the possibility that *ColIb* carries another anti-restriction gene cannot be ruled out, although the product of this gene would have to act *in-cis*. These assumptions will be discussed further in Chapter Eight.

One other observation made during this work was that transfer of the *Eco*RI R-M genes by conjugation decreases, only slightly, the viability of recipient cells. Presumably, any killing is due to the detrimental effects of the endonucleolytic activity of the *Eco*RI restriction enzyme on the unmodified

host genome of the recipient cell. Construction of a restriction mutant of NTP14 for use in a control experiment would have been useful in confirming that killing of recipient cells is due to the *EcoRI* restriction enzyme and no other plasmid-encoded function. The results of transferring this mutant plasmid to unmodified recipient cells could have been used to compare with the results obtained using wild type NTP14 (Table 6.2.3). Unfortunately, due to the time constraints of the project this useful NTP14 mutant was not made. As an alternative control, a strain harbouring recombinant plasmid pJH16 was used as a source of a recipient with a modified genome. However, to what extent the strains genome is methylated as a result of harbouring such a plasmid is not known. Despite this, results in Table 6.2.3 suggest that the slight decrease in viability of recipient cells acquiring plasmid NTP14 is due to the acquisition of genes encoding the *EcoRI* restriction endonuclease. These results will not be discussed further in this thesis.

Chapter Seven

Determination and Analysis of the Nucleotide Sequence Of the ColIb Leading Region

7.1 Introduction

This chapter is concerned with the determination and analysis of the nucleotide sequence of the ColIb leading region. Although peripheral to the main body of this thesis, such an investigation was useful in providing an interesting insight into a relatively uncharacterised region of a conjugative plasmid.

The leading region is defined as the first segment of a conjugative plasmid to enter the recipient cell during bacterial conjugation. To date, only three genes have been identified within this region of ColIb, which are *ssb*, *psiB* and *ardA*. Although not essential for conjugation, the products of these genes have been shown to play an important ancillary role in plasmid DNA metabolism (Loh *et al.*, 1989; Jones *et al.*, 1992; Read *et al.*, 1992; see Chapter One, section 1.6). Each gene is expressed early in a transient burst in the newly infected transconjugant cell promoting establishment of the immigrant plasmid (Howland *et al.*, 1989; Jones *et al.*, 1992; Althorpe *et al.*, 1999). How the expression of these genes is enhanced during conjugation is unknown. However, a number of hypotheses exist (Chapter One, section 1.6.5).

The aim of this chapter is to use the sequence data for a number of purposes including determining the genetic organisation of the region, identifying more potential protein coding regions and, through sequence comparisons, determining their possible function or evolutionary ancestry. It was also hoped that indications of the possible mechanism for enhanced expression of these genes during conjugation might be revealed.

The work described in this chapter is the result of a joint effort with Dr S. Bates, a post-doctoral researcher in this laboratory. The recombinant

plasmids constructed by myself and S. Bates for sequencing the ColIb leading region are described in Fig. 7.1. Of the 11.7 kb of sequence data generated, 6.4 kb was obtained during this work and the other 5.3 kb by S. Bates. Analysis of the sequence data was a joint effort. With the permission of Dr S. Bates, all data obtained during the collaboration is presented within this chapter. Experiments described include the strategy used to clone and sequence 11.7 kb of the ColIb leading region and the computer-based analyses employed to identify interesting features within the resulting sequence data.

No attempts were made to identify promoters within the leading region by examining the sequence data for the classical -10 and -35 RNA polymerase binding and recognition consensus sequences. Studies of numerous collections of promoters identified in *E. coli* has led to a divergence from the original consensus sequences proposed by Pribnow (-10) and Maniatis (-35) where a base substitution at each position is possible. Such degeneracy has led to the conclusion that a promoter can be found every 200 bp of sequence. Use of the consensus sequences to identify promoters has to be treated with caution and any candidates must be considered arbitrary until functional data are obtained (Hénaut and Danchin, 1996).

7.2 Results and Discussion

7.2.1 Strategy for determining the nucleotide sequence of the ColIb leading region

The leading region of ColIb is found located adjacent to the *oriT* within the largest *EcoRI* (E1) fragment of the plasmid (Fig. 7.1 A). The length of the leading region is estimated to be ~ 17 kb from *oriT* to the genes encoding UV protection functions (*imp*). To generate a library of subclones suitable for DNA sequencing, smaller restriction fragments of the region were isolated from either pCRS4 or pCRS3 (Rees, 1986; Rees *et al.*, 1987) and ligated into the appropriate cleavage sites of cloning vector pIC19H (Fig. 7.1 C). Recombinants

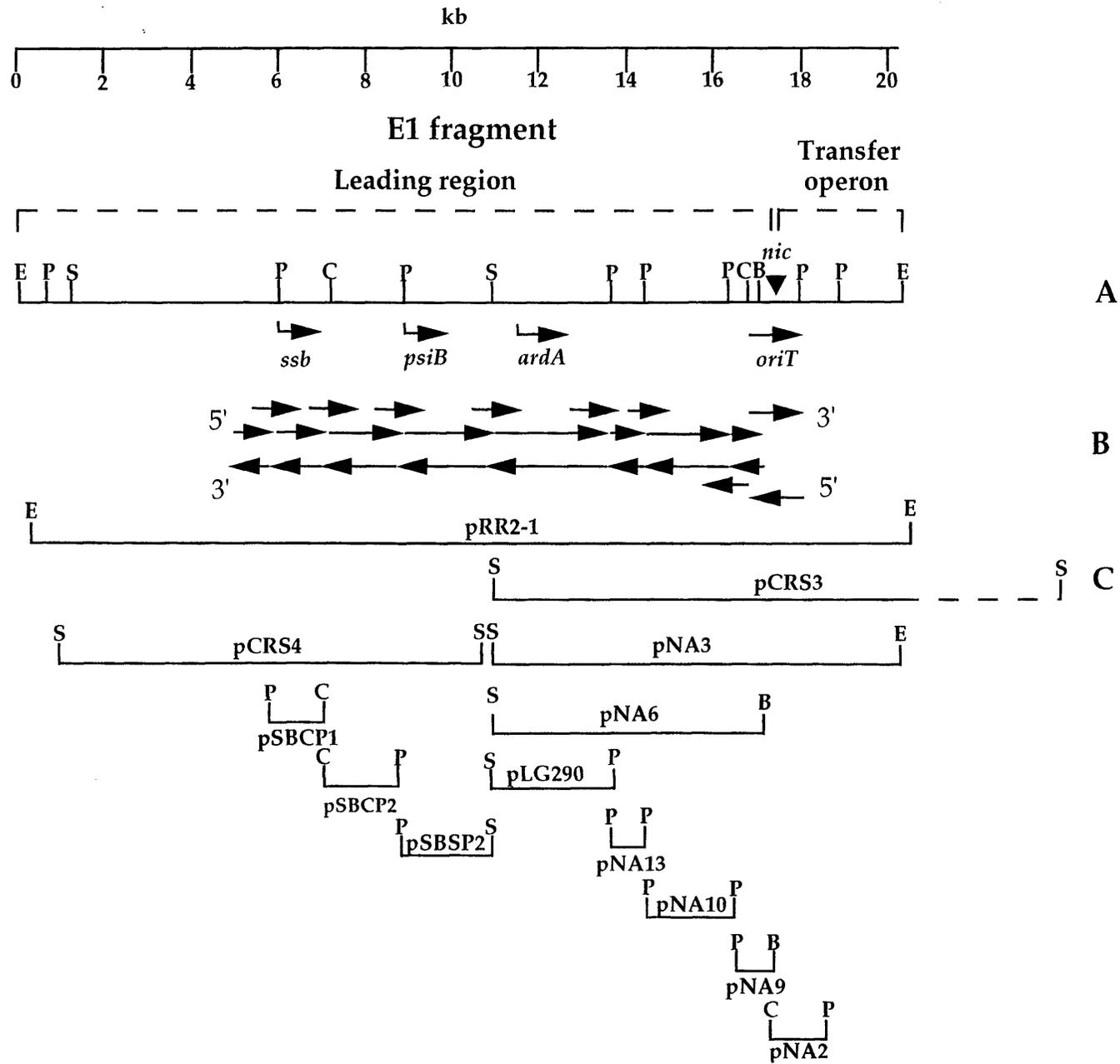
Fig. 7.1. Scheme showing the construction of a small library of subclones used to generate the 11.7 kb nucleotide sequence of the ColIb leading region.

(A) Physical map of the large *EcoRI* (E1) fragment of ColIb, which encompasses the leading region and part of the transfer region including *oriT* and the *oriT* operon. Restriction enzyme cleavage sites are indicated by B (*BglII*), C (*ClaI*), E (*EcoRI*), P (*PstI*) and S (*SalI*). Also indicated are the positions of *ssb*, *psiB*, *ardA* and the *nic* site within *oriT*.

(B) Schematic diagram showing the strategy devised to determine the nucleotide sequence of both DNA strands of the ColIb leading region. One arrow (5' → 3') may represent more than one sequencing reaction to generate data for one strand of the corresponding subclone. Primers used in these sequencing reactions are in Chapter Two, Table 2.1.4.

(C) Library of subclones constructed to generate sequence data for the ColIb leading region. Restriction enzyme cleavage sites are indicated. Subclones generated from pCRS4 (Rees, 1986) and the resulting sequence data were by S. Bates. Subclones generated from recombinant pNA3 and resulting sequence data was part of this work with the exception of pLG290 (Read *et al.*, 1992). The source of the insert for pNA3 was pCRS3 (Rees *et al.*, 1987). Subclones pRR2-1 (Roscoe, 1996) pNA6 and pCRS4 were used to generate sequence data over restriction enzyme cleavage sites. The *ArdA* phenotypes of pNA6 and pLG290 were examined and the results are shown in Appendix I.

Fig. 7.1



pLG290 (Read *et al.*, 1992) and pNA2 (Althorpe *et al.*, 1999; Chapter Four, section 4.2.1) were isolated during other studies.

Attempts were made by S. Bates to subclone the region upstream of *ssb*. However, all recombinants isolated resulted in rearrangements of the *SalI-PstI* insert. Such a phenomenon has also been described by Howland *et al.* (1989) during the cloning of S4, which is a *SalI* fragment internal of the E1 fragment of ColIb (Fig. 7.1 A). Unfortunately, due to time constraints this problem was not resolved and consequently no sequence data are available beyond 400 bp upstream of *ssb*.

Sequence reactions were performed as described in Chapter Two using either M13/pUC universal forward or reverse primers or a specially designed primer as described in Table 2.1.4. Each subclone was sequenced completely on both strands according to the strategy shown in Fig. 7.1 B. Generation of sequence data over restriction enzyme cleavage sites was achieved using specifically designed primers and one of the larger subclones, pCRS4, pNA6 or pRR2-1.

The sequence data generated from these reactions, combined with the data generated for the *oriT* region (Chapter Four, Fig. 4.2) were compiled using the GEL set of programs from the GCG package to give a contig of 11,712 bp in length. The raw sequence data was then annotated and is shown in Fig. 7.3. In addition, this data can also be found in the EMBL database accession number AJ238399. Much of this work is also published in Bates *et al.* (1999).

7.2.2 Identification of putative protein coding regions (*orfs*) in the ColIb leading region

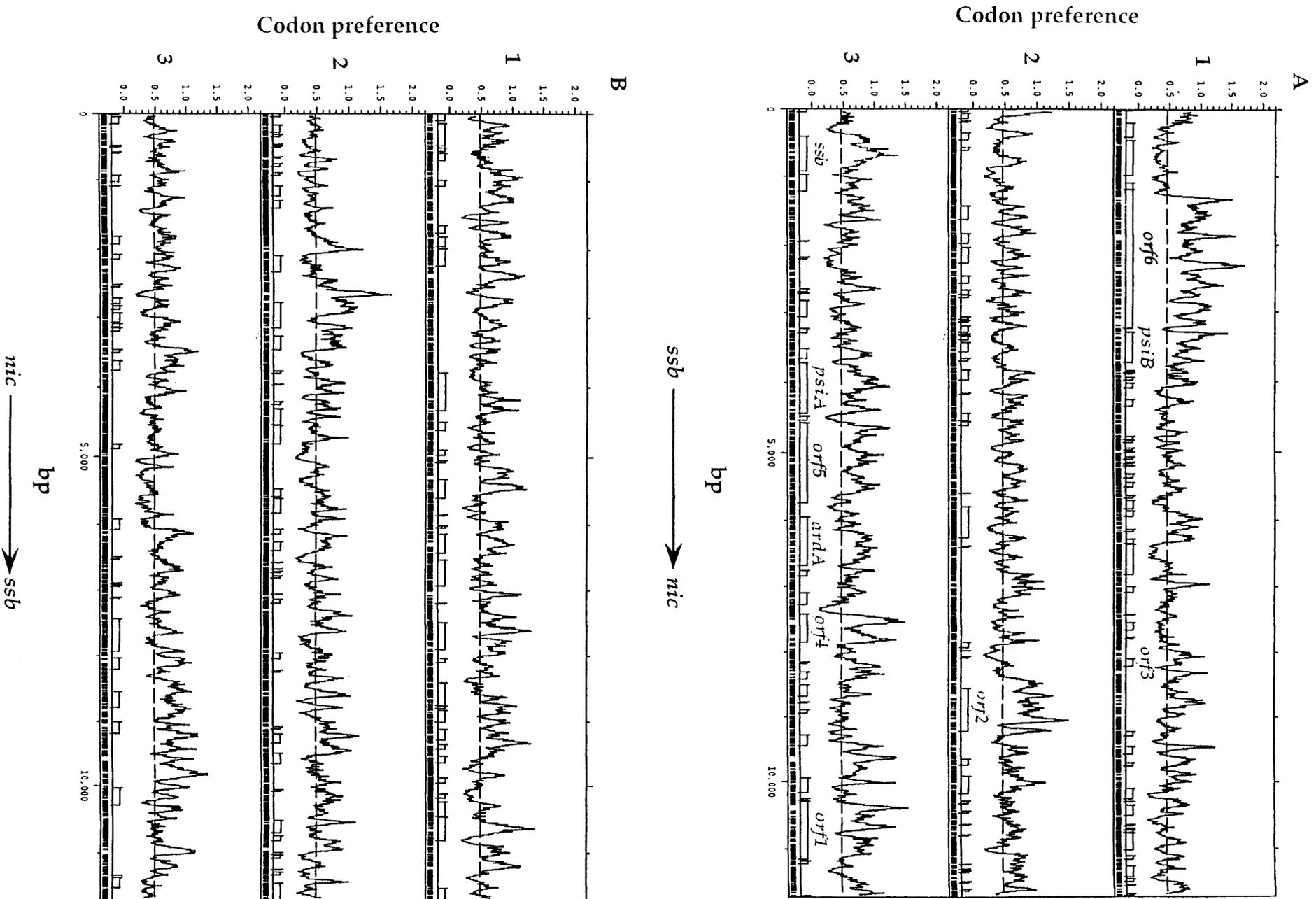
To identify putative protein coding regions within the leading region of ColIb the sequence data generated was analysed using the GCG program CODONPREFERENCE and the *E. coli* codon usage table of highly expressed genes (Gribskow *et al.*, 1984; Hénaut and Danchin, 1996). Furthermore, the translation table was altered so that as well as ATG, GTG start codons were

Fig. 7.2. Protein coding regions encoded by ColIb leading region.

Putative protein coding regions (*orfs*) and previously identified genes (*ssb*, *psiB* and *ardA*) were identified by analysis of the ColIb leading region sequence data using the GCG program CODONPREFERENCE (window 25 bp) and the *E. coli* codon usage table of highly expressed genes. Only *orfs* with ATG or GTG start codons were searched for.

Panel **A** shows the analysis running from *ssb* towards the *nic* site of the *oriT*. Panel **B** shows the analysis running in the opposite orientation from the *nic* site to *ssb*. Note that the potential *orfs* identified, and *ssb*, *psiB* and *ardA* are all transcribed towards *oriT*.

Fig. 7.2



incorporated into the analyses. Both orientations of the 11.7 kb sequence were analysed and the CODONPREFERENCE statistical plots and *orfs* identified are shown in Fig. 7.2. In total, ten *orfs* (Fig. 7.2 and Table 7.2.1) were identified, three of which include previously identified genes *ssb*, *psiB* and *ardA*. The most common start codon was ATG found for eight of the ten *orfs*, where the GTG start codon was found for *orf2* and *orf3*. TGA was the only stop codon found.

Adding to the authenticity of these potential *orfs*, good matches to the Shine-Dalgarno consensus sequence (3'-AUUCCUCCACUGCAU-5') (Mathews and van Holde, 1990) for ribosome binding were located for each *orf* between 4-8 bp upstream of each start codon. These ribosome-binding sites (*rbs*) are indicated in Fig. 7.3.

The most striking feature of these newly identified *orfs* is that like *ssb*, *psiB* and *ardA*, they are all orientated in the same direction towards *oriT*. Such a finding raises the possibility that these genes might be transcribed as part of an operon rather than as single separate units. In addition, the biological significance of this organisation of the leading region may be that the DNA strand destined for transfer (T-strand) forms the template strand for mRNA synthesis. This may be important to ensure that each gene is expressed soon after infection of the recipient cell by the immigrant plasmid.

The coding regions of all the *orfs* identified were translated using the computer program 'Gene Jockey II' (BIOSOFT). Both the nucleotide and amino acid sequences of these coding regions were compared to other sequences present in the databases EMBL, GenBank and SwissProt, available via the World Wide Web (Chapter Two).

7.2.3 Similarities between DNA and protein sequences found in the leading region to those present in the databases

Attempts were made to determine the potential function and evolutionary history of the *orfs* identified within the leading region of ColIb.

Analyses involved comparing specific regions of either the nucleotide or predicted amino acid sequences of the leading region to those available within the databases EMBL, GenBank and SwissProt. The GCG programs used for these comparisons were FASTA, BLAST and TFASTA (Pearson and Lipman, 1988; Altschul *et al.*, 1990; Hénaut and Danchin, 1996). However, once potential homologues had been identified, the GCG program BESTFIT was often used to determine the amount of similarity and identity between the two sequences. Potential secondary structures within protein sequences were identified using the GCG program MOTIF.

In the first place, sequence comparisons involved the previously identified genes *ssb*, *psiB* and *ardA* to establish if any new homologues could be found within the databases to those found when the genes were first sequenced (Howland *et al.*, 1989; Jones *et al.*, 1992; Delver *et al.*, 1991).

No new homologues were found for both *ssb* and *psiB*, despite the fact that a number of plasmids are known to carry *ssb*- and *psiB*-like sequences (Golub *et al.*, 1988). However, a number of errors were found in the original sequence of *psiB*, which on correction resulted in both the nucleotide and predicted amino acid sequence forming a better alignment with the two other sequenced *psiB* genes of plasmid F and R6-5 (Dutreix *et al.*, 1988). F and R6-5 PsiB proteins are the same size as ColIb PsiB (144 residues) and display 85.4% and 84.6% identity to the ColIb protein respectively. Such alignments suggest that a stronger relationship exists between the genes of ColIb, F and R6-5 than previously suggested by Jones *et al.* (1992).

Through cross-hybridisation studies homologues of the ColIb *ardA* gene have been found on a number of enterobacterial plasmids including members of the I complex (I1-B-K), IncN and IncFV plasmids (Chilley and Wilkins, 1995). Sequence similarities have also been demonstrated between ColIb *ardA* and those of plasmids R16 (IncB), pKM101 (IncN) and F0*lac* (IncFV), displaying 94.0%, 60.5% and 56.4% identity, respectively. During our analyses we found two new homologues of ColIb *ardA*, which were Orf79 of plasmid pMT1 of

Yersinia pestis (causative agent of bubonic plague) and Orf18 of the conjugative transposon Tn916. The predicted amino acid sequences of these ArdA-homologues share 55% and 25% identity with that of Collb ArdA, respectively (Hu *et al.*, 1998; EMBL accession number AF053947; Clewell *et al.*, 1995).

Analysis of the *orf* found within the sequence at base positions 8002-7283 (Fig. 7.3) revealed a similarity to the *psiA* genes of plasmid F and R6-5 (Loh *et al.*, 1989). Like *psiA* of F and R6-5, Collb *psiA* also overlaps the TGA stop codon of the *psiB* gene, which suggests that these two *orfs* are co-ordinately expressed from the same transcript. The Collb PsiA protein, which is predicted to be 239 amino acid residues in length, shares 63.4% identity to the corresponding proteins of F and R6-5 with the majority of amino acid changes located in the C-terminal region of the peptide. The function of PsiA of plasmid F and R6-5 is unknown although unlike PsiB it is not believed to be involved in SOS inhibition as PsiB confers the major Psi phenotype (Bailone *et al.*, 1988).

The largest *orf* identified within the sequence was *orf6*, which can be found located between *ssb* and *psiB* at base positions 10527-8488 (Fig. 7.3). The predicted amino acid sequence of Orf6 is 679 amino acid residues with a molecular weight of 74 kDa (Table 7.2.1). Comparisons of the sequence of Orf6 with those found in the databases identified only one full-length homologue, Orf81 of *Yersinia pestis* plasmid pMT1 (53% identity). The function of Orf81 is unknown (Hu *et al.*, 1998). Other homologues identified included members of the SpoOJ family of proteins, whose functions include DNA partitioning, and the KorB transcriptional repressor of IncP plasmid RP4. These proteins share 20-30% identity with Orf6 over 200 amino acids residues. From these comparisons, it is predicted that Orf6 might be a DNA binding protein, although the protein exhibits an overall net negative charge (Table 7.2.1). However, further evidence to support this notion is that the amino acid sequence of Orf6 contains a potential Helix-Turn-Helix (HTH; residues 186-205) of the AraC family type. Furthermore, a potential P-loop structure is identified

Table 7.2.1. Orfs identified in the ColIb leading region.

ORF	Co-ordinates from <i>nic</i>	Amino acid Residues	Mr (kDa) †	Net charge	Motifs §	Notes
<i>ssb</i>	11316-10789	175	19.2	3	ssDNA binding	ssDNA-binding protein
<i>orf6</i>	10529-8490	679	74.3	-18	P-loop HTH motif coiled coil	DNA binding?
<i>psiB</i>	8435-8001	144	15.9	-8	-	Inhibitor of SOS induction
<i>psiA</i>	8004-7285	239	27.7	3	-	Unknown
<i>orf5</i>	7155-5980	391	43.9	40	-	Possible transposase
<i>ardA</i>	5531-5030	166	19.2	-25	-	Antirestriction protein
<i>orf4</i>	4301-3867	144	16.5	2	-	Unknown
<i>orf3</i>	3775-3590	88	10.5	6	-	Unknown
<i>orf2</i>	3444-2503	313	35.8	-9	-	<i>yadD</i> homologue
<i>orf1</i>	1409-561	282	32.2	-3	-	Unknown/methylase?

† Mr was predicted using the PEPTIDESORT program from the GCG package version 9.2 (Genetics computer group, 1997)

§ Motifs were detected using the GCG program MOTIFS searching against the PROSITE database. Predicted helix turn helix (HTH) using the GCG programs HTHSCAN and COILSCAN respectively. Table is taken from Bates *et al.*, (1999).

at residues 178-185, which is indicative of an ATP/GTP binding site motif. Both these regions of Orf6 have a net positive charge (Table 7.2.1).

There is suggestive evidence that expression of *orf6* is enhanced during conjugative transfer of ColIb, since a protein of similar size (82-kDa) to the predicted Orf6 was found to increase in relative concentration during bacterial conjugation (Jones *et al.*, 1992).

The *orf5* locus can be found at base position 7153-5978 in the sequence (Fig. 7.3). The predicted amino acid sequence of Orf5 is 391 residues with a high net positive charge (Table 7.2.1). Comparisons of the sequence of Orf5 with other proteins present in the databases revealed extensive homologies to the protein-coding regions found within a number of unusual insertion elements (IS) all related to IS891. In fact, similarities were detected over the full length of Orf5 to the probable transposases of IS elements IS1253 of *Dichelobacter nodosus* (41% identity, 52% similarity; Billington *et al.*, 1996; EMBL: U34772); IS1341 of thermophilic bacterium PS3 (39% identity, 51% similarity, Murai *et al.*, 1995); IS605 (*tnpB*) of *Helicobacter pylori* cag pathogenicity island (34% identity, 45% similarity, Censini *et al.*, 1996; EMBL: U95957) and IS891 of Cyanobacterium *Anabaena* sp. (39% identity, 44% similarity; Bancroft and Wolk, 1989; EMBL: M24855). Hence, the product of *orf5* appears to be a transposase (Table 7.2.1). Our analysis also revealed that like the IS elements mentioned above, the N-terminal 120 amino acids of Orf5 have a high degree of homology to VirE (81% identity) and VsdF (79% identity), which are encoded by virulence-associated plasmids from *Salmonella typhimurium* and *Salmonella dublin*, respectively (Gulig *et al.*, 1992; Krause *et al.*, 1991).

No inverted repeat sequences that are classically associated with the ends of IS elements could be found surrounding *orf5*. However, as found for IS1253, an imperfect direct repeat sequence with internal dyad symmetry, possibly directing transposase recognition, was found flanking *orf5* (Fig. 7.3). Interestingly, the location of *orf5* corresponded to a region of the sequence that

Fig. 7.3. 11.7 kb DNA sequence of the ColIb leading region.

Features shown include the ATG / GTG start and TGA stop codons of *ssb*, *psiB*, *psiA*, *ardA*, *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, and *orf6*. The associated ribosome binding site (*rbs*) of each *orf* is underlined. Amino acid sequences deduced from the DNA sequences of all *orfs* are presented where the start of each translated amino acid is in uppercase. All *orfs* are transcribed in the rightwards direction (5' → 3'). Restriction enzyme cleavage sites are *Bgl*II, *Cla*I, *Pst*I and *Sal*I.

The *ssi* sites (indicated by single lines) are putative promoters for rightwards transcription of the transferring DNA strand. These *ssi* sites were identified by their similarity in sequence to *Frpo*, which functions as a single-stranded promoter in plasmid F (Masai and Arai, 1997; EMBL: D90178).

The *nic* site is indicated (▲) and the direction of transfer of the plasmid to the recipient during conjugation from this site is from right to left (5'-3').

Imperfect direct repeats flanking *orf5* in the T-strand are indicated (/ \) and are:

```
5945 5' -TGCTCCCCGCCCTG - TCGGGCGAGGCTTCCC-3' 5974
      . . . . .
7233 5' - TCCTCCCCACCCGCATTGGGCGAGGGTTCTC-3' 7263
```

Fig. 7.3.

TC
--
ssi3

11712 AGACAGCGCGGCACTGGACGCGAGGACGTGTGGACTGCCACACACTCCTTTGACGCGGA 11655

11654 CTACCGCGTCGTCAGAAGGGAGGACGGCGGTGGAGACGCCGACAAAGGGCGACACCTGA 11595

11594 AAGGCCCCCGTCTGCATTGGGGGAGGGCAAGGTAGACAGGTCCCCTCTACGTGGCCGG 11535

11534 TGTTTTAAAAATCGGTACTTGGGCCGTTT'TTAAACGCAGGCGGGGGCCTGTCGACAGG 11475

11474 GCACTCCGCCATTTGCCCGGGGTAGCCTTTCCAGGTGACCCCCCTCTTGGGGACTT 11415

AAGGGGCACATGAACGTCCCCTCTCCGGT
-----+-----+-----+ 11355

11414 CACCTGCCGTCCTCCTCGAAAGCGAGGAGACTTCCCCGTGTAAGTGCAGGGCAGAGGCCA

rbs *ssb* →

ATCCGGTTTATCTCACATCTTTCAGTCAGGAGGTTCTTATGAGTGCACGTGGTATCAACA
11354 -----+-----+-----+-----+-----+-----+-----+ 11295

TAGGCCAAATAGAGTGTAGAAAGTCAGTCCTCCAAGAATACTCACGTGCACCATAGTTGT

MetSerAlaArgGlyIleAsnLys

AGGTCATCCTCGTCGGGCGTCTGGGCAATGATCCGGAGGTCCGTTACATCCCCAACGGTG
11294 -----+-----+-----+-----+-----+-----+-----+ 11235

TCCAGTAGGAGCAGCCCGCAGACCCGTTACTAGGCCTCCAGGCAATGTAGGGGTTGCCAC

ValIleLeuValGlyArgLeuGlyAsnAspProGluValArgTyrIleProAsnGlyGly
PstI

GCGCAGTGGCAAAC[ETGCAGGTGGCCACGTCAGAAAGCTGGCGGACAAACAGACGGGGG
11234 -----+-----+-----+-----+-----+-----+-----+ 11175

CGCGTCACCGTTTGGACGTCCACCGGTGCAGTCTTTCGACCGCGTGTGTTGTCTGCCCCC

AlaValAlaAsnLeuGlnValAlaThrSerGluSerTrpArgAspLysGlnThrGlyGlu

AGATGCGGGAGCAGACGGAATGGCACCGTGTGTTGTCTGTTTCGGCAAGCTCGCGGAAGTGG
11174 -----+-----+-----+-----+-----+-----+-----+ 11115

TCTACGCCCTCGTCTGCCTTACCGTGGCACAACACGACAAGCCGTTTCGAGCGCCTTACC

MetArgGluGluThrGluTrpHisArgValValLeuPheGlyLysLeuAlaGluValAla

CAGGTGAATATCTGCGCAAGGGCGCGCAGGTCTACATCGAGGGGCAACTCCGTACCCGTA
11114 -----+-----+-----+-----+-----+-----+-----+ 11055

GTCCACTTATAGACGCGTTCCCGCGCTCCAGATGTAGCTCCCCGTTGAGGCATGGGCAT

GlyGluTyrLeuArgLysGlyAlaGlnValTyrIleGluGlyGlnLeuArgThrArgSer

GCTGGGACGACAACGGCATCACCCGCTACATCACTGAAATCTTGTTAAGACCACGGGCA
11054 -----+-----+-----+-----+-----+-----+-----+ 10995

CGACCCTGCTGTTGCCGTAGTGGGCGATGTAGTACTTTAAGAACAATTCTGGTGCCCGT

TrpAspAspAsnGlyIleThrArgTyrIleThrGluIleLeuValLysThrThrGlyThr

CCATGCAGATGCTGGGGAGTGACCCACAGCAGAACGCTCAGGCGCAACCGAAGCCTCAGC
10994 -----+-----+-----+-----+-----+-----+-----+ 10935

GGTACGTCTACGACCCCTCACGTGGTGTCTGCGAGTCCGCGTTGGCTTCGGAGTGC

MetGlnMetLeuGlySerAlaProGlnGlnAsnAlaGlnAlaGlnProLysProGlnGln

10934 AGAATGGGCAGCCACAGAGTGTCTGACGCGACGAAAAAAGGTGGCGCGAAAACGAAAGGCC 10875
 -----+-----+-----+-----+-----+-----+-----+-----+
 TCTTACCCGTCGGTGTCTCACGACTGCGCTGCTTTTTTCCACCCGCTTTTGCTTCCGG

AsnGlyGlnProGlnSerAlaAspAlaThrLysLysGlyGlyAlaLysThrLysGlyArg

10874 GTGGACGTAAGGCCGCGCAGCCCGAGCCTCAGCCGCAAACGCCGAGGGGGAGGATTACG 10815
 -----+-----+-----+-----+-----+-----+-----+-----+
 CACCTGCATTCCGGCGCTCGGGCTCGGAGTCGGCGTTTGGCGCCTCCCCCTCCTAATGC

GlyArgLysAlaAlaGlnProGluProGlnProGlnThrProGluGlyGluAspTyrGly

10814 GGTTTTCAGACGACATCCCGTTCTGATCGGGCTGACTGTGACAACGCCCCGCCCTGTGC 10755
 -----+-----+-----+-----+-----+-----+-----+-----+
 CCAAAAGTCTGCTGTAGGGCAAGACTAGCCCGACTGACACTGTTGGCGGGCGGGACACG

PheSerAspAspIleProPhe*

10754 GGGGCATCACCCGAGATATGAGGATGAGCGAATATTTTCAGAATACTTCAGGGACTGCCGG 10695
 -----+-----+-----+-----+-----+-----+-----+-----+
 CCCCCTAGTGGCCTCTATACTCCTACTCGCTTATAAAGTCTTATGAAGTCCCTGACGGCC

10694 ACGGCTCCTTTACCCGCGAACAGGCGGAAGCCGTTGCCGCACAGTACCGGAACGTCTTTA 10635
 -----+-----+-----+-----+-----+-----+-----+-----+
 TGCCGAGGAAATGGGCGCTTGTCCGCCTTCGGCAACGGCGTGTTCATGGCCTTGCAGAAAT

10634 TCGAGGATGACCACGGAGAACATTTTCGCCCTGGTTGTCCGTAATAACGGAACAATGGTCT 10575
 -----+-----+-----+-----+-----+-----+-----+-----+
 AGCTCCTACTGGTGCCTCTTGTAAGGCGGACCAACAGGCATTATTGCCTTGTACCAGA

rbs *orf 6* →

10574 GGCCTACCTGGAATTTTGGAGACGGTGCCTGGTACTGGATGAACATGCCATCCGTGATT 10515
 -----+-----+-----+-----+-----+-----+-----+-----+
 CCGCATGGACCTTAAACTCCTGCCACGGCCATGACCTACTTGGTACGGTAGGCACTAA

MetProSerValIle

10514 TCGGGATTCCTAAGTAAGAGAGGTGCCGGACGCGCGAACGTCCGGCAGGACATAAGCAAT 10455
 -----+-----+-----+-----+-----+-----+-----+-----+
 AGCCCTAAGAAATTCATTCTCTCCACGGCTGCGCGCTTGCAGGCCGCTCCTGTATTCTGTTA

SerGlyPheLeuSerLysArgGlyAlaGlyArgAlaAsnValArgGlnAspIleSerAsn

10454 TATAAGGGGATGATTATGCCTGTAACGAAGTGTGAACCAGAAACCACCCGCAAAGCAAGC 10395
 -----+-----+-----+-----+-----+-----+-----+-----+
 ATATTCCTACTAATAACGGACATTGCTTACACTTGGTCTTGGTGGGCGTTTCGTTCG

TyrLysGlyMetIleMetProValThrLysCysGluProGluThrThrArgLysAlaSer

10394 CGTAAATCTGCAAAAACGAGGAAACTGCCCTGTCTGCCCTGCTGGCGCAGACGGAGGAA 10335
 -----+-----+-----+-----+-----+-----+-----+-----+
 GCATTTAGACGTTTTTGGCTCCTTTGACGGGACAGACGGGACGACCGCGTCTGCCTCCTT

ArgLysSerAlaLysThrGlnGluThrAlaLeuSerAlaLeuLeuAlaGlnThrGluGlu

10334 GTGAGCGTGCCGCTGGCCTCACTGATTAAATCACCGCTGAATGTGCGCACGGTGGCGTAT 20275
 -----+-----+-----+-----+-----+-----+-----+-----+
 CACTCGCACGGCGACCGGAGTGACTAATTTAGTGGCGACTTACACGCGTGCCACGGCATA

ValSerValProLeuAlaSerLeuIleLysSerProLeuAsnValArgThrValProTyr

10274 TCTGTGAGTCCGTCAGCGAAGTGGCTGGGTCCATTAAGGGCGTTGGCCTGTGCAGAAT 10215
 -----+-----+-----+-----+-----+-----+-----+-----+
 AGACACCTCAGGCAGTCGCTTGACCGACCCAGGTAATCCCGCAACCGGACAACGTCTTA

SerValGluSerValSerGluLeuAlaGlySerIleLysGlyValGlyLeuLeuGlnAsn

10214 CTGGTCTGCATACCCTGCCTGGTACCGTTACGGTGTGCCCGCAGGCGGTGCCCGACTG 10155
 -----+-----+-----+-----+-----+-----+-----+-----+
 GACCAGACGATGGGACGGACCACTGGCAATGCCACAGCGGCGTCCGCCAGCGGCTGAC

LeuValValHisThrLeuProGlyAspArgTyrGlyValAlaAlaGlyGlyArgArgLeu

10154 GCAGCACTCAACATGCTGGCAGAGCGCAACATCCTTCCGGCTGACTGGCCTGTCCGCGTG 10095
 -----+-----+-----+-----+-----+-----+-----+-----+
 CGTCGTGAGTTGTACGACCGTCTCGGTTGTAGGAAGGCCGACTGACCGGACAGGCGCAC

CTGCTGGAAGCAGAAATGAAGCTGATGCGCTGCATGGCGAAGGTCAGAGCGTGGACGCCG 9315
 9374 -----+-----+-----+-----+-----+-----+-----+ 9315
 GACGACCTTCGTCTTTACTTCGACTACGCGACGTACCGCTTCCAGTCTCGCACCTGCGGC
 LeuLeuGluAlaGluMetLysLeuMetArgCysMetAlaLysValArgAlaTrpThrPro
 GAGATGCGTACCGGAAGCGGTGTGGTGGTGTCTTGGCGTTACGGTAATGTGTGTGTCCAG 9255
 9314 -----+-----+-----+-----+-----+-----+-----+ 9255
 CTCTACGCATGGCCTTCGCCACACCACCACAGGACCGCAATGCCATTACACACACAGGTC
 GluMetArgThrGlySerGlyValValValSerTrpArgTyrGlyAsnValCysValGln
 CGTGGTGTGCAGTTGCGCAGTGAAGATGACGCGGCTGACCGCACGGAACAGGTGCAGGAG 9195
 9254 -----+-----+-----+-----+-----+-----+-----+ 9195
 GCACCACACGTCAACGCGTCACTTCTACTGCGCCGACTGGCGTGCCTTGTCCACGTCCCTC
 ArgGlyValGlnLeuArgSerGluAspAlaAlaAspArgThrGluGlnValGlnGlu
 AAAGCATCAGTGGAGGAAATCAGTCTGCCACTGCTGACGAAAATGTCTTCCGAACGCACG 9135
 9194 -----+-----+-----+-----+-----+-----+-----+ 9135
 TTTCTAGTACCTCCTTTAGTCAGACGGTGACACTGCTTTTACAGAAGGCTTGCGTGC
 LysAlaSerValGluGluIleSerLeuProLeuLeuThrLysMetSerSerGluArgThr
 CTGGCAGTCCAGGCGGCACTCATGCAGCAACCGGACAAATCTCTGGCACTGCTGGCATGG 9075
 9134 -----+-----+-----+-----+-----+-----+-----+ 9075
 GACCGTCAAGTCCGCGTGAGTACGTCTGGCCTGTTTAGAGACCGTGACGACCGTACC
 LeuAlaValGlnAlaAlaLeuMetGlnGlnProAspLysSerLeuAlaLeuLeuAlaTrp
 ACGCTCTGCCTGAATGTGTTTGGCAGCGGAGCGTACAGTAAACCAGCACAAATCAGCCTG 9015
 9074 -----+-----+-----+-----+-----+-----+-----+ 9015
 TCGGAGACGGACTTACACAAACCGTCGCCTCGCATGTCAATTTGGTCGTGTTAGTCGGAC
 ThrLeuCysLeuAsnValPheGlySerGlyAlaTyrSerLysProAlaGlnIleSerLeu
 GAATGTAAACATTATTCGCTGACCAGCGATGCGCCGTCAGGGAAGGAAGGTGCCGCGTTC 8955
 9014 -----+-----+-----+-----+-----+-----+-----+ 8955
 CTTACATTTGTAATAAGCGACTGGTCGCTACGCGGAGTCCCTTCCCTCCACGGCGCAAG
 GluCysLysHisTyrSerLeuThrSerAspAlaProSerGlyLysGluGlyAlaAlaPhe
 ATGGCGCTGATGGCAGAAAAGCCCGTCTTGGCGCCCTGTTACCGGAGGGATGGTCACGG 8895
 8954 -----+-----+-----+-----+-----+-----+-----+ 8895
 TACCGCGACTACCGTCTTTTTTCGGGAGAACCGGGACAATGGCCTCCCTACCAGTGCC
 MetAlaLeuMetAlaGluLysAlaArgLeuAlaAlaLeuLeuProGluGlyTrpSerArg
 GACATGACGACGTTCCCTGTCGCTCAGCCAGGAGGTGCTGTTATCCCTGCTCAGTTTCTGC 8835
 8894 -----+-----+-----+-----+-----+-----+-----+ 8835
 CTGTACTGCTGCAAGGACAGCGAGTCCGCTCCACGACAATAGGGACGAGTCAAAGACG
 AspMetThrThrPheLeuSerLeuSerGlnGluValLeuLeuSerLeuLeuSerPheCys
 ACCGCGTGCAGCATCCACGGCGTTCAGACCCGCGAGTGTGGTCACACGTCACGCAGCCCCG 8775
 8834 -----+-----+-----+-----+-----+-----+-----+ 8775
 TGGCGCACGTCGTAGGTGCCGCAAGTCTGGGCGCTCACACCAGTGTGCAGTGCCTCGGGC
 ThrAlaCysSerIleHisGlyValGlnThrArgGluCysGlyHisThrSerArgSerPro
 CTGGATACGCTGGAGAGCGCCATCGGTTTTACATGCGCGACTGGTGGCAGCCGACAAAA 8715
 8774 -----+-----+-----+-----+-----+-----+-----+ 8715
 GACCTATGCGACCTCTCGCGGTAGCCAAAAGTGTACGCGCTGACCACCGTCCGCTGTTTT
 LeuAspThrLeuGluSerAlaIleGlyPheHisMetArgAspTrpTrpGlnProThrLys
 GCAAACCTTCTTCGGACACCTGAAAAGCCACAAATTATCGCAGCCCTGAATGAGGCCGGA 8655
 8714 -----+-----+-----+-----+-----+-----+-----+ 8655
 CGTTTGAAGAAGCCTGTGGACTTTTTTCGGTGTTTAATAGCGTCCGGACTTACTCCGGCCT
 AlaAsnPhePheGlyHisLeuLysLysProGlnIleIleAlaAlaLeuAsnGluAlaGly
 CTGTCCGGTGCCGCACGGGACGCGAGAAGATGAAGAAAGGTGATGCGGCTGAACATGCA 8595
 8654 -----+-----+-----+-----+-----+-----+-----+ 8595
 GACAGGCCACGGCGTCCCTGCGCCTTCTACTTCTTCCACTACGCCGACTGTACGT
 LeuSerGlyAlaAlaArgAspAlaGluLysMetLysLysGlyAspAlaAlaGluHisAla

8594 GAGCACCATATGAAAGACAACCGCTGGGTTCAGGCTGGATGTGTGCTCCACATCCACAG 8535
 -----+-----+-----+-----+-----+-----+-----+-----+
 CTCGTGGTATACTTTCTGTTGGCGACCCAAGGTCCGACCTACACACGAGGTGTAGGTGTC
 GluHisHisMetLysAspAsnArgTrpValProGlyTrpMetCysAlaProHisProGln
 8534 ACAGATGCCACTGAACGCACCGATAACCTGGCTGATGCCGCCCTGATGTAACAACCACACCG 8475
 -----+-----+-----+-----+-----+-----+-----+-----+
 TGTCTACGGTGACTTGCCTGGCTATTGGACCGACTACGGCGGACTACTTGTGGTGTGGC
 ThrAspAlaThrGluArgThrAspAsnLeuAlaAspAlaAla* *
 CCCC GCCG GAGACGGGGCGGCAGCAAGGGAGATACCGTGTGAAACTGAACTGACCCTG
 8474 -----+-----+-----+-----+-----+-----+-----+-----+ 8415
 GGGCGGCCTCTGCCCCGCCGTCGTTCCCTCTATGGCACTACTTTTACTTGACTGGGAC
 MetLysThrGluLeuThrLeu
 PstI
 8414 AATGCCCTGTCAGTCCATGAACGCACAGGAATATGAAGATATCCGTGCTGCGGGAAGCGAT 8355
 -----+-----+-----+-----+-----+-----+-----+-----+
 TTACGGGACGTCAGGTACTTGCCTGTCCTTATACTTCTATAGGCACGACGCCCTTCGCTA
 AsnAlaLeuGlnSerMetAsnAlaGlnGluTyrGluAspIleArgAlaAlaGlySerAsp
 8354 ATGCGCCGTAATCTCACTCACGAGGTGATGCGTGAAGTGGACGCACCGGCTAACTGGATG 8295
 -----+-----+-----+-----+-----+-----+-----+-----+
 TACGCGGCATTAGAGTGAGTGTCCACTACGCACTTCACCTGCGTGGCCGATTGACCTAC
 MetArgArgAsnLeuThrHisGluValMetArgGluValAspAlaProAlaAsnTrpMet
 8294 ATGAATGGCGAGTATGGCAGTGAGTTCGGGGCTTTTTCCCGTCCAGGTTTCGTTTCACG 8235
 -----+-----+-----+-----+-----+-----+-----+-----+
 TACTTACCGCTCATAACCGTCACTCAAGCCCCGAAAAAGGGCAGGTCCAAGCAAAGTGC
 MetAsnGlyGluTyrGlySerGluPheGlyGlyPhePheProValGlnValArgPheThr
 8234 CCAGCCCACGAACGTTTCCACCTGGCATTATGTTCCGCCGGAGACGTCTCTCAGCTCTGG 8175
 -----+-----+-----+-----+-----+-----+-----+-----+
 GGTGGGTGCTTGCAAAGGTGACCGTAATACAAGCGGCCCTCTGCAGAGAGTCGAGACC
 ProAlaHisGluArgPheHisLeuAlaLeuCysSerProGlyAspValSerGlnLeuTrp
 8174 ATGCTGGTTCTGGTGAATTGTGGTGGACAGCCTTTCGCCGTCGTTTCAGGTGCAACATATC 8115
 -----+-----+-----+-----+-----+-----+-----+-----+
 TACGACCAAGACCACCTAACACCACCTGTCGAAAGCGGCAGCAAGTCCACGTTGTATAG
 MetLeuValLeuValAsnCysGlyGlyGlnProPheAlaValValGlnValGlnHisIle
 8114 TTCACGCCTGTCGCTATCAGTCACACGCTGGCGCTTGCCGCGACACTTGATGCGCAGGGG 8055
 -----+-----+-----+-----+-----+-----+-----+-----+
 AAGTGCGGACAGCGATAGTCAGTGTGCGACCGCAACGGCGCTGTGAATAACGCGTCCCC
 PheThrProValAlaIleSerHisThrLeuAlaLeuAlaAlaThrLeuAspAlaGlnGly
 8054 TACAGTGTAAACGACATCATCCATATCCTGATGGCAGAAAGGAGGTGAGGATGAGCGCAC 7995
 -----+-----+-----+-----+-----+-----+-----+-----+
 ATGTCACAATTGCTGTAGTATAGGACTACCGTCTTCCCTCCAGTCCGTACTCGCGTG
 TyrSerValAsnAspIleIleHisIleLeuMetAlaGluGlyGlyGlnAla*
 MetSerAlaArg
 7994 GTTACGGGCACTGATCCCCCTCAGCGCAGAGCAACAGGCCCATGCAGGCTGTGGCTG 7935
 -----+-----+-----+-----+-----+-----+-----+-----+
 CAAGTGCCCGTACTAGGGGAGTCGCGTCTCGTTGTCCGGCGGTACGTCCGACACCGAC
 SerArgAlaLeuIleProLeuSerAlaGluGlnGlnAlaAlaMetGlnAlaValAlaVal
 7934 TCACAGAACAACGTCGTCGTCAGGGACGCACACTTTCAGCATGGCCTTATGCCAGCGCTT 7875
 -----+-----+-----+-----+-----+-----+-----+-----+
 AGTGTCTTGTGTCAGCAGCAGTCCCTGCGTGTGAAAGTCGTACCGGAATACGGTCCGGAA
 ThrGluGlnArgArgArgGlnGluArgThrLeuSerAlaTrpProTyrAlaSerAlaPhe
 7874 TCTTTCGCTGCCTGAATGGCAGCCGCCGATTTTCGCTGACCGATCTCCGCTTTTTTGCTC 7815
 -----+-----+-----+-----+-----+-----+-----+-----+
 AGAAAGCGACGGACTTACCGTCCGGCGCCTAAAGCGACTGGCTAGAGCGAAAAACGAG

PheArgCysLeuAsnGlySerArgArgIleSerLeuThrAspLeuArgPhePheAlaPro

7814 CTGCGCTGACGAAGGAGGAATTTTCATGGCAACCGTCTCCTGTGGCTGGCTGCCGTGGATA
-----+-----+-----+-----+-----+-----+-----+ 7755
GACGCGACTGCTTCTCCTTAAAGTACCGTTGGCAGAGGACACCGACCGACGGCACCTAT

AlaLeuThrLysGluGluPheHisGlyAsnArgLeuLeuTrpLeuAlaAlaValAspLys

7754 AACTGATTGAAAGTTTTGGTGAAGTCTGTGTTCTTCCCCTGCCATCCGATGCTGGGCATC
-----+-----+-----+-----+-----+-----+ 7695
TTGACTAACTTTCAAAAACCACTTCAGACACAAGAAGGGGACGGTAGGCTACGACCCGTAG

LeuIleGluSerPheGluGluValCysValLeuProLeuProSerAspAlaGlyHisArg

7694 GTCTGTCCCCGTCCGTTCTTTTCGTGAAGGTGAGCGGCGTCGTCAGAAAACCACGCTGA
-----+-----+-----+-----+-----+-----+ 7635
CAGACAAGGCGAGGCAAGGAAAAGCACTTCCACTCGCCGACGAGTCTTTTGGTGCAGCT

LeuPheProSerValProPheArgGluGlyGluArgArgArgGlnLysThrThrLeuThr

7634 CAGAGCAGAAATACAGCCGCCAGCGGGAACGTGAGGCAGAACGACGGGAACCTGGAATACC
-----+-----+-----+-----+-----+-----+ 7575
GTCTCGTCTTATGTGCGCGGTGCGCCCTGCACTCCGTCTGTGTCCTTGACCTTATGG

GluGlnLysTyrSerArgGlnArgGluArgGluAlaGluArgArgGluLeuGluTyrGln

7574 AGACATGTTTTGCTCAGGCGCAGATTGACCTTGCCTTTCATACTCCC GCCACGGTTCGGAA
-----+-----+-----+-----+-----+-----+ 7515
TCTGTACAAAACGAGTCCGCGTCTAACTGGAACGCAAAGTATGAGGGCGGTGCCAGCCTT

ThrCysPheAlaGlnAlaGlnIleAspLeuAlaPheHisThrProAlaThrValGlySer

7514 GCTGGTTGTCCCGCTGGTCTGGTGTGTTGAGGAGCATGATCTGGAACGATTTTCTGGG
-----+-----+-----+-----+-----+-----+ 7455
CGACCAACAGGGCGACCAGACCACAACAACCTCCTCGTACTAGACCTTTGTAAAAGACCC

TrpLeuSerArgTrpSerGlyValValGluGluHisAspLeuGluThrIlePheTrpGly

7454 GGTGGTGCAGGGCGTTTTCCATCACTGTCATCATTTGACCGGTTTTTCTGGCAGGAGGAAC
-----+-----+-----+-----+-----+-----+ 7395
CCACCACGCCCGCAAAGGTAGTGACAGTAGTAAACTGGCCAAAAGACCGTCTCCTTG

TrpCysGlyArgPheProSerLeuSerSerPheAspArgPhePheTrpGlnGluGluPro

7394 CACTCTGGCGGCTGATTTTTGAAGCCGGTGAGGCCGGTCTGGTGCACCGGTACAGATAC
-----+-----+-----+-----+-----+-----+ 7334
GTGAGACCGCCGACTAAAACCTTCGGCCACTCCGGCCAGCACCGTGGCCATGTCTATG

LeuTrpArgLeuIlePheGluAlaGlyGluAlaGlyArgGlyAlaProValGlnIleArg

7334 GTGCACTTGAGCAGTGGATGATCCCGAACAAGCTGGAGAACGCAATA TGATGAAATCAGC
-----+-----+-----+-----+-----+-----+ 7275
CACGTGAACCTCGTCACCTACTAGGGCTTGTTCGACCTCTTGCCTTATACTACTTTAGTCG

AlaLeuGluGlnTrpMetIleProAsnLysLeuGluAsnAlaIle* *

7274 GAAAAGCCGCCGAGAACCCTCGCCCAATGCGGGTGGGGAGGAAAGGCGGAAAAGCCCGAA
-----+-----+-----+-----+-----+-----+ 7215
CTTTTCGGCGGCTCTTGGGAGCGGGTTACGCCACCCCTCCTTTCCGCCTTTTCGGGCTT

direct repeat rbs orf 5

7214 GGGCTTTAAAAATATTATTACGGTGTAGCATAAAAACCTGTACCAATGAGCGGTGACTA
-----+-----+-----+-----+-----+-----+ 7155
CCCGAAATTTTTATAATAATGCCACAATCGTATTTTTGGACATGGTTACTCGCCACTGAT

Met



7154 TGTTAAGAGCAACAAAATACGCATCTATCCTACTCCTGAACAGGCTGCATTTCTCAACG
-----+-----+-----+-----+-----+-----+ 7095
ACAATTCTCGTTGTTTTTATGCGTAGATAGGATGAGGACTTGTCCGACGTAAAGAGTTGC

LeuArgAlaThrLysIleArgIleTyrProThrProGluGlnAlaAlaPheLeuAsnAla

7094 CCCAGTTCGGTGCAGTACGTTTTGCGTACAACAAAGCTCTTCACATCAAAAAACACGCTT
-----+-----+-----+-----+-----+-----+ 7035
GGGTCAAGCCACGCCATGCAAAACGCATGTTGTTTCGAGAAGTGTAGTTTTTTGTGCGAA

GlnPheGlyAlaValArgPheAlaTyrAsnLysAlaLeuHisIleLysLysHisAlaTyr

7034 ACCAGAGACACGGAGTAAATTTAAGTCCGCGTAACGATCTGAAACCGTTGTTGTCTGTGG
 -----+-----+-----+-----+-----+-----+-----+ 6975
 TGGTCTCTGTGCCTCATTTAAATTCAGGCGCATTGCTAGACTTTGGCAACAACAGACACC

 GlnArgHisGlyValAsnLeuSerProArgAsnAspLeuLysProLeuLeuSerValAla

 6974 CCAAAAAGTCCCGTAAATATGCCTGGCTCAAAGAATTTGACTCAATGGCATTGCAACAGG
 -----+-----+-----+-----+-----+-----+-----+ 6915
 GGTTTTTTCAGGGCATTATACGGACCGAGTTTCTTAAACTGAGTTACCGTAACGTTGTCC

 LysLysSerArgLysTyrAlaTrpLeuLysGluPheAspSerMetAlaLeuGlnGlnAla

 6914 CGGTAATCAATATTGATGTGGCATTTCCTCAACTTTTTTAATCCGAAGCTGAAAGCGCGGT
 -----+-----+-----+-----+-----+-----+-----+ 6855
 GCCATTAGTTATAACTACCCGTAAGAGTTGAAAAAATTAGGCTTCGACTTTCGCGCCA

 ValIleAsnIleAspValAlaPheSerAsnPhePheAsnProLysLeuLysAlaArgPhe

 6854 TCCCGACGTTTAAACGCAAACACGGAAAACAATCGAGCTATCACTGTGTGGGATTAAGG
 -----+-----+-----+-----+-----+-----+-----+ 6795
 AGGGCTGCAAATTTGCGTTTGTGCCTTTTGTAGCTCGATAGTGACACAGCCCTAATTC

 PrpThrPheLysArgLysHisGlyLysGlnSerSerTyrHisCysValGlyIleLysVal

 6794 TGCTGGATAATGCCATCAAAATCCCAAACACTGGCCCCGATAGAAGCACGCCTTCATCGTG
 -----+-----+-----+-----+-----+-----+-----+ 6735
 ACGACCTATTACGGTAGTTTTAGGGTTTGACCGGGGTATCTTCGTGCGGAAGTAGCAC

 LeuAspAsnAlaIleLysIleProLysLeuAlaProIleGluAlaArgLeuHisArgGlu

 6734 AACTTCATGGAAAGCTGAAAAGCATCACTGTCAACCCGTTCTGCAACGGGGAAAATACTACG
 -----+-----+-----+-----+-----+-----+-----+ 6675
 TTGAAGTACCTTTCGACTTTTCGTAGTGACAGTGGGCAAGACGTTGCCCTTTATGATGC

 LeuHisGlyLysLeuLysSerIleThrValThrArgSerAlaThrGlyLysTyrTyrAla

 6674 CTTCATTCTTTGTGACGATGGCGCTTCCGCACCAGAGAAAACCCACGTACCTGAAGGAAG
 -----+-----+-----+-----+-----+-----+-----+ 6615
 GAAGGTAAGAAACACTGCTACCGGAAGGCGTGGTCTCTTTGGGTGCATGGACTTCCTTC

 SerIleLeuCysAspAspGlyAlaSerAlaProGluLysProThrTyrLeuLysGluGlu

 6614 AAAAGATTACAGGTCTGGATATGGGACTGGAGCACTATGCCATCACTTCTGATGCTGTAA
 -----+-----+-----+-----+-----+-----+-----+ 6555
 TTTTCTAATGTCCAGACCTATACCCTGACCTCGTGATACGGTAGTGAAGACTACGACATT

 LysIleThrGlyLeuAspMetGlyLeuGluHisTyrAlaIleThrSerAspAlaValLys

 6554 AAGTTCCGAATCCTCGCCATCTTATCAATGTAAGTCGTAATCTGCGACGCAAACAAAAG
 -----+-----+-----+-----+-----+-----+-----+ 6495
 TTCAAGGCTTAGGAGCGGTAGAATAGTTACATTCAGCATTAGACGCTGCGTTTGTTTTTTC

 ValProAsnProArgHisLeuIleAsnValSerArgAsnLeuArgArgLysGlnLysAla

 6494 CATTATCCCGTAAGCAAAGGGAAGCGGAATCGTAAAAAGGCCAGAATCCGTCTTGCGG
 -----+-----+-----+-----+-----+-----+-----+ 6435
 GTAATAGGGCATTTCGTTTTCCCTTCGCGCTTAGCATTTTTCCGGTCTTAGGCAGAACGCC

 LeuSerArgLysGlnLysGlySerAlaAsnArgLysLysAlaArgIleArgLeuAlaAla

 6434 CCTTACACGAACGGGTGGCGAATGCCCGTGTGATTTTCAACACAAGCTCTCTTGCGCGA
 -----+-----+-----+-----+-----+-----+-----+ 6375
 GGAATGTGCTTGCCACCGCTTACGGGCACGACTAAAAGTTGTGTTCGAGAGAACGCGCT

 LeuHisGluArgValAlaAsnAlaArgAlaAspPheGlnHisLeuSerCysAlaIle

 Sall
 ↓
 6374 TTTCGACGAAAACCAAGCGGTAATTGTGGAGACGCTGAAAACAGCCAACATGATGAAAA
 -----+-----+-----+-----+-----+-----+-----+ 6315
 AACAGCTGCTTTTGGTTCGCCATTAACACCTCTGCGACTTTTGTGCGTTGTACTACTTTT

 ValAspGluAsnGlnAlaValIleValGluThrLeuLysThrAlaAsnMetMetLysAsn

 6314 ACCACAATCTGGCAAGAGCGATAGGTGATGCTGGCTGGCACAGCTTCATCACAAGCTGG
 -----+-----+-----+-----+-----+-----+-----+ 6255
 TGGTGTAGACCGTTCTCGCTATCCACTACGACCGACCGTGTGAAGTAGTGTTCGACC

HisAsnLeuAlaArgAlaIleGlyAspAlaGlyTrpHisSerPheIleThrLysLeuGlu
 6254 AGTACAAGGCAGCGACAAAAGGCGTTCACCTGGTAAACTCGACCAGTGGTTTGCCAGTT
 -----+-----+-----+-----+-----+-----+-----+ 6195
 TCATGTTCCGTCGCTGTTTTCCGCAAGTGGACCATTTTGAGCTGGTCACCAAACGGTCAA

 TyrLysAlaAlaThrLysGlyValHisLeuValLysLeuAspGlnTrpPheAlaSerSer
 6194 CGAAAACGTGCCATTGTTGTGGTTACAAAATGCCGAAATGCCACTTCATAAACGCATCT
 -----+-----+-----+-----+-----+-----+ 6135
 GCTTTTGCACGGTAACAACACCAATGTTTTACGGCCTTTACGGTGAAGTATTTGCGTAGA

 LysThrCysHisCysCysGlyTyrLysMetProGluMetProLeuHisLysArgIleTrp
 6134 GGCGATGCCCTGAGTGTGGAGCAGAGCACGCCGATATCAATGCCGCACTCAACATCC
 -----+-----+-----+-----+-----+-----+ 6075
 CCGCTACGGGACTCACACCTCGTCTCGTGTGGCGCTATAGTTACGCCGTGAGTTGTAGG

 ArgCysProGluCysGlyAlaGluHisAspArgAspIleAsnAlaAlaLeuAsnIleArg
 6074 GGCAAAAAGGAATACTGGAATAAAGGCGGGGACTCGTCTGCTCTGCCCATGGAGGCC
 -----+-----+-----+-----+-----+-----+ 6015
 CCGTTTTTCCCTTATGACCTTGATTTCCGCCGCCCTGAGCAGCAGAGACGGGTACCTCCGG

 GlnLysGlyIleLeuGluLeuLysAlaAlaGlyLeuValValSerAlaHisGlyGlyGln
 6014 AGCGTAAATCCGTCGCACAGACGGTTGCGGCCTGAGGAAGTGGGAAGCCTCGCCCGACAGG
 -----+-----+-----+-----+-----+-----+ 5955
 TCGCATTTAGGCAGCGTGTCTGCCAACGCCGACTCTTACCCTTCGGAGCGGGCTGTCC

 ArgLysSerValAlaGlnThrValAlaAla* **direct repeat**
 5954 GCGGGGAGCAGTCACGGGTGGTCTGGAGGTCATGCGGGCGTGTCTCTGCACTCGCCGGAA
 -----+-----+-----+-----+-----+-----+ 5895
CGCCCTCGTCAGTGCCACCAGACCTCCAGTACGCCGCACAGGAGAGCTGAGCGGCCT

 TAAGGAAGTCGCCGGCGGCTCCGCTTTTACCCGGCCATGCGGGGCATGGCCTTGTGGGTT
 5894 -----+-----+-----+-----+-----+-----+ 5835
 ATTCCTCAGCGGCCCGGAGGCGAAAATGGGCCGTACGCCCGTACCGGAACACCCAA

 TTCAGCTCTGTGGCCTCAGCGTCGTGTGCGGGCTGTGCCGTGCCTCCATCTTAGCCGGGC
 5834 -----+-----+-----+-----+-----+-----+ 5775
 AAGTCGAGACACCGGAGTCGCAGCACACGCCCGACACGGCACGGAGGTAGAATCGGCCCG

 TGGCAGGGATGCAAGGGTACGCTTCGCCGCTGCGGTCACCCGGTCCCTCCTTCCCCTCTG
 5774 -----+-----+-----+-----+-----+-----+ 5715
 ACCGTCCCTACGTTCCCATGCGAAGCGGCGACGCCAGTGGGCCAGGGAGGAAGGGCAGAC

 CCGTGATTTCCGGTCGTTACCCGCTCAGATTTGCGGGTCTCACCCCTCCAACCTCCCGC
 5714 -----+-----+-----+-----+-----+-----+ 5655
 GGCATAAAAGGCCAGCAAGTGGGCGAGTCTAAAGCGCCAGAGTGGGAGGTTGAGGGCG

 CAGCCGTGTGCGTCAGGCTGCGGCTTCCCTTGCGCCCTGTGCATCCCCGCCTTATCGCCC
 5654 -----+-----+-----+-----+-----+-----+ 5595
 GTCGGCACACGCAGTCCGACGCCAAGGAACCGGGACACGTAGGGGCGGAATAGCGGG

 GGCTTTTATGGAGGCACGGCACCCCGGTTGCCGAGAATGTGACTATCGAGGATTCG
 5594 -----+-----+-----+-----+-----+-----+ 5535
 CCGAAAATACCTCCGTGCCGTGGCGGGCCACGGCGTCTTGTACTGATACTCCTAAGC

ardA →
 5534 GGAATGCTGTTGTTGCACCTGCTGTATACGTTGGAACCTGGCACAAATACAACCTGTGGA
 -----+-----+-----+-----+-----+-----+ 5475
 CCTTACAGACAACAACGTGGACGACATATGCAACCTTGGACCGTGTATGTTGACACCT

 MetSerValValAlaProAlaValTyrValGlyThrTrpHisLysTyrAsnCysGly
 5474 AGCATCGCCGAACGCTGGTTTGCACCTGACCAGTGTGATGATGAGCGGACTTTTTCGCC
 -----+-----+-----+-----+-----+-----+ 5415
 TCGTAGCGGCTTGCAGACAACTGGACTGGTGCAAACTACTACTCGCGCTGAAAAGCGG

 SerIleAlaGluArgTrpPheAspLeuThrThrPheAspAspGluArgAspPhePheAla

GCCTGCCGTGCTCTTACCAGGATGAAGCCGATCCTGAACTGATGTTTCAGGATTATGAG
 5414 -----+-----+-----+-----+-----+-----+-----+ 5355
 CGGACGGCAGGAGAAGTGGTCTACTTCGGCTAGGACTTGACTACAAAGTCCTAATACTC
 AlaCysArgAlaLeuHisGlnAspGluAlaAspProGluLeuMet PheGlnAspTyrGlu
 GGATTCCTCCGGGGAATATGGCCTCTGAATGCCATATCAACTGGGCCTGGGTGAAGGCTTC
 5354 -----+-----+-----+-----+-----+-----+-----+ 5295
 CCTAAGGGCCCCTTATACCGGAGACTTACGGTATAGTTGACCCGGACCCAACCTCCGAAG
 GlyPheProGlyAsnMetAlaSerGluCysHisIleAsnTrpAlaTrpValGluGlyPhe
 CGCCAGGCACGGGATGAAGGCTGCGAAGAGGCTTATCGTCTCTGGGTGGAGGATACCGGT
 5294 -----+-----+-----+-----+-----+-----+-----+ 5235
 GCGGTCCGTGCCCTACTTCCGACGCTTCTCCGAATAGCAGAGACCCACCTCCTATGGCCA
 ArgGlnAlaArgAspGluGlyCysGluGluAlaTyrArgLeuTyrValGluAspThrGly
 GAGACGGATTTTGACACCTTCCGCGATGCCTGGTGGGGCGAGGCTGACAGTGAGGAGGCT
 5234 -----+-----+-----+-----+-----+-----+-----+ 5175
 CTCTGCCTAAAACCTGTGGAAGGCGCTACGGACCACCCCGCTCCGACTGTCACTCCTCCGA
 GluThrAspFPheAspThrPheArgAspAlaTrpTrpGlyGluAlaAspSerGluGluAla
 TTTGCGGTGAGTTCGCCAGTGATACCGCCTGCTGGCTGACGTGCCGGAGACGGTGGCG
 5174 -----+-----+-----+-----+-----+-----+-----+ 5115
 AAACGCCAACTCAAGCGGTCACTATGGCCGGACGACCGACTGCACGGCCTCTGCCACCGC
 PheAlaValGluPheAlaSerAspThrGlyLeuLeuAlaAspValProGluThrValAla
 CTCTATTTGACTATGAGGCGTATGCGCGGGATTTATTCCTGGACTCCTTCACCTTTATT
 5114 -----+-----+-----+-----+-----+-----+-----+ 5055
 GAGATAAACTGATACTCCGCATACGCGCCCTAAATAAGGACCTGAGGAAGTGGAATAA
 LeuTyrPheAspTyrGluAlaTyrAlaArgAspLeuPheLeuAspSerPheThrPheIle
 GACGGTCATGTGTTCCGTCGGTGAATTACTCCCCGCTCCGGCGGGGTATGGCCTGCCA
 5054 -----+-----+-----+-----+-----+-----+-----+ 4995
 CTGCCAGTACACAAGGCAGCCACTAAATGAGGGGGCGAGGCCGCCCCATACCGGACGGT
 AspGlyHisValPheArgArg*
 GGCCGCCGAAAACCTGGCAGGCTTACGCCTGCCTGAAATGCAGGAGGGCCGCGCCTGC
 4994 -----+-----+-----+-----+-----+-----+-----+ 4935
 CCGCGGCCTTTTGGGACCGTCCGAATGCGGACGGACTTACGTCTCCCGGCGCGGACG
 GGCGTGCCCTTCTGCCCCGTCATGGTGTGAGGCATTTTCTGTATGTGTTCTGTTC
 4934 -----+-----+-----+-----+-----+-----+-----+ 4875
 CCGCGACGGGAAGGACGGGCAGTACCACACTCCGTA AAAAGACAATACAGCAAAGACAAC
 TTCTGATTTCCGGAGGCGGGACACCTTCGGCTCCCGGCTGCCTGGCGGCATCCGGTGT
 4874 -----+-----+-----+-----+-----+-----+-----+ 4815
 AAGACTAAAGCCTCCGCCCTGTGGAAGCCGAGGGCCGACGGACCGCCGTAGGCCAGCA
 CAGGCCGACTCCGACAGGCAGCTGCTGGCGACCCGCCAGTCTCCGCCGAACCCGCTGG
 4814 -----+-----+-----+-----+-----+-----+-----+ 4755
 GTCCGGGCTGAGGCTGTCCGTCGACACCGCGTGGGCGGTGACAGGCGGCTTGGGCGACC
 CGCGGTTTCGGGCTGCCGCTACGGTCCGATGTCTGAAAAACCAGGCTTGCAGCGACCGTT
 4754 -----+-----+-----+-----+-----+-----+-----+ 4695
 GCGCCAAAGCCCGACGGCGATGCCAGGCTACAGACTTTTGGTCCGAACGTCGCTGGCAA
 CCGGTCGCGCGCCCTGTGCGCGCGGAAACGGCGC
 4694 -----+-----+-----+-----+-----+-----+-----+
 GGCCAGCGCGCGGGGACAGCGCGGCTTTGCCGCGAGGGACGTGTGGACTGCAACACAT 4635

ssi2

4634 TCCGGTGACGCGGACTACCGCTCAGTTGGAAGGCGCGGTGGGGACGCCGACAAAGG 4575

4574 GCGACACCTGAAAGGCCCGACCTCATTGGTGTAGGGCAAAGTCGTCAGGTCCCCTTCT 4515

4514 ATATGGCAGGTGTTTTAAAAATCGGTATTTGGGCCGTTTTTTTAAAACGCAGGCGGGGGCC 4455

4454 TGTCGACGGGGCACTGCACCACTGCCTGCGGGGTAGCCTTTCCAGGTGACGCCCTCTTT 4395

4394 GGGGCTTCTTACCTGCCGTCTTCCCGAAAGCGGGGAGGCCCTCGTTGGCCAGCCCCG
GGGAGCAACCGGTCGGGCC
-----+-----+ 4335

4334 CGGCGACCTGATTCAGCAGAAGGAGATGTTTTTCATGACAGTCAGCAGCAGATTCCGTT
-----+-----+-----+-----+-----+-----+-----+ 4275
GCCGCTGGACTAAGTCGTCTTCTCTACAAAAGTACTGTCAGTCGTCTGTAAAGCAA

MetThrValSerSerThrIleSerVal

4274 TTTTGGCCGACGGGGTGTTCGTACCGTTTACTGCCACCTGCACGGGAGCCGACCTGG
-----+-----+-----+-----+-----+-----+-----+ 4215
AAAACGGCGCTGCCCCACAAGGCATGGCAAATGACGGTGGACGTGCCCTCGGCTGGACC

PheCysArgAspGlyValPheArgThrValTyrCysHisLeuHisGlyGluProThrTrp

4214 AACGGTCGCATCCTGCATACTCACTATGCCACCGGTCAGCAGGCCGAAGCCCTGGTTGAA
-----+-----+-----+-----+-----+-----+-----+ 4155
TTGCCAGCGTAGGACGTATGAGTGATACGGTGGCCAGTCGTCCGGCTTCGGGACCAACTT

AsnGlyArgIleLeuHisThrHisTyrAlaThrGlyGlnGlnAlaGluAlaLeuValGlu

4154 CACGGTGATATCCGTTGCCCTCGGTCCCCGTTGTGACAAACCCGCGGGCATACTTCAG
-----+-----+-----+-----+-----+-----+-----+ 4095
GTGCCACTATAGGCAACGGAGCCAGGGGCAACACTGTTTGGGCGGCCCGTATGCGAAGTC

HisGlyAspIleArgCysLeuGlyProArgCysAspLysProAlaGlyHisThrLeuGln

4094 AACCCGGTGGACGGTGTGACGGCTTATTACGGACGTGACAGTGGTTTCCGGATGGACAGT
-----+-----+-----+-----+-----+-----+-----+ 4035
TTGGGCCACCTGCCACACTGCCAATAATGCCTGCACTGTCACCAAAGGCCTACCTGTCA

AsnProValAspGlyValThrAlaTyrTyrGlyArgAspSerGlyPheArgMetAspSer

4034 GAGGCGCGTGAGTACCGTTCTTTCAGGGAGGCTATTGCCACTGAAAGCACTGAAGAGGTG
-----+-----+-----+-----+-----+-----+-----+ 3975
CTCCGCGCACTCATGGCAAGAAAGTCCCTCCGATAACGGTGACTTTCGTGACTTCTCCAC

GluAlaArgGluTyrArgSerPheArgGluAlaIleAlaThrGluSerThrGluGluVal

3974 CGCTTCCATTATGTGTTTCATCGACGGCTACTGGAAGGTGATGTACCGCACGCCGAAGGC
-----+-----+-----+-----+-----+-----+-----+ 3915
GCGAAGGTAATACACAAGTAGCTGCCGATGACCTTCCACTACATGGCGTGCGGCCTTCCG

ArgPheHisTyrValPheIleAspGlyTyrTrpLysValMetTyrArgThrProGluGly

3914 TGGAAGATGAAAGCGCTCGCGCTGGCACTGCGTGTCCGAAATGAAAAAAGGCGGG
-----+-----+-----+-----+-----+-----+-----+ 3855
ACCTTCTACTTTTCGCGAGCGCGACCGTGACGCAGCGACAGGCTTACTTTTTTCCGCC

TrpLysMetLysAlaLeuAlaLeuAlaLeuArgArgCysProLys*

3854 GCGATAAACCTGCCTCTCTCCGCCGTTACTTCCCGCCAGGAAGAGCGGCAGTCATCAA
-----+-----+-----+-----+-----+-----+-----+ 3795
CGCTATTTGGGACGGAGAGGCGCAATGAAGGGCGGTCCTTCTCGCCGTCAGTAGTT

orf 3 →

3794 CAACATAAGGAGTAATATTCTGAGACCATCAATTATCTTCGCAACCGCCGAGTATGTAA
-----+-----+-----+-----+-----+-----+-----+ 3735
GTTGTATTCTCATTATAACACTCTGGTAGTTAATAGAAGCGTTGGCGGCTCATACATTT
ValArgProSerIleIlePheAlaThrAlaGluTyrValLys

GCGTCTGCGTGAAGAGTGCCTGCGGGAGAATAAACCCCTGCACCGCCATACCCGCTTCAG
 3734 -----+-----+-----+-----+-----+-----+-----+ 3675
 CGCAGACGCACTTCTCACGGACGCCCTTATTTGGGGACGTGGCGGTATGGGGCAAGTC
 ArgLeuArgGluGluCysLeuArgGluAsnLysProLeuHisArgHisThrArgPheArg

 ACGTGAGGAGCTGGCACAGGATGAGATTAACCCGGACGCTCTGGCGATGAGCGGCCATAT
 3674 -----+-----+-----+-----+-----+-----+-----+ 3615
 TGCAGTCTCGACCGTGTCTACTCTAATTGGGCCTGCAGGACCGCTACTCGCCGGTATA
 ArgGluGlnLeuAlaGlnAspGluIleAsnProAspValLeuAlaMetSerGlyHisIle

 PstI
 CGCCAGACGCTGCGAGTGGAGCAGAAGCGGGTGGCGTATCCCCGCTATGAAAGTCAGCGAATG
 3614 -----+-----+-----+-----+-----+-----+-----+ 3555
 GCGGTCTGCGACGTCACCTCGTCTTCGCCACGCATAGGGGCGATACTTTCAGTCGCTTAC
 AlaArgArgCysSerGluGlnLysArgValArgIleProAlaMetLysValSerGluTrp

 GGGCCACCTGCTCCGCGCGCTTAAAATAGAGCGGGTCTGCCACTGATTAACCCTTGCCGT
 3554 -----+-----+-----+-----+-----+-----+-----+ 3495
 CCCGGTGGAGGCGCGCAACTTTATCTCGCCAGACGGTACTAATTGGGAACGGCA
 GlyHisLeuLeuArgAlaLeuGluIleGluArgValCysHis *
 CTGCCTGTGCAGGCGGCAGGATACGTTTCAGACATCTCACCAGGAGGAACCGTAGTAAGA
 3494 -----+-----+-----+-----+-----+-----+-----+ 3435
 GACGGACACGTCGCCCGTCTATGCAAGTCTGTAGAGTGGTCTCCTTGGCACTCATTCT

 ValSerLysLys

 AGAAAAACACCACAACGCCACGCCGATGATGCCGCTTCCGGTCGTTCCCTGGCGAATC
 3434 -----+-----+-----+-----+-----+-----+-----+ 3375
 TCTTTTGTGGTGTGCGGGTGCGGCGTACTACGGCGCAAGGCCAGCAAGGACCGCTTAG

 LysAsnThrThrThrProThrProHisAspAlaAlaPheArgSerPheLeuAlaAsnPro

 CCGACGTCGCCAGAGATTTTCTGGAAGTGCATCTTCCTGCGGAGTACCGGCAGTTGTGCG
 3374 -----+-----+-----+-----+-----+-----+-----+ 3315
 GGCTGCAGCGGTCTCTAAAAGACCTTGACGTAGAAGGACGCCTCATGGCCGTC AACACGC

 AspValAlaArgAspPheLeuGluLeuHisLeuProAlaGluTyrArgGlnLeuCysAsp

 ACCTGTCCACGCTGAAGCTGGAACCCGCCACCTTTGTTGAGCCGGACCTGCATCAGTACG
 3314 -----+-----+-----+-----+-----+-----+-----+ 3255
 TGGACAGGTGCGACTTTCGACCTTGGGCGGTGGAACAACCTCGGCCTGGACGTAGTCATGC

 LeuSerThrLeuLysLeuGluProAlaThrPheValGluProAspLeuHisGlnTyrAla

 CCAGCGATATCCTCTGGAGCGTAAAACCCGGGGTGAAGATGGTTATGTTTATACGC
 3254 -----+-----+-----+-----+-----+-----+-----+ 3195
 GGTGCTATAGGAGACCTCGCACTTTTGGTGGCCCCACTTCTACCAATACAAATATGCG

 SerAspIleLeuTrpSerValLysThrThrGlyGlyGluAspGlyTyrValTyrThrLeu

 TCATCGAGCACCAGACACCGAAAATCTGTACATGCCTTTTCGCATGTTACGCTACAGTG
 3194 -----+-----+-----+-----+-----+-----+-----+ 3135
 AGTAGCTCGTGGTCTCGTGGCTTTTAGACATGTACGGAAAAGCGTACAATGCGATGTAC

 IleGluHisGlnSerThrGluAsnLeuLysMetProPheArgMetLeuArgTyrSerVal

 TGGCGGCGATGCAGAGACATCTGGAGCAGCACAACCGTTGCCACTGGTCATTCGGGTGC
 3134 -----+-----+-----+-----+-----+-----+-----+ 3075
 ACCGCCGCTACGCTCTGTAGACCTCGTGTGTTTGAACGGTGACCAGTAAGGCCACG

 AlaAlaMetGlnArgHisLeuGluGlnHisLysThrLeuProLeuValIleProValLeu

 TGTTCTATCACGGTGAGCGCAGCCCGTACCCGTACAGCATGAACTGGCTGGACTGTTTTG
 3074 -----+-----+-----+-----+-----+-----+-----+ 3015
 ACAAGATAGTCCACTCGCGTCCGGCATGGGCATGTCGTACTTGACCGACCTGACAAAAC

 PheTyrHisGlyGluArgSerProTyrProTyrSerMetAsnTrpLeuAspCysPheGlu

 AGAATCCGGCGCTGCGGCTAAAATATACACAAAGCCGTTCCGCTGGTTGATATCACTG
 3014 -----+-----+-----+-----+-----+-----+-----+ 2955
 TCTTAGGCCCGAACCGGATTTTATATGTGTTTCGGCAAAGGCGACCAACTATAGTGAC

 AsnProAlaLeuAlaAlaLysIleTyrThrLysProPheProLeuValAspIleThrVal

2114 GACATCCACCTGTTCCGCATTTTCCCAGGCGTGCCATGTTCTCTGTGTCACGCCAAATTC 2055
 -----+-----+-----+-----+-----+-----+-----+ 2055
 CTGTAGGTGGACAAGGCGTAAAAGGGTCCGCACGGTACAAGAGACACAGTGC GGTTTAAAG

2054 TCTGGCCGCTTTTTTCGCGTGACCAGAGCATGCTTTTTTCGCCAGATTTCGCAGTTCCCAGCC 1995
 -----+-----+-----+-----+-----+-----+-----+ 1995
 AGACCGGCGAAAAAGCGCACTGGTCTCGTACGAAAAAGCGGTCTAAGCGTCAAGGGTCGG

1994 GGTCATAAAACACCTTTGAAAAAGTGAAAAAACTTCACTTTATTTTATCACCATAACCAC 1935
 -----+-----+-----+-----+-----+-----+-----+ 1935
 CCAGTATTTTGTGGAACTTTTCACTTTTTTGAAGTGAATAAAATAGTGGTATTGGTG

1934 CAAATCAGCGCGAAAAATAAGTGTGAAAAAGTGAAGGAATTTCACTTTTTTTGATTGTAG 1875
 -----+-----+-----+-----+-----+-----+-----+ 1875
 GTTTAGTCGCGCTTTTTATTCACACTTTTTCACTTCCTTAAAGTGAAAAAAACTAACATC

1874 GTGTCGTGCTGTTTTCTCTGTTTTCACCTGTGCGGGAGTGGTGTGCCAGTGACCACCACA 1815
 -----+-----+-----+-----+-----+-----+-----+ 1815
 CACAGCACGACAAAAGAGACAAAAGTGGACACGCCCTCACCACACGGTCACTGGTGGTGT

1814 TGCCATTGTTGCGCCGTCGTTTTTCTTC *ssi1*
 -----+-----+-----+-----+-----+-----+-----+ 1755
 ACGGTAACAACGCGGCAGCAAAAAGAAGTCAGCGCGCCTACGCCGCGAGGGACGTGTGG 1755

1754 AATGCCACACATTCCGGTGACGCGGACTACCGCGTTCGACAGAGACGGACAGACGGCGGCG 1695

1694 TGACGCCGACAAAGAGGCGTCTACCTGAAAGCCCCCATGTATGTGCTGGTGTAGGGCAA 1635

1634 AGTCAGCAGTCCCCTCTCTACATGGTTCGCTGTTTTGAAAAACGGCGTCCGCAGTTTTTG 1575

1574 AAAACGCAGACGGGGCCTGTTGACGGGGCACTGCACCACTCGCAGGGGGCAGCCTTTCC 1515

1514 AGGTGACGCCCTTTAGGGGCTTCTGGGCTGCCGTCTTCCTCGAAGCGAGGAGGCCCTT 1455
 -----+-----+-----+-----+-----+-----+-----+ 1455
 ATGACCTGGCTGACCCGGCAGTCGTGACGATGAGGAGACAATGTGATGATCAGACTTTA
 TACTGGACCGACTGGGCCGTCAGCACTGCTACTCCTCTGTTACACTACTTAGTCTGAAAT
 MetAsnGlnThrLeu

1394 CCCACTGCTGACCTGAATACTGCCGGCAGCAGATGTTATTCCGTCTGTGGCCATCGAC 1335
 -----+-----+-----+-----+-----+-----+-----+ 1335
 GGGTGACGACTGGACTTATGACGGCCGTCGCTGTCTACAATAAGGCAGACACCGGTAGCTG
 ProThrAlaAspLeuAsnThrAlaGlyThrThrAsnValIleProSerValAlaIleAsn

1334 CGCATCATCGCGCAGCGTAACGAAGGTATTGCACTGTTTCATGCAGGCGATGGAATGCCTG 1275
 -----+-----+-----+-----+-----+-----+-----+ 1275
 GCGTAGTAGCGCTCGCATTGCTTCCATAACGTGACAAGTACGTCCGCTACCTTACGGAC
 ArgIleIleAlaGlnArgAsnGluGlyIleAlaLeuPheMetGlnAlaMetGluCysLeu

1274 GCGACAGCGCGCAGGATTCTGCTTGATGCGTCAGGTGATATTTTTCTTTACGGTTTTGAA 1215
 -----+-----+-----+-----+-----+-----+-----+ 1215
 CGCTGTCGCGCTCCTAAGACGAACCTACGCAGTCCACTATAAAAAGAAATGCCAAAACCTT

AlaThrAlaArgArgIleLeuLeuAspAlaSerGlyAspIlePheLeuTyrGlyPheGlu
 GACTGCGTGACCGACTCCGTTCCGGTGCATGGATAAACCGGAAGAAGCGAAAAGGAATATC
 1214 -----+-----+-----+-----+-----+-----+-----+ 1155
 CTGACGCACTGGCTGAGGCAAGCCACGTACCTATTTGGCCTTCTTCGCTTTTCCTTATAG

AspCysValThrAspSerValArgCysMetAspLysProGluGluAlaLysArgAsnIle
 ACCCGTCTTGCCGACCGGAAAATCTGGGACCGCTGATGACGGATACGGGGATGTACACC
 1154 -----+-----+-----+-----+-----+-----+-----+ 1095
 TGGGCAGAACGGCTGGCCTTTTAGACCCTGGCGGACTACTGCCTATGCCCTACATGTGG

ThrArgLeuAlaAspArgLysIleTrpAspArgLeuMetThrAspThrGlyMetTyrThr
 TTCATGAGTTCATGTCTAGCGTGTAGTGGAAACAGCCAGCTGATGAGCGACACCTGTCTCT
 1094 -----+-----+-----+-----+-----+-----+-----+ 1035
 AAGTACTCAAGTACAGTCGCACTACTCACCTTGTTCGGTTCGACTACTCGCTGTGGACAGGA

PheMetSerSerCysGlnArgAspGluTrpAsnSerGlnLeuMetSerAspThrCysPro
 GAAATCACCCCTGGACAATGTACTGGCAACTTTTCGCCATCTGAATGCCAGTAAGATGCAG
 1034 -----+-----+-----+-----+-----+-----+-----+ 975
 CTTTAGTGGGACCTGTTACATGACCGTTGAAAAGCGGTAGACTTACGGTCATTCTACGTC

GluIleThrLeuAspAsnValLeuAlaThrPheArgHisLeuAsnAlaSerLysMetGln
 ACATTTGAACAGGGACTGATTGATGTCTACCGGAAATTGTCATGGGATTACAGAACCAAT
 974 -----+-----+-----+-----+-----+-----+-----+ 915
 TGTAACCTGTCCCTGACTAACTACAGATGGCCTTTAACAGTACCCTAATGTCTTGGTTA

ThrPheGluGlnGlyMetIleAspValTyrArgLysLeuSerTrpAspTyrArgThrAsn
 AATCCCTGCCGCTCTGGCAAGAGAATCATTATTGAAAACCTGCTGTACCGCTGGAGTAAC
 914 -----+-----+-----+-----+-----+-----+-----+ 855
 TTAGGGACGGCAGACCCGTTCTCTTAGTAATAACTTTTGGACGACATGGCGACCTCATTTG

AsnProCysArgLeuGlyLysArgIleIleIleGluAsnLeuLeuTyrArgTrpSerAsn
 PstI
 GGGCGTGTTACGCTGGACTGTCAGCGGACGGGAGGCACTGGATGACCTGGTACGTCCTGTTT
 854 -----+-----+-----+-----+-----+-----+-----+ 795
 CCCGCACAATGCGACCTGACGTGCGCTGCCCTCCGTGACCTACTGGACCATGCAGGCAAA

GlyArgValThrLeuAspCysSerGlyArgGluAlaLeuAspAspLeuValArgProPhe
 TATCTGCTGGAGGGGCGCAACGTTTCCTGACTTCCGGAGCAGCATCGGGGCGCAGTACGGT
 794 -----+-----+-----+-----+-----+-----+-----+ 735
 ATAGACGACCTCCCCGCTTGCAAGGACTGAAGGCCTCGTCGTAGCCCCGCGTCATGCCA

TyrLeuLeuGluGlyArgAsnValProAspPheArgSerSerIleGlyAlaGlnTyrGly
 GAATTTATCGGGAACGGCGACAATGTCCGTAAGCTGTTAGAAGGGGAATATTTTACGGTG
 734 -----+-----+-----+-----+-----+-----+-----+ 675
 CTTAAATAGCCCTTGCCGCTGTTACAGCCATTCGACAATCTTCCCCTTATAAAATGCCAC

GluPheIleGlyAsnGlyAspAsnValGlyLysLeuLeuGluGlyGluTyrPheThrVal
 CGTGGCTACCAGAAAGGGACCGTACACATTGTCTTTAAGCGTCCCTGACCTTGTGAAAAA
 674 -----+-----+-----+-----+-----+-----+-----+ 615
 GCACCGATGGTCTTCCCTGGCATGTGTAACAGAAATTCGCAGGACTGGAACAACCTTTTT

ArgGlyTyrGlnLysGlyThrValHisIleValPheLysArgProAspLeuValGluLys
 CTGAATGATATTATTGCACGGCATTATCCTGGTTCATTGCCGCCACGAGTCTGATTAAAC
 614 -----+-----+-----+-----+-----+-----+-----+ 555
 GACTTACTATAATAACGTGCCGTAATAGGACCAAGTAACGGCGGTGCTCAGACTAATTTG

LeuAsnAspIleIleAlaArgHisTyrProGlySerLeuProProArgVal*
 AGAAAAGCCCTGGTATTTATGCCGGGCTTTTTTTCTGTAATCATCATAATATCAGAACA
 554 -----+-----+-----+-----+-----+-----+-----+ 495
 TCTTTTCGGGACCATAAATACGGCCCCGAAAAAAGACATTAGTAGTATTATAGTCTTGT

494 TGATGTGACATTGTCACATCAATTGTTGTTTCATATCGATGCCATCTTTCCTGGCTGAATC 435
-----+-----+-----+-----+-----+-----+-----+
ACTACACTGTAACAGTGTAGTTAACAACAAGTATAGCTACGGTAGAAAGGACCGACTTAG

ClaI

434 TGTCTTTTGTAGATGCTGTTTTAGAGGAAGAACGGCGGGATGGCTTTTTCGATTCTGCTTC 375
-----+-----+-----+-----+-----+-----+-----+
ACAGAAAAATCTACGACAAAATCTCCTTCTTGCCGCCCTACCGAAAAAGCTAAGACGAAG

374 CCGAATAATGTCACGAAGTGCCTTCTCTCCTACTTCATACCCTGCATTTTGTAGCGTGTC 315
-----+-----+-----+-----+-----+-----+-----+
GGCTTATTACAGTGCTTACCGAAGAGAGGATGAAGTATGGGACGTAAAAACTCGCACAG

314 ACGAACATCTGCCAGTGTATAACCCCTTGTCTTACCAGAACATAATGTCATCACGAAT 255
-----+-----+-----+-----+-----+-----+-----+
TGCTTGTAGACGGTCACATATTGGGAAACAGGAATGGTCTTGTATTACAGTAGTGCTTA

BglIII

254 GGCATCAAGAAAAATCACGCAATGTTTTCCGTTGAGCGGTGAGATCTGGCAATCCGATAA 195
-----+-----+-----+-----+-----+-----+-----+
CCGTAGTTCTTTTAGTGCGTTACAAAAGGCAACTCGCCACTCTAGACCGTTAAGGCTATT

194 AGCAGCTTTTGCCAGCTGGATGTCATCATCTGAGTAAAATTTTTAGAAAGCCATTGGCAC 135
-----+-----+-----+-----+-----+-----+-----+
TCGTCGAAAACGGTCGACCTACAGTAGTAGACTCATTTTAAAAAATCTTCGGTAACCGTG

134 TTTCTCTCCGTTTCATGTTGAAGTCTGATTTGATCATCAAAAAGTACTCCTGGCAGGCT 75
-----+-----+-----+-----+-----+-----+-----+
AAAGAGAGGGCAAAGTACAACCTCAGACTAAACTAGTAGTTTTTACTGAGGACCGTCCGA

74 GCGCCCTGCCTCTGCCGTCTGCATAAGACTATGATGCACAAAAATAACAGGCTATAATGG 15
-----+-----+-----+-----+-----+-----+-----+
CGCGGGACGGAGACGGCAGACGTATTCTGATACTACGTGTTTTTATTGTCCGATATTACC

14 CCTGAAAAACGGGA 1
-----+-----+
GGACTTTTTGCCCT

▲
nic site

was distinguished by its low GC content (see section 7.2.4; Fig. 7.5). However, it remains to be determined if such a putative IS element is still active.

Comparisons of the predicted amino acid sequence of *orf2* to sequences present in the databases revealed that Orf2 shares homology (47-50% identity) to a family of hypothetical proteins that have been identified in *E. coli* through the genome sequencing project (Blattner *et al.*, 1997). The functions of these proteins YADD, YFAD, YHGA and YJIP are unknown. One other homologue (55.6% identity) to Orf2 is the plasmid-encoded Orf73 of *Yersinia pestis* plasmid pMT1. Again, the function of this protein of pMT1 is unknown.

The predicted amino acid sequence of *orf1* displayed a weak homology (29% identity, 35% similarity) over 197 amino acid residues of the C-terminal region to a plasmid-encoded restriction methylase gene of IncH plasmid R478 (EMBL: U60283). However, no other similar homologues were found. Finally, the predicted amino acid sequence of *orf4* showed no significant homology to other proteins available in the databases and *orf3* with *E. coli* hypothetical protein YDFB.

The potential functions of all these *orfs* identified within the leading region of ColIb are summarised in Table 7.2.1.

7.2.4 G + C content of the ColIb leading region

The overall G + C content of the 11.7 kb leading region of ColIb was 55.7% (Fig. 7.5), which is higher than the predicted value of 51-52% for the genome that the plasmid was originally isolated in, namely *Shigella sonnei* P9 (Ochman and Lawrence, 1996).

The G + C content drops significantly below the average at base positions 7500-6000 to 45 % (Fig. 7.5). On the basis that the GC contents of closely related organisms tend to be similar, such a finding is indicative of an insertion into the backbone of the leading region during its evolution. Interestingly, this region corresponds to the position of *orf5* and the putative IS element we have identified (Section 7.2.3).

Fig. 7.4. Alingment of the nucleotide sequences of *ssi* elements in the ColIb T-strand and *Frpo*.

Line 1, *ssi3*; line 2, *ssi2*; line 3, *ssi1*; line 4, *Frpo*. Nucleotides that are homologous are boxed. Coordinates in the leading region sequence are *ssi3*, 2-331; *ssi2*, 7036-7361; *ssi1*, 9928-10256. *Frpo* is taken from Masai and Arai (1997). Figure is taken from Bates *et al.*, (1999).

Fig.7.4

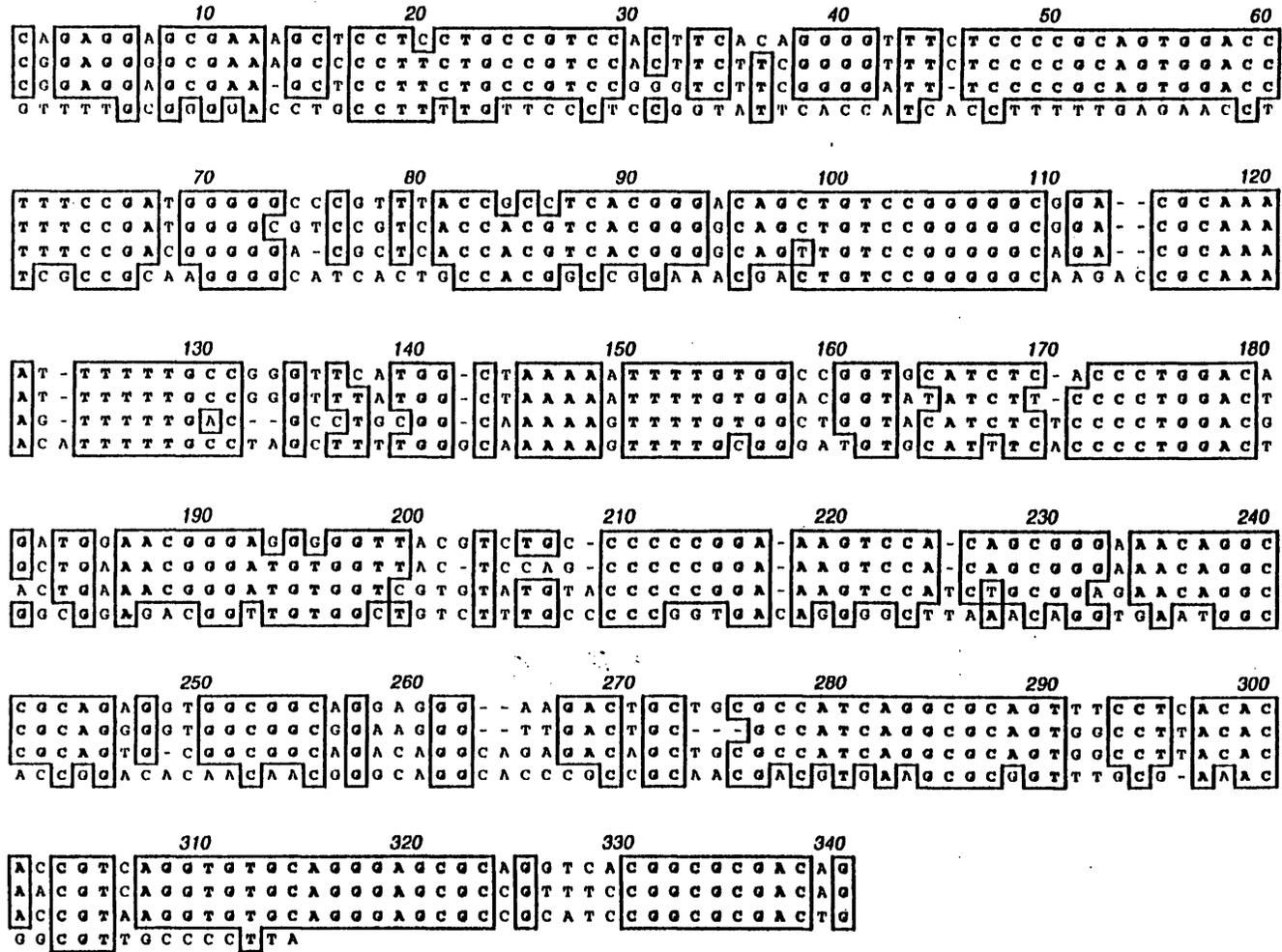
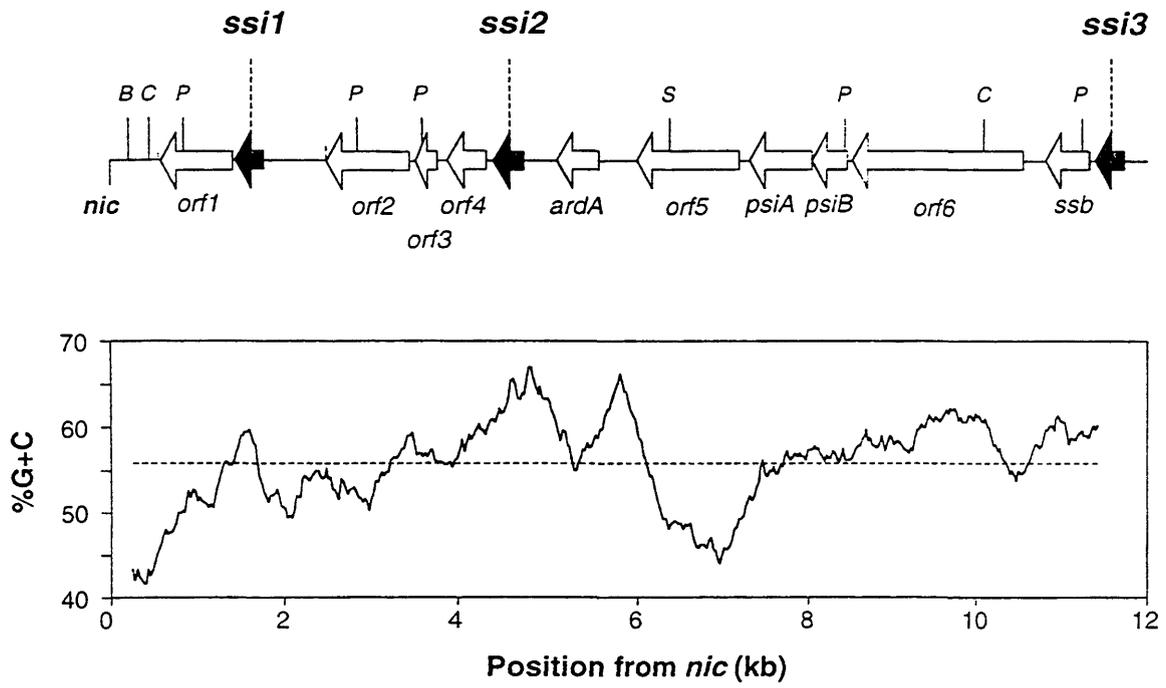


Fig. 7.5. G + C content of the ColIb leading region.

Panel shows G+C% profile of the 11.7 kb of the ColIb leading region. The horizontal line indicates the mean value of 55.7% G+C. The GC profile was calculated using the programme WINDOW (window = 500; shift = 25) followed by STATPLOT. Above the panel is a genetic map of the leading region showing its organisation and position of genetic loci in relation to the G+C% content. Indicated on the map is the *nic* site of the *oriT*, the putative *orfs* identified, the *ssi* sequences and *ssb*, *psiB* and *ardA*. Restriction enzyme cleavage sites are also indicated; B, *BglII*; C, *ClaI*; P, *PstI*; S, *SalI*. Note that at base positions 7500 - 6000 the G+C content drops significantly to 45% which would be indicative of an insertion into the backbone of the leading region during its evolution (*orf5*).

Fig. 7.5



7.2.5 Similarity of the ColIb leading region to other conjugative plasmids

Sequence data are only available for the leading regions of IncP plasmids RP4 and R751 (Pansegrau *et al.*, 1994a; Thorsted *et al.*, 1998) and regions of plasmid F. Through our comparisons, no significant homologies were detected between the leading region of ColIb and the genomes of RP4 and R751.

Extensive nucleotide sequence similarities between ColIb and plasmid F were found extending from the Shine-Dalgarno sequence upstream of *ssb* to 2-bp downstream of *psiA*. Over this region, there is approximately 80% conservation between ColIb and F at the nucleotide sequence level. However, it is important to note that the sequence of the corresponding region of F has not been published in full particularly the region between *ssb* and *psiB*. No homology to F was detected beyond the *psiA* gene. It remains to be determined if any homology exists further upstream of ColIb *ssb* to plasmid F.

The similarities between ColIb and plasmid F has led us to conclude that the region from *ssb* to *psiA* may exist as a conserved cassette-like module, which may also be present on a number of other plasmids. Cross-hybridisation to plasmids from Inc groups FI, FII, FIV, FVI, K, I1, Y, 9 and B shows that *ssb* and *psiB* genes exist on a number of conjugative plasmids all of which display a high degree of identity with each other (Golub *et al.*, 1988). The proposed *ssb-psiA* module is separated from the rest of the leading region of ColIb by a putative IS element (*orf5*).

The only other part of the ColIb leading region that bears similarity with the leading regions of other conjugative plasmids is the *ardA* gene (Section 7.2.3).

7.2.6 Examination of the ColIb leading region for type I restriction enzyme cleavage sites

The leading region of ColIb has been examined for type II restriction enzyme cleavage sites through restriction enzyme analysis of the plasmid

Table 7.2.2 Type I restriction enzyme cleavage sites found within 11.7 kb of the leading region of Collb.

Family	Enzyme	Recognition sequence	Number sites	Position in sequence
K	<i>EcoKI</i>	AACN ₆ GTGC	1	9646-9658
	<i>EcoB</i>	TGAN ₈ TCCT	-	
	<i>EcoD</i>	TTAN ₇ GTCY	-	
	<i>StySB</i>	GAGN ₆ RTAYG	2	2438-2452; 7286-7299
	<i>StySP</i>	AACN ₆ GTRC	1	9646-9658
A	<i>EcoA</i>	GAGN ₇ GTCA	2	2144-2163; 7643-7657
	<i>EcoE</i>	GAGN ₇ GTCA	-	
C	<i>EcoR124</i>	GAAN ₆ RTCG	3	4248-4266; 6461-6473; 10796-10803
	<i>EcoDXXI</i>	TCAN ₇ RTTC	1	7086-8000

Type I restriction enzyme cleavage sites in the leading region of Collb were determined by comparing the sequence data available with the recognition sequence of the enzyme using the GCG program FINDPATTERNS. Restriction enzyme recognition sequences were obtained from REBASE via the World Wide Web (<http://rebase.neb.com/rebase/rebase.html>). N = unspecified base, R = A/G, Y = C/T.

(Rees *et al.*, 1987). Several of these sites are indicated in Fig. 7.3. However, *in vitro* analysis of any DNA for type I cleavage sites is made difficult by the nature of the cleaving reactions for this particular class of enzyme (Chapter One, section 1.5.2) and their limited availability. The most appropriate way to determine the number of type I cleavage sites is to probe any sequence data that is available for the enzymes recognition sequence.

ColIb has been estimated to have approximately 7 *EcoKI* restriction enzyme cleavage sites (B. M. Wilkins, personal communication). From studies with *ardA* and the *ardA*-independent restriction evasion process mediated by ColIb, it is of interest to this project to determine how many of these sites can be found within the leading region of ColIb. Consequently, the sequence data was probed for the *EcoKI* recognition sequence AACNNNNNNGTGC (Wilson and Murray, 1991). Our analyses revealed that only one of the seven *EcoKI* cleavage sites can be found within the sequenced region of the ColIb leading region at base positions 9646-9658 in *orf6* (Table 7.2.2; Fig. 7.3).

Finally, the sequence data of the ColIb leading region was also probed for the cleavage sites of a number of other type I restriction cleavage sites. The enzymes chosen and the results obtained are in Table 7.2.2.

7.2.7 Repeat sequences within the leading region

Three copies of a ~328-bp repeat sequence were identified within the sequence data available for the leading region of ColIb. Each repeat is positioned in the same orientation approximately 60 nucleotides upstream of either *ssb*, *orf4* or *orf1* and the three repeats share between 80 and 85% identity to each other over their entire length. The repeats were labelled *ssi1*, *ssi2* or *ssi3* and are found at base positions 1787-1459, 4678-4353, 11711-11381, respectively (Fig. 7.3). Comparisons of the nucleotide sequence of each *ssi* to those present in the databases identified one homologue belonging to plasmid

F, originally termed *ssiD*, which has now been renamed *Frpo*. Each *ssi* shares ~60% identity over 160-bp of their central portion to the *Frpo* sequence of plasmid F, as shown in Fig. 7.4 (Nomura *et al.*, 1991; Masai and Arai, 1997; EMBL: D90178).

The *Frpo* sequence was originally identified in the leading region of the plasmid F as a single-stranded DNA synthesis initiation sequence, hence the name *ssiD*. Early studies with the single-stranded phage M13 carrying *ssiD* showed that the locus plays an active role in priming DNA synthesis on single-stranded viral DNA (Nomura *et al.*, 1991). It was later found by Masai and Arai (1997) that the repeat locus also had additional properties that allow it to function as a novel type of promoter active in single-stranded form. Consequently, the locus was renamed *Frpo*.

The promoter forms in single-stranded DNA through the formation of secondary structure. The close identity that exists between the *ssi* sequences found in the leading region of *ColIb* and *Frpo* of plasmid F suggests a similar role in *ColIb*. It is envisaged that such sequences may act as single-stranded promoters that form on the transferring plasmid strand and facilitate transient expression of the leading region genes in the recipient before regeneration of a complementary strand through DNA synthesis.

The formation of secondary structure and action of these *ssi* sequences as promoters involved in enhancing expression of leading region genes during conjugative transfer of the plasmid is discussed further in Chapter Eight.

7.3 Summary

Generation and analysis of the 11.7 kb of nucleotide sequence data for the *ColIb* leading region led to the identification of the following features:

* Identification of ten potential protein coding regions, three of which have been previously characterised (*ssb*, *psiB*, *ardA*).

- * Three of the novel *orfs* identified include an F-like *psiA* gene, another which encodes a potential DNA binding protein (*orf6*) and one which is a member of the *E. coli* family of genes known as YHGA (*orf2*).
- * Identification of an unusual IS element and a related Orf (*orf5*), the amino acid sequence of which bears similarity to a number of putative transposase enzymes.
- * All *orfs* identified within the leading region are orientated in the same direction towards *oriT* (Fig. 7.2). Such an observation raises the question of whether the genes present in the leading region of Collb are transcribed as part of an operon.
- * The *orfs* identified from *ssb* to *psiA* of Collb and plasmid F appear to be conserved as a cassette-like module that may be present on a number of other conjugative plasmids.
- * Three of the *orfs* (*ardA*, *orf6* and *orf2*) identified had homology to three *orfs* present on *Yersinia pestis* plasmid pMT1. However, there is no general similarity in the genetic organisation of these two plasmids.

Finally, the most interesting finding was the identification of three repeat sequences (~328 bp) present in the leading region of Collb which appear to be analogous to *Frpo*. It is envisaged that the Collb *ssi* sequences, like *Frpo*, act as single-stranded promoters. Such findings have led us to assume a model explanation as to how expression of the leading region genes on Collb are enhanced during conjugative transfer of the plasmid (Chapter Eight).

Chapter Eight

General Discussion

The fact that ColIb has a narrow replication-maintenance range encompassing genera such as *Escherichia*, *Salmonella*, *Shigella* and *Klebsiella* from the Enterobacteriaceae family raises the question of why the plasmid has evolved two separate strategies for restriction avoidance. However, the finding that strains of *Escherichia coli* encode more than a hundred known R-M systems (<http://rebase.neb.com/rebase/rebase.html>) not only highlights the ubiquity of these systems in nature but suggests that the ability of ColIb to overcome type I and type II restriction barriers must be a significant if not essential aid to transmission. Whether ColIb has recognition sites for all these enzymes and whether the restriction-avoidance mechanisms encoded by ColIb are active against all these R-M systems has yet to be determined.

The resistance of ColIb to both type I and type II restriction enzymes during its transfer by conjugation involves two distinctive mechanisms. One mechanism involves a specialised plasmid-encoded antirestriction gene known as *ardA*, the product of which is active against type I restriction enzymes. Another mechanism, which acts independently of *ardA*, alleviates restriction of ColIb from both type I and type II enzymes in the recipient in second or subsequent rounds of transfer. The main thrust of this thesis was to investigate further these two ColIb restriction-evasion mechanisms. The use of a specially designed genetic system that exploited the sensitivity of IncP β plasmid R751 to restriction was instrumental in allowing a number of observations to be made about these two mechanisms and helped to rule out hypotheses which were made for the *ardA*-independent mechanism. The genetic system relied on measuring ColIb-mediated rescue of R751 from destruction by *EcoKI* (I) and *EcoRI* (II) during conjugation. For simplicity, conclusions and observations made for the ColIb *ardA*-independent

mechanism will be discussed first followed by those made for the *ArdA* system. Other work described in this thesis involved an investigation into the regulatory mechanism governing leading region gene expression during conjugation. The proposed model will be discussed last.

The *ColIb* restriction-avoidance mechanism which was identified using an *ardA* mutant of *ColIb* was originally described by Read *et al.* (1992). It was postulated by Read *et al.* (1992) that amplified transfer of 93-kb *ColIb* might overwhelm the restriction enzymes in the recipient by substrate saturation, regardless of whether they are type I or type II enzymes. Such a scenario should result in alleviation of restriction *in trans*. However, data presented in this thesis are not in agreement with this 'substrate saturation' hypothesis. Furthermore, results suggest that one mechanism is not responsible for alleviating restriction from both type I and type II restriction enzymes as proposed by Read *et al.* (1992). Such a conclusion was drawn from the observation that unlike the anti-type II mechanism, *ColIb*-mediated alleviation of type I restriction occurs independently of any plasmid transfer and appears to be expressed constitutively and *in trans* (Fig. 3.6 B). For clarity, observations made for these two mechanisms will be discussed separately.

The possibility that *ColIb* alleviates type II restriction by overwhelming the enzymes by substrate saturation during amplified transfer was ruled out by data in Chapter Three (Fig. 3.3 A). The key finding was that *ColIb Δ rd* confers no significant protection on restriction-sensitive plasmid R751 when the two plasmids are co-transferred together to a restricting recipient. If amplified transfer of *ColIb* from the donor to the recipient resulted in substrate saturation of restriction enzymes, then restriction should be alleviated *in trans* and the transmission frequency of R751 should approach that observed in non-restricting matings. However, data in Fig. 3.3 (A) clearly shows this not to be the case. It also seems unlikely from these results that amplified transfer of *ColIb Δ rd* changes the physiology of the recipient such that there is a transient

breakdown of restriction. Again, such a scenario should result in alleviation of restriction in the recipient *in trans*.

The finding that ColIb cannot alleviate type II restriction *in trans* suggests that the plasmid must avoid destruction from such enzymes by a *cis*-acting mechanism (see Chapter Six). A number of phages are known to encode anti-restriction mechanisms that are *cis*-acting. One such mechanism involves the random incorporation of unusual bases into the genome of the phage, such that the DNA is not recognised as unmethylated, the signal for cleavage by many restriction enzymes (Krüger and Bickle, 1983; Bickle and Krüger, 1993). These unusual bases are quite often modified versions of either adenine or cytosine residues as these are commonly the substrates for most methyltransferase enzymes within restriction enzyme recognition sequences. Such bases are made at the level of nucleotide metabolism and are incorporated into the genome of the phage only. Consequently, phages demonstrate resistance to restriction; yet, their presence in a cell has no effect on restriction *in trans*. It seems unlikely that ColIb employs a similar mechanism of restriction-avoidance as the plasmid is sensitive to restriction enzyme cleavage *in vitro*. Furthermore, R751 DNA should also be modified when present in the same cell as ColIb by using the same nucleotides.

The possibility that ColIb carries another antirestriction gene(s) cannot be ruled out by any data in this thesis. One idea is that such a gene(s) encodes a non-specific methyltransferase enzyme that randomly methylates the genome of the plasmid. Such a gene must only be expressed during conjugation, as ColIb DNA isolated from vegetatively growing cells is sensitive to the effects of restriction *in vitro* and *in vivo* during transformation (Rees *et al.*, 1987; Read *et al.*, 1992). If the gene is limited to expression during conjugation, it seems likely that the gene should be situated within the leading region of ColIb. Examination of the sequence data generated for the leading region of ColIb (Chapter Seven) identified only one potential *orf* that might encode a methylase. The product of *orf1*, displays weak similarity to a putative

restriction methylase of IncH plasmid R478 (U60283) of *Serratia marcescens*. However, the region of homology (29% identity, 35% similarity over C-terminal 197 amino acids) does not encompass the N-6 adenosine methylase signature typical of DNA methylases. No other *orfs* identified within the complete nucleotide sequence of ColIb were found to have homology to other restriction methylase genes (Fig. 8.1).

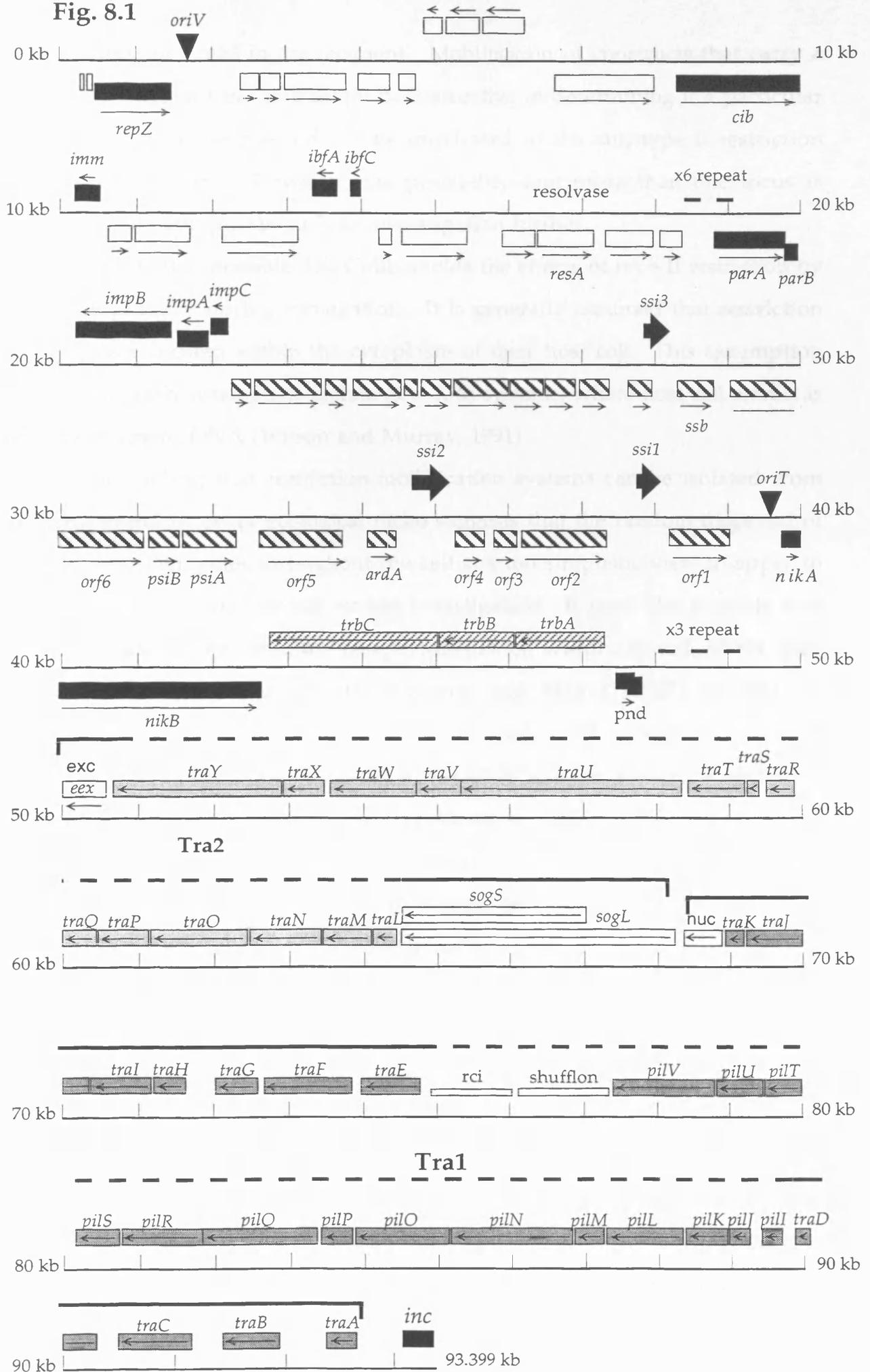
It seems unlikely from these predictions made for Orf1 that the putative protein methylates ColIb during conjugation to protect the plasmid from destruction by type II restriction enzymes. Furthermore, if the enzyme methylates *in trans*, timing of expression of *orf1* and subsequent methylation of ColIb would be an important factor to consider as no cross-protection is conferred on R751 during co-transfer of the two plasmids (see Chapter Three). Further analyses of *orf1* are required before its role in the avoidance of type II restriction enzymes by ColIb can be ruled out. Experiments that might be instructive would involve cloning *orf1* into a vector where it can be expressed constitutively and examining the ability of the resulting construct to rescue either transferring R751 or a restriction sensitive strain of phage λ from the *in vivo* effects of type II restriction enzymes. Such a construct could also be used to determine if *orf1* has a role in protecting ColIb ArdA⁻ from type I restriction enzymes.

Design of an experiment to reveal details about an anti-restriction mechanism that is believed to be *cis*-acting is difficult. Further complications are that such a mechanism appears to only be active during the process of conjugation, as shown by the inability of ColIb to evade type II restriction during the process of transformation (Read *et al.*, 1992). One line of investigation could involve cloning restriction fragments of ColIb and examining them for their ability to confer protection against the effects of *EcoRI* upon the resulting construct. The vector chosen would require *EcoRI* cleavage sites and the ColIb *oriT* locus. Such an analysis could only be utilised if after its mobilisation by the ColIb *nic* mutant (pNA11) the vector is sensitive

Fig. 8.1. Genetic organisation of IncI1 plasmid Collb.

The genetic map was deduced from the complete nucleotide sequence of the plasmid, which was recently made available (Sampei and Mizobuchi; EMBL: AB021078). The genome is represented by the lines marked in kb coordinates. All open reading frames (*orfs*) identified are shown as rectangles. The transcriptional directions of each *orf* is indicated by an arrow and those running from left to right are below the line and those from right to left are above the line. The *orfs* that are part of the Tra1 region of the plasmid are  rectangles, with the exception of the shufflon, *rci* locus and EDTA-resistant nucleasenuc. The *orfs* that form the Tra2 region of the plasmid are  rectangles, with the exception of *sog* and *eex*. The Tra1 and Tra2 regions are indicated on the map, broken lines highlight *orfs* that encode pili; these are the flexible pilus in Tra1 and the rigid pilus in Tra2. Loci with defined functions are indicated by  rectangles. The *orfs* comprising the leading region of the plasmid, some of which were identified during this work, are indicated by  rectangles. Putative *orfs* with no defined function are unlabelled. Locations of *oriT* and *oriV* are indicated. The locations of potential single-stranded promoters (*ssi1*, *ssi2* and *ssi3*) identified during this work are also shown  .

Fig. 8.1



to the effects of *EcoRI* in the recipient. Mobilisation of constructs that carry a particular fragment of *ColIb* might be instructive in determining if a particular gene or locus of the plasmid can be implicated in the anti-type II restriction evasion mechanism. However, the possibility that more than one locus is involved may complicate such an investigation further.

Finally, it is possible that *ColIb* avoids the effects of type II restriction by its mode of entry during conjugation. It is generally assumed that restriction enzymes are located within the cytoplasm of their host cell. This assumption arises from early attempts to isolate new R-M systems, where host cell extract is able to fragment DNA (Wilson and Murray, 1991).

The finding that restriction-modification systems can be isolated from bacteria found in every ecological niche suggests that the random dispersal of all restriction enzymes throughout the cell is a too simplistic view to apply to every known system without further investigation. It might be possible that certain restriction enzymes are compartmentalised within the cell, rather than evenly distributed throughout. Kohring and Mayer (1987) showed by immuno-localisation that between 70-90% of total *EcoRI* restriction enzyme is found in the cell envelope of *E. coli*, whilst 60-70% of the methyltransferase is present within the cytoplasm. Such results are in agreement with the 'cellular defence' hypothesis described in Chapter One, section 1.5.6, where the methylase protects the cells own DNA from destruction and the restriction enzyme part cleaves any foreign DNA entering the cell.

If certain restriction enzymes are compartmentalised within the cell, particularly in the periplasmic space, then such findings by Kohring and Mayer (1987) could explain why some cloned R-M genes in *E. coli* are viable in the r^+m^- configuration. Furthermore, it may be possible that *ColIb* itself avoids the effects of *EcoRI* restriction during conjugation by determining a DNA transport system that ensures transferring DNA does not come into contact with the *EcoRI*-rich periplasmic space. First, this could be the outcome of the plasmid mating-pair formation system, which ensures direct entry of transferring DNA

into the cytoplasm of the recipient. Bacteriophage λ is sensitive to restriction by *EcoRI*. During the initial stages of infection by the phage, the genome is injected as duplex DNA into the periplasmic space before it traverses the inner membrane by a passive sugar uptake mechanism (Dürrenberger *et al.*, 1991). Presumably during transfer of R751, the plasmid enters the periplasmic space.

According to the model proposed by Dürrenberger *et al.* (1991) described in Chapter One, section 1.3.1, plasmid DNA transferring by the process of enterobacterial conjugation enters the periplasm of the recipient cell before traversing the inner membrane by a similar mechanism to that of phage λ . However unlike λ , plasmid DNA at this stage is still in single-stranded form. Restriction enzymes are thought to cleave single-stranded DNA only when canonical duplex structures are formed (Nishigaki *et al.*, 1985; Chapter One, section 1.5.4). Perhaps during ColIb-mediated conjugation conversion of the plasmid into duplex DNA is delayed until entry to the cytoplasm of the recipient. In contrast, perhaps complementary strand synthesis of R751 occurs concurrently during transfer within the periplasmic space, making the plasmid susceptible to cleavage. Such hypotheses are difficult to test. If synthesis of the second strand is delayed until entry to the cytoplasm, this might explain why the plasmid carries a *psiB* gene. Single-stranded DNA entering the cytoplasm will induce the SOS response (Jones *et al.*, 1992). R751 does not carry a similar *psiB* gene and it has yet to be tested if transfer of the plasmid induces the SOS response.

If the findings by Kohring and Mayer (1987) are correct and ColIb does in fact inject its DNA directly into the cytoplasm of the recipient, then mobilisation of any restriction sensitive vector by the ColIb *nic* mutant (pNA11) should avoid the effects of *EcoRI* restriction in the recipient. During work described in this thesis (Chapter Four), pNA11 was often used to mobilise the cloned ColIb *oriT* plasmid (pNA2). Mobilised pNA2 was not sensitive to the effects of *EcoRI* restriction during its transfer (data not shown). However, whether this finding is due to protection by the ColIb transport system or to the

relative insensitivity of pNA2 to *EcoRI* through carriage of only one cleavage site is not known. These experiments need to be repeated using a similar vector carrying enough *EcoRI* cleavage sites to make it a suitable assay of restriction. Unfortunately, time was not available during this work to pursue this investigation further.

The 'substrate saturation' hypothesis as a means of explaining how Collb *ArdA*⁻ avoids the effect of type I restriction enzymes can be ruled out by data presented in this thesis (Chapter Three). The substrate saturation scenario relies on amplified transfer of Collb*drd* from the donor to the restricting recipient. However, Collb*drd* *ArdA*⁻ plasmid is still able to alleviate the effects of type I enzymes when already established in the restricting recipient (Fig. 3.6 **B**).

One hypothesis that has yet to be ruled out is that either transfer of Collb*drd* *ArdA*⁻ or presence of the plasmid in the restricting recipient induces a change in cell physiology, which in turn leads to a breakdown of restriction. Interestingly as early as 1965, Glover and Coulson described experiments involving the breakdown of type I restriction due to changes in cell physiology. Such experiments involved the conjugative transfer of the *EcoKI* R-M genes from an *E. coli* K-12 Hfr donor to an *E. coli* B F⁻ recipient. The restriction phenotype is delayed in the recipient for 15 generations whilst the modification phenotype is observed immediately (Prakash-Cheng and Ryu, 1993). This delayed expression of restriction is due to the activity of ClpX and ClpP, which make up the ClpXP protease of *E. coli*. ClpXP modulates restriction activity by operating post-transcriptionally during assembly of the *EcoKI* complex (see Chapter One, section 1.5.3). ClpX and ClpP form part of the *E. coli* heat shock response that is associated with changes in cell physiology.

It may be possible that either conjugative transfer of Collb *ArdA*⁻ (Fig. 3.4) or just the presence of the plasmid in *E. coli* (Fig. 3.6 **B**) is able to stimulate either ClpXP or another protease, which in turn modulates type I enzyme restriction activity. In contrast, Collb may encode its own protease, although

examination of the complete nucleotide sequence of the plasmid, which is currently available (AB021078), identified no ORFs that have potential protease activity. However, it should be noted that several of the ORFs identified are still without any predicted function (Fig. 8.1).

Unfortunately, the hypothesis that Collb stimulates activity of ClpXP was not tested during this work. It is predicted that if Collb stimulates ClpXP activity, which in turn modulates restriction by *EcoKI*, transfer of the plasmid to *clpX* and *clpP* mutant strains of *E. coli* K-12 should result in a significant reduction in the plasmids transmission, compared to that shown in Fig. 3.2 B. Such a hypothesis should be easily tested by obtaining the appropriate strains described.

Finally, it remains possible that Collb *ArdA* encodes in addition to *ardA* another anti-restriction gene(s) active against type I enzymes and is expressed constitutively. Recent findings by Rastorguev *et al.* (1998) showed that IncI1 plasmid R64 carries a gene that encodes a moderate anti-restriction phenotype active specifically against type I enzymes. The gene identified is not homologous to *ardA*, although the protein does share a similar nine amino acid motif common to *ArdA* and other anti-restriction proteins (see Chapter One, section 1.7). The R64 gene is located adjacent to the plasmids streptomycin and tetracycline resistance genes, downstream from *oriV*. Comparison of the nucleotide sequence of the gene with those present in the databases revealed only one homologue, that of the *arsR* gene of IncFI plasmid R773. *ArsR* is a repressor of the arsenical resistance operon *arsRDABC*. However, R64 carries no such operon as shown by the sensitivity of cells carrying the plasmid to treatment with arsenical chemicals (Rastorguev *et al.*, 1998). Analysis of the complete nucleotide sequence of Collb revealed that the plasmid does not carry a similar gene.

The *arsR* gene of R64 was identified by cloning fragments of the plasmid and examining the resulting recombinants for their ability to confer protection on restriction sensitive phage λ . It would be of interest to this work to apply a

similar strategy to Collb, with the aim of identifying a region of the plasmid that is responsible for constitutive alleviation of restriction by type I enzymes.

Zygotic induction of Collb leading region genes, *psiB* and *ssb*, has been demonstrated (Jones *et al.*, 1992; Chapter One, section 1.6.4). Only recently, evidence has been obtained to show that transcription of *ardA* is enhanced in the recipient cell during the process of conjugation. The technique used was a competitive RT-PCR assay, which is able to quantify mRNA transcript levels of a particular gene. A full description of the details relating to this technique is available in Althorpe *et al.* (1999). During conjugation, *ardA* mRNA transcripts were measured and compared to those from vegetatively growing cells. Transcript levels of *ardA* increased in a short pulse during conjugation, which began at about five-six minutes. Transcript levels later diminished as transcription is turned down and then existing transcripts were degraded. Comparisons of these results with the time of entry of Collb*drd* to the recipient cell, nine minutes, suggest that transcription of *ardA* begins before completion of the first round of plasmid transfer (Althorpe *et al.*, 1999).

The burst of *ardA* transcription during conjugation in comparison to that detected in vegetatively growing cells suggests that like *psiB* and *ssb*, *ardA* is also subject to zygotic induction. This view is further supported by data presented in this thesis (Chapter Three and Chapter Five), which has also been published in conjunction with the transcriptional analysis of the *ardA* gene in Althorpe *et al.* (1999). In experiments involving the use of the R751 restriction assay, ArdA-mediated restriction alleviation was observed to be optimal when encoded by a transferring Collb*drd* plasmid rather than an equivalent plasmid resident in the restricting recipient (Fig. 3.5; Fig. 3.6 C). Such results imply that expression of *ardA* is activated during conjugation.

Transcriptional activation of *ardA* in the recipient was indicated by the strong protection conferred by a co-transferring Collb*drd* plasmid (Fig. 3.5 A) but lack of even basal protection by the Collb*drd nic* mutant (Chapter Four and Chapter Five, Fig. 5.3 A) immobilised in the donor. Thus, it is inferred that

neither *ardA* mRNA or ArdA protein are transported from the donor to recipient cells. R751 is protected to an intermediate extent by a co-transferring *Collbdrd* ArdA mutant (Fig. 3.4 A), but this effect is attributed to the *ardA*-independent restriction-avoidance response described already in this chapter.

Some interesting observations about the timing of expression of *ardA* can be made from the studies conducted on *ardA* transcription during conjugation as measured by the competitive RT-PCR assay (Althorpe *et al.*, 1999). Transcription of *ardA*, in the recipient, extends over 10 minutes or more in the cell population. Transfer of *Collb* from the donor to the recipient is not synchronised and transcription in a single pair of conjugating cells is likely to be less than 10 minutes. From measuring and analysing the time of entry of the plasmid, the onset of *ardA* transcription can be estimated. The rate of DNA transfer in Hfr matings is estimated to be 45-kb min⁻¹ at 37°C (Bachmann *et al.*, 1976). Assuming transfer of *Collb* occurs at the same rate, transfer of the 100-kb plasmid (pLG221) should be completed within 2.5 minutes. Completion of the first round of transfer in a bacterial population is marked by a surge of transconjugant production. In the case of *Collb*, this surge occurs after nine minutes. The level of *ardA* transcripts increase in abundance after five-six minutes, which suggests that transcription of this gene is initiated within a minute of entry of the plasmid. Such timing indicates that transcription occurs shortly after entry of the leading region to the recipient cell.

One intriguing question is what is the regulatory mechanism employed by *Collb* to allow enhanced expression of leading region genes, *psiB*, *ssb* and *ardA*, in a transient burst shortly after conjugative transfer of the plasmid. It is known that plasmid DNA is transferred to the recipient cell in single-stranded form, yet, it seems likely from the observations made above that these genes are expressed before transfer of the plasmid has finished. By sequencing the leading region of *Collb*, the aim of Chapter Seven was to use such data not

only to map genetic loci, but to look for regulatory sequences that might give an indication of the mechanism governing leading region gene expression.

For all plasmids examined, *ardA* and *psiB-ssb* genes are orientated in their leading regions such that the DNA strand that is thought to be transferred to the recipient cell during conjugation also corresponds to the template strand for transcription of these genes (Fig. 7.5). A number of putative *orfs* were identified from the sequence data generated, the predicted functions of which are described in Chapter Seven, section 7.2.3. Interestingly, all the *orfs* identified were orientated in the same transcriptional orientation as *ardA*, *psiB* and *ssb* (Fig. 7.5).

Perhaps the most interesting finding was the identification of three repeat sequences of ~328-bp in length, which were named *ssi1*, *ssi2* and *ssi3*. The repeats share between 80-85% sequence identity with each other over their entire length. Remarkably, each of these repeats aligns with only one other sequence in the databases, a similar sequence found within the leading region of plasmid F, originally called *ssiD*, now renamed *Frpo* (Fig. 7.4).

The F locus was originally isolated by Nomura *et al.* (1991) and shown to be a single-stranded initiation sequence capable of supporting synthesis of primers for DNA replication, hence the name *ssiD*. Later work carried out by Masai and Arai (1997) showed that the *ssiD* locus has additional properties which allow it to function as a novel type of promoter active only in single-stranded form. Consequently, the locus was renamed *Frpo*. The elaborate stem-loop structure proposed by Masai and Arai (1997) extends over ~300 nucleotides in the single strand and includes a -35 and -10 sequence and a transcription start site (Fig. 8.3). Masai and Arai (1997) found that *Frpo* is only active as a promoter in single-stranded DNA. However, dsDNA can be activated for transcription from such promoters by denaturation. The ColIb *ssi* sequences have predicted secondary structure similar to that of *Frpo* (Bates *et al.*, 1999; Fig. 8.2). The putative -10 regions of the ColIb *ssi* sequences contain numerous mis-matches, which may aid promoter function by facilitating

Fig. 8.2. Comparison of the proposed secondary structure for *Frp* with the sequences of the *ColIb ssi* sites.

Figure is taken from Bates *et al.* (1999). The *Frp* structure is from Masai and Arai (1997) with the exception that opposing G and T residues are paired. The arrow indicates leftward extension from the 5' end of the RNA. *Frp** is a thermodynamically favoured variant of the sequence. *ssi* structures were calculated using the *mfold* 3.0 RNA folding program (Mathews *et al.*, 1999), which was obtained via the World Wide Web (<http://www.ibc.wustl.edu/zuker/>). Upper case letters indicate nucleotides that conform with the consensus sequences of the -35 and -10 regions (5'-TTGACA-3' and 5'-TATAAT-3'). Figure was drawn by B. M. Wilkins.

Fig. 8.2.

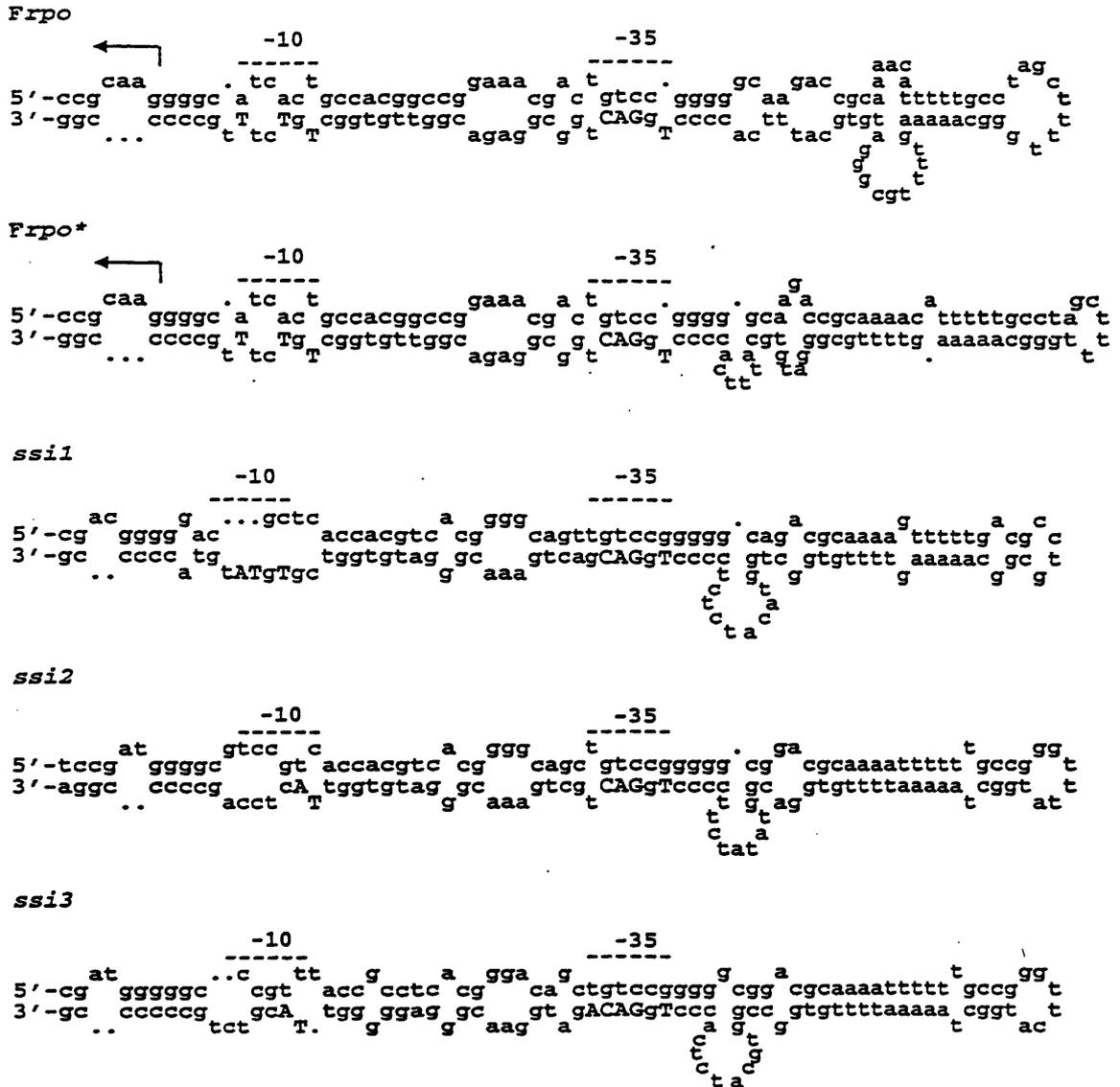


Fig. 8.3. Schematic diagram depicting the proposed secondary structure of the *Frp* locus.

The figure is based on the proposed structure formed in single-stranded DNA at the *Frp* locus of plasmid F (Masai and Arai, 1997). The structure acts as a novel type of promoter capable of directing transcription by RNA polymerase from ssDNA templates. Sites of interest include the -35, -10 and transcription start site. The stem loop structure cannot form in dsDNA as it would be an imperfect duplex. The loops, due to mispairing are seen to be essential for function. Figure is derived from that in Masai and Arai (1997) and was drawn by Chris Cane.

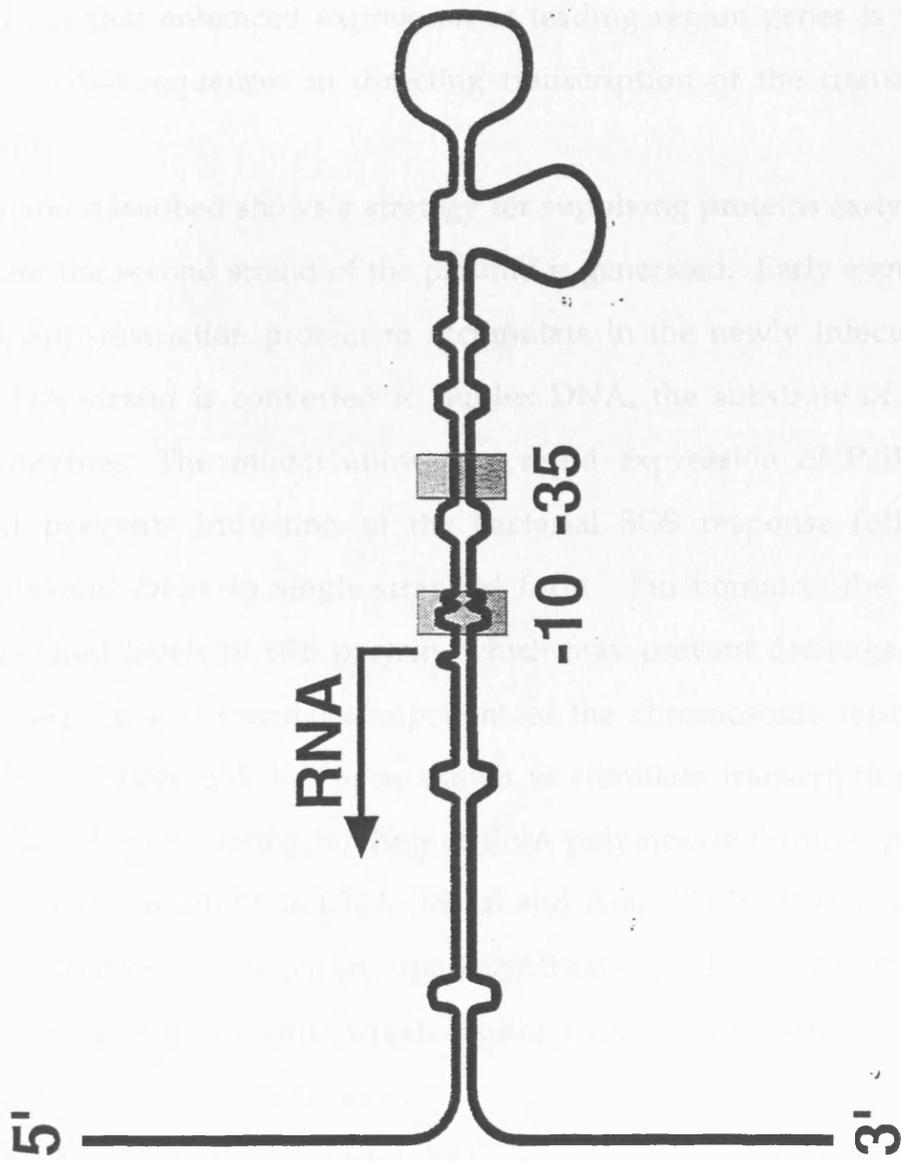


Fig. 8.3.

melting of the duplex structure and formation of the open complex necessary for transcription (Bates *et al.*, 1999).

Importantly, *Frpo* and the *Collb ssi* sequences are orientated in their respective leading regions such that the promoter structure will form in the DNA strand that is transferred to the recipient. This feature leads to the appealing model that enhanced expression of leading region genes is due to activity of *Frpo*-like sequences in directing transcription of the transferring plasmid strand.

The model described shows a strategy for supplying proteins early in the recipient before the second strand of the plasmid is generated. Early expression allows ArdA anti-restriction protein to accumulate in the newly infected cell before the DNA strand is converted to duplex DNA, the substrate of type I restriction enzymes. The model allows for rapid expression of PsiB as a function that prevents induction of the bacterial SOS response following transfer of plasmid DNA in single-stranded form. Furthermore, the model allows for elevated levels of SSB protein, which may prevent drainage of the cellular counterpart, an essential component of the chromosome replication apparatus. In addition, SSB has been shown to stimulate transcription from *Frpo*, presumably by facilitating binding of RNA polymerase through protein-protein interactions on template DNA (Masai and Arai, 1997). It is envisaged that these promoters are silenced upon synthesis of the complementary plasmid strand in the recipient, which would explain how leading region genes are expressed in a transient burst.

Design of an experiment to test if the *Collb ssi* sequences are able to act as single-stranded promoters is possible. A *Collb* plasmid carrying a promoterless *lacZ* operon fusion within the *ssb* gene has already been created in this laboratory (Jones *et al.*, 1992). The plasmid would provide a useful source of a DNA fragment carrying the *ssb::lacZ* fusion and the upstream *ssi3* sequence for cloning into a single-stranded vector based on M13. Confirmation that *ssi3* acts as a promoter would be achieved if elevated levels of β -galactosidase activity

were detected in cells carrying the single-stranded DNA construct. As a useful control, a similar vector carrying the Collb fragment inserted in the opposite orientation should result in no elevated levels of β -galactosidase. Alternatively, the Ω fragment described in Chapter Four, could be cloned upstream of *ardA* in a construct carrying only the leading region (see Fig. 7.5). It is predicted that if zygotic induction of *ardA* is governed by *ssi3* found upstream of the gene, then the presence of Ω will disrupt transcription and prevent enhanced expression of *ardA*. Consequently, mobilisation of the construct would result in no ArdA-mediated alleviation of type I restriction in the recipient cell. However, this experiment does not confirm that *ssi3* is the promoter.

In conclusion, despite complications encountered, a number of hypotheses for the Collb anti-type I and anti-type II restriction mechanisms have been tested and rejected by this work. Unfortunately, it was beyond the scope of this thesis to determine the molecular basis of these two plasmid evasion processes, although a number of new hypotheses have been generated which are described in this chapter. In addition, data in this thesis were not only useful in providing supporting evidence that expression of *ardA* is enhanced by mating (Althorpe *et al.*, 1999) but has played a major role in the generation of a model that explains how leading region genes might be zygotically induced in *E. coli* during conjugation (Bates *et al.*, 1999). If the Collb *ssi* regions are confirmed as being promoters active in single-stranded DNA form, then it would be interesting to see if they are also active in strains other than *E. coli*.

Appendix I

Determination of the R_{KI} and R_{RI} phenotypes of *E. coli* strains using λvir

<i>E. coli</i> host	λvir	pfu ml ⁻¹	e.o.p
NM816N	$\lambda vir.0$	5.0×10^8	1.0
GI65N	"	1.7×10^6	0.003
GI65N(pLG221)	"	1.2×10^8	0.24
GI65N(pLG292)	"	5.0×10^7	0.10
GI65N(pNA11)	"	3.5×10^8	0.70
GI65N(pNA12)	"	8.0×10^7	0.16
GI65N(pNA6)	"	5.3×10^8	1.06
GI65(pLG290)	"	4.9×10^8	0.98
GI65N	$\lambda vir.K$	1.0×10^{12}	1.0
GI65N(NTP14)	"	1.0×10^9	0.001
GI65N(pNA16)	"	2.1×10^{12}	2.1
GI65N(pNA17)	"	1.0×10^9	0.001

100 μ l of *E. coli* cells (10^7 ml⁻¹) were mixed with 3 ml of SNA and overlaid onto a BLA plate. 10 μ l of each diluted bacteriophage suspension was spotted onto the surface of the plate and allowed to adsorb. The plates were then incubated overnight at 37°C. The number of plaques formed were examined and the plaque forming units (pfu) per ml and efficiency of plating (e.o.p) were calculated. An e.o.p value of 1.0 is equivalent to 100% plating efficiency. λvir is a virulent strain of λ , where $\lambda vir.0$ carries no methylation, $\lambda vir.K$ carries *EcoKI* methylation.

Literature Cited

Achtman, M., Morelli, G., and Schwuchow, S. (1978a) Cell-cell interactions in conjugating *Escherichia coli*: role of F pili and fate of mating aggregates. *J Bacteriol* 135: 1053-1061.

Achtman, M., Schwuchow, S., Helmuth, R., Morelli, G., and Manning, P. A. (1978b) Cell-cell interactions in conjugating *Escherichia coli*: Con⁻ mutants and stablization of mating aggregates. *Mol Gen Genet* 164: 171-183.

Alfoldi, L., Jacob, F., and Wollman, E. L. (1957) Zygoose letale dans les croisements entre souches colicinogenes et non colicinogenes. *Acad Sci* 244: 2974-2976.

Althorpe, N. J., Chillely, P. M., Thomas, A. T., Brammar, W. J., and Wilkins, B. M. (1999) Transient transcriptional activation of the IncI1 plasmid anti-restriction gene (*ardA*) and SOS inhibition gene (*psiB*) early in conjugating recipient bacteria. *Mol Microbiol* 31: 133-142.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Amábile-Cuevas, C. F., and Chicurel, M. E. (1992) Bacterial plasmids and gene flux. *Cell* 70: 189-199.

Anthony, K. G., Sherburne, C., Sherburne, R., and Frost, L. S. (1994) The role of the pilus in recipient cell recognition during bacterial conjugation mediated by F-like plasmids. *Mol Microbiol* 13: 939-953.

Arber, W., and Dussoix, D. (1962) Host specificity of DNA produced by *Escherichia coli* I. Host controlled modification of bacteriophage λ . *J Mol Biol* 5: 18-36.

Backmann, B. J. (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* 36: 525-557.

Backmann, B. J., Low, K. B., and Taylor, A. L. (1976) Recalibrated linkage map of *Escherichia coli* K-12. *Bacteriol Rev* 40: 116-167.

Bagdasarian, M., Bailone, M. M., Manning, P. A., Lurz, R., Timmis, K. M., and Devoret, R. (1986) An inhibitor of SOS induction, specified by a plasmid locus in *Escherichia coli*. *Proc Natl Acad Sci USA* 83: 5723-5726.

Bagdasarian, M., Bailone, A., Angulo, J. F., Scholz, P., Bagdasarian, M., and Devoret, R. (1992) PsiB, an anti-SOS protein, is transiently expressed by the F sex factor during its transmission to an *Escherichia coli* K-12 recipient. *Mol Microbiol* 6: 885-893.

Bailone, A., Backman, A., Sommer, S., Celerier, J., Bagdasarian, M. M., Bagdasarian, M., and Devoret, R. (1988) PsiB polypeptide prevents activation of RecA protein in *Escherichia coli*. *Mol Gen Genet* 214: 389-395.

Balzer, D., Pansegrau, W., and Lanka, E. (1994) Essential motifs of relaxase (TraI) and TraG proteins involved in conjugative transfer of plasmid RP4. *J Bacteriol* 176: 4285-4295.

Bancroft, I., and Wolk, C. P. (1989) Characterisation of an insertion sequence (IS891) of novel structure from the Cyanobacterium *Anabaena* sp. Strain M-131. *J Bacteriol* 171: 5949-5954.

- Barcus, V. A., and Murray, N. E.** (1995) Barriers to recombination: restriction. In *Population Genetics of Bacteria*. Baumberg, S., Young, J. P. W., Wellington, E. M. H. and Saunders, J. R. (eds). Cambridge: Cambridge University Press, pp. 31-58.
- Bates, S., Cashmore, A. M., and Wilkins, B. M.** (1998) IncP plasmids are usually effective in mediating conjugation of *Escherichia coli* and *Saccharomyces cerevisiae*: Involvement of the Tra2 mating system. *J Bacteriol* **180**: 6538-6543.
- Bates, S., Roscoe, R. R., Althorpe, N. J., Brammar, W. J., and Wilkins, B. M.** (1999) Expression of leading region genes on IncI1 plasmid Collb-P9: Genetic evidence for single-stranded DNA transcription. *Microbiology* **145**: 2655-2662.
- Belogurov, A. A., Delver, E. P., and Rodzevich, O. V.** (1992) IncN plasmid pKM101 and IncI1 plasmid Collb-P9 encode homologous antirestriction proteins in their leading regions. *J Bacteriol* **174**: 5079-5085.
- Belogurov, A. A., Delver, E. P., and Rodzevich, O. V.** (1993) Plasmid pKM101 encodes two non-homologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J Bacteriol* **175**: 4843-4850.
- Belogurov, A. A., and Delver, E. P.** (1995) A motif conserved among the type I restriction-modification enzymes and antirestriction proteins: a possible basis for mechanism of action of plasmid-encoded antirestriction functions. *Nucleic Acids Res* **23**: 785-787.
- Bickle, T. A., Brack, C., and Yuan, R.** (1978) ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc Natl Acad Sci USA* **75**: 3099-3103.
- Bickle, T. A., and Krüger, D. H.** (1993) Biology of DNA restriction. *Microbiol Rev* **57**: 434-450.
- Billington, S. J., Sinistaj, M., Cheetham, B. F., Ayres, A., Moses, E. K., Katz, M. E., and Rood, J. I.** (1996) Identification of a native *Dichelobacter nodosus* plasmid and implications for the evolution of the *vap* region. *Gene* **172**: 111-116.
- Birmbiom, H. C., and Doly, J.** (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl Acids Res* **7**: 1513-1523.
- Bischofberger, N., Ng, P. G., Webb, T. R., and Matteucci, M. D.** (1987) Cleavage of single-stranded oligonucleotides by *EcoRI* restriction endonuclease. *Nucl Acids Res* **15**: 709-716.
- Blattner, F. R., Plunkett, G., Bloch, C. A., Ferna, N. T., Burland, V., Riley, M., Colladovides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y.** (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1470.
- Boulnois, G. J., and Wilkins, B. M.** (1978) A ColI-specified product, synthesised in newly infected recipients, limits the amount of DNA transferred during conjugation of *Escherichia coli* K-12. *J Bacteriol* **133**: 1-9.
- Boyd, C. A., Archer, J. A. K., and Sherratt, D. J.** (1989) Characterisation of the ColE1 mobilization region and its protein products. *Mol Gen Genet* **217**: 488-498.
- Bradley, D. E.** (1980) Morphological and serological relationships of conjugative pili. *Plasmid* **4**: 155- 176.
- Bradley, D. E.** (1983) Derepressed plasmids of incompatibility group I₁ determine two different morphological forms of pilus. *Plasmid* **9**: 331-334.

Bradley, D. E. (1984) Characteristics and function of thick and thin conjugative pili determined by transfer derepressed plasmid of incompatibility groups I₁, I₂, I₅, B, K and Z. *J Gen Microbiol* **130**: 1489-1502.

Bukhari, A. I., Shapiro, J. A., and Adhya, S. L. (1977) Bacterial plasmids. In *DNA insertion elements, and episomes*. New York: Cold Springer Harbor, pp.621-624.

Burckhardt, J., Weisemann, J., and Yuan, R. (1981a) Characterisation of the DNA methylase activity of the restriction enzyme from *Escherichia coli* K*. *J Biol Chem* **256**: 4024-4032.

Burckhardt, J., Weisemann, J., Hamilton, D. L., and Yuan, R. (1981b) Complexes formed between the restriction endonuclease *EcoK* and heteroduplex DNA. *J Mol Biol* **153**: 425-440.

Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rapuoli, R., and Covacci, A. (1996) *Cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* **93**: 14648-14653.

Chabbert, Y. A., Roussel, A., Witchitz, J. L., San-Son-Le Pors, M. J., and Courvalin, P. (1979) Restriction endonuclease generated patterns of plasmids belonging to incompatibility groups II, C, M and N: Application to plasmid taxonomy and epidemiology. In *Plasmids of Medical, Environmental and Commercial importance*. Timmis, K. N., and Pühler, A. (eds). North Holland / Amsterdam / New York: Elsevier, pp. 183-193.

Chandrasegaran, S., and Smith, H. O. (1988) Amino acid sequence homologies among twenty-five restriction endonucleases and methylases. In *Structure and Expression. From proteins to ribosomes*. Sarma, R. H., and Sarma, M. H. (eds). New York: Adenine press Inc, pp.149-156.

Chase, J. W., Merrill, B. M., and Williams, K. R. (1983) F-sex factor encodes a single-stranded DNA binding-protein (SSB) with extensive sequence homology to *Escherichia coli* SSB. *Proc Natl Acad Sci USA* **80**: 5480-5484.

Chatfield, L. K., Orr, E., Boulnois, G. J., and Wilkins, B. M. (1982) DNA primase of plasmid ColIb is involved in conjugal DNA synthesis in donor and recipient bacteria. *J Bacteriol* **152**: 1188-1195.

Chilley, P. M., and Wilkins, B. M. (1995) Distribution of the *ardA* family of antirestriction genes on conjugative plasmids. *Microbiology* **141**: 2157-2164.

Clewell, D. B., Flannagan, S. E., and Jaworski, D. D. (1995) Unconstrained bacterial promiscuity - the Tn916 - Tn1545 family of conjugative transposons. *Trends Microbiol* **3**: 229-236.

Cohen, A., Fisher, W. D., Curtiss III, R., and Adler, H. (1968) DNA isolated from *Escherichia coli* minicells mated with F⁺ cells. *Proc Natl Acad Sci USA* **61**: 61-68.

Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1972) Non-chromosomal antibiotic resistance in bacterial genetic transformation of *Escherichia coli* R-factor DNA. *Proc Natl Acad Sci USA* **69**: 2110-2114.

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

Cowan, G. M., Gann, A. A. F., and Murray, N. E. (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56**: 103-109.

- Cram, D., Ray, A., O'Goram, L., and Skurray, R. (1984) Transcriptional analysis of leading region in F plasmid transfer. *Plasmid* **11**: 221-233.
- Curtiss, R., Caro, L. G., Allison, D. P., and Stallions, D. R. (1969) Early stages of conjugation in *Escherichia coli*. *J Bacteriol* **100**: 1091-1104.
- Daniel, A. S., Fuller-Pace, F. V., Legge, D. M., and Murray, N. E. (1988) Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. *J Bacteriol* **170**: 1775-1782.
- Datta, N., and Hughes, V. (1983) Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. *Nature* **306**: 616-617.
- Davis, E. J., and Henry, J. (1982) Conjugal transfer replication of R64*drd11* plasmid DNA in the donor cells of *Escherichia coli* K-12. *Mol Gen Genet* **187**: 305-309.
- Delver, E. P., Kotova, V. U., Zavilgelsky, G. B., and Belogurov, A. A. (1991) Nucleotide sequence of the gene (*ard*) encoding the antirestriction protein of plasmid Collb-P9. *J Bacteriol* **173**: 5887-5892.
- Di Lorenzo, L., Frost, L. S., Finlay, B. B., and Paranchych, W. (1991) Characterisation of the *oriT* region of the IncFV plasmid pED208. *Mol Microbiol* **5**: 780-785.
- Dixon, D. A., and Kowalczykowski, S. C. (1993) The recombination hotspot χ is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* **73**: 87-96.
- Dreiseikelmann, B. (1994) Translocation of DNA across bacterial membranes. *Microbiol Rev* **58**: 293-316.
- Dryden, D. T. F., Cooper, L. P., Thorpe, P. H., and Byron, O. (1997) The *in vitro* assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry* **5**: 1065-1076.
- Dürrenberger, M. B., Villiger, W., and Baechli, T. (1991) Conjugational junctions: morphology of specific contacts in conjugating *Escherichia coli* bacteria. *J Struct Biol* **107**: 146-156.
- Dutreix, M., Backman, A., Celerier, J., Bagdasarian, M. M., Sommer, S., Bailone, A., Devoret, R., and Bagdasarian, M. (1988) Identification of *psiB* genes of plasmids F and R6-5. Molecular basis for *psiB* enhanced expression in plasmid R6-5. *Nucl Acids Res* **16**: 10669-10679.
- Dybvig, K., and Yu, H. (1994) Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*. *Mol Microbiol* **12**: 547-560.
- Endlich, B., and Linn, S. (1985a) The DNA restriction endonuclease of *Escherichia coli* B. I. Studies of the DNA translocation and the ATPase activities. *J Biol Chem* **260**: 5720-5728.
- Endlich, B., and Linn, S. (1985b) The DNA restriction endonuclease of *Escherichia coli* B. II. Further studies of the structure of DNA intermediates and products. *J Biol Chem* **260**: 5729-5738.
- Falkow, S., Tompkins, L. S., Silver, R. P., Guerry, P., and Le Blanc, D. J. (1971) The replication of R-factor DNA in *Escherichia coli* K-12. *Ann NY Acad Sci* **182**: 153-171.
- Falkow, S., Hedges, R. W., Guerry, R., and Datta, N. (1974) Polynucleotide sequence relationships among plasmids of the I compatibility complex. *J Gen Microbiol* **85**: 65-76.

- Finlay, B. B., and Paranchych, W. (1986) Nucleotide sequence of the surface exclusion genes *traS* and *traT* from the *Inf_olac* plasmid pED208. *J Bacteriol* **166**: 713-721.
- Finnegan, J., and Sherratt, D. (1982) Plasmid ColE1 conjugal mobility: The nature of *bom*, a region required *in cis* for transfer. *Mol Gen Genet* **185**: 344-351.
- Firth, N., Ippen-Ihler, K., and Skurray, R. A. (1996) Structure and function of the F factor and mechanism of conjugation. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. 2nd edn. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 2377-2401.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J-F., Dougherty, B. A., Merrick, J. M., McKenny, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J. O., Scott, J., Shirley, R., Liu, L. I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, T., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Sandek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Grehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.
- Frédéricq, P. (1965) Genetics of colicinogenic factors. *Zentbl Bakt Parasit Kde* **196**: 142-151.
- Frost, L. S., Ippen-Ihler, K., and Skurray, R. A. (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**: 162-210.
- Fürste, J. P., Pansegrau, W., Ziegelin, G., Kroeger, M., and Lanka, E. (1989) Conjugative transfer of promiscuous IncP plasmids: interactions of plasmid-encoded products with the transfer origin. *Proc Natl Acad Sci USA* **86**: 1771-1775.
- Furuya, N., and Komano, T. (1991) Determination of the nick site at *oriT* of IncI1 plasmid R64: global similarity of *oriT* structures of IncI1 and IncP plasmids. *J Bacteriol* **173**: 6612-6617.
- Furuya, N., Nisioka, T., and Komano, T. (1991) Nucleotide sequence and functions of the *oriT* operon in IncI1 plasmid R64. *J Bacteriol* **173**: 2231-2237.
- Furuya, N., and Komano, T. (1994) Surface exclusion gene of IncI1 plasmid R64: Nucleotide sequence and analysis of deletion mutants. *Plasmid* **32**: 80-84.
- Furuya, N., and Komano, T. (1995) Specific binding of the NikA protein to one arm of 17-base-pair inverted repeat sequences within the *oriT* region of plasmid R64. *J Bacteriol* **177**: 46-51.
- Furuya, N., and Komano, T. (1997) Mutational analysis of the R64 *oriT* region: Requirement for precise location of the NikA-binding sequence. *J Bacteriol* **179**: 7291-7297.
- Glover, S. W., and Colson, C. (1965) The breakdown of the restriction mechanism in zygotes of *Escherichia coli*. *Genet Res* **6**: 153-155.
- Golub, E. I., and Low, B. K. (1985) Conjugative plasmids of enteric bacteria from many different incompatibility groups have similar genes for single-stranded DNA-binding proteins. *J Bacteriol* **162**: 235-241.
- Golub, E. I., and Low, B. K. (1986) Derepression of single-stranded DNA-binding protein genes on plasmids derepressed for conjugation, and complementation of an *E. coli* *ssb'* mutation by these genes. *Mol Gen Genet* **204**: 410-416.

- Golub, E. I., Bailone, A., and Devoret, R. (1988) A gene encoding an SOS inhibitor is present in different conjugative plasmids. *J Bacteriol* **170**: 4392-4394.
- Gottesman, S. (1996) Proteases and their targets in *Escherichia coli*. *Ann Rev Genet* **30**: 465-506.
- Gottesman, S., and Maurizi, M. R. (1992) Regulation by proteolysis: Energy-dependent proteases and their targets. *Microbiol Rev* **56**: 592-621.
- Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E., and Rosenberg, J. M. (1981) Sequence analysis of the DNA encoding the *EcoRI* endonuclease methylase. *J Biol Chem* **256**: 2143-2153.
- Gribskov, M., Devereux, J., and Burgess, R. R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nucl Acids Res* **12**: 539-549.
- Gulig, P. A., Caldwell, A. L., and Chiodo, V. A. (1992) Identification, genetic analysis and DNA sequence of a 7.8 kb virulence region of the *Salmonella typhimurium* virulence plasmid. *Mol Microbiol* **6**: 1395-1411.
- Haase, J., Lurz, R., Grahn, A. M., Banford, D. H., and Lanka, E. (1995) Bacterial conjugation mediated by plasmid RP4 - RSF1010 mobilization, donor specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *J Bacteriol* **177**: 4779-4791.
- Hadi, S. M., Bickle, T. A., and Yuan, R. (1975) The role of S-adenosylmethionine in the cleavage of deoxyribonucleic acid by the restriction endonuclease from *Escherichia coli* K. *J Biol Chem* **250**: 4159-4164.
- Ham, L. M., Cram, D., and Skurray, R. (1989) Transcriptional analysis of the F plasmid surface exclusion region: mapping of *traS*, *traT* and *traD* transcripts. *Plasmid* **21**: 1-8.
- Hama, C., Takizawa, T., Moriwaki, H., Urasaki, Y., and Mizobuchi, K. (1990) Organisation of the replication control region of plasmid ColIb-P9. *J Bacteriol* **172**: 1983-1991.
- Hanahan, D. (1983) Studies of transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557-580.
- Harrington, L. C., and Rogerson, A. C. (1990) The F pilus of *E. coli* appears to support stable DNA transfer in the absence of wall-wall contact between cells. *J Bacteriol* **172**: 7263-7264.
- Hartskeerl, R. A., Bergmans, J. E. N., Kamp, M. C., and Hoekstra, W. P. M. (1983) Cloning of an exclusion-determining fragment of the IncI plasmid, R144. *Plasmid* **10**: 11-20.
- Hartskeerl, R. A., and Hoekstra, W. P. M. (1984) Exclusion in IncI1-type *Escherichia coli* conjugations: the stage of conjugation at which exclusion operates. *Antonie van Leeuwenhoek, J Microbiol* **50**: 113-124.
- Hartskeerl, R., Tommassen, J., and Hoekstra, W. P. M. (1985a) Relationship between the proteins encoded by the exclusion determining locus of the IncI plasmid R144 and the cellular localization of these proteins in *Escherichia coli* K-12. *Mol Gen Genet* **200**: 138-144.
- Hartskeerl, R., Zuidweg, E., Geffen, M., and Hoekstra, W. P. M. (1985b) The IncI1 plasmids R144, R64 and ColIb belong to one exclusion group. *J Gen Microbiol* **131**: 1305-1311.

Hartskeerl, R. A., and Hoekstra, W. P. M. (1985) Exclusion by the IncI plasmid R144 determined by measuring DNA transfer in *Escherichia coli* conjugations. *J Gen Microbiol* 131: 2263-2267.

Hartskeerl, R., Overduin, P., Hoekstra, W., and Tommassen, J. (1986) Nucleotide sequence of the exclusion-determining locus of IncI plasmid R144. *Gene* 42: 107-111.

Havekes, L., Tommassen, J., Hoekstra, W. P. M., and Lugtenberg, B. (1977) Isolation and characterisation of *Escherichia coli* K-12 F⁻ mutants defective in conjugation with an I-type donor. *J Bacteriol* 129: 1-8.

Hayes, W. (1968) Temperate bacteriophages and lysogeny. In *The Genetics of Bacteria and their Viruses: Studies in Basic Genetics and Molecular Biology*. (2nd ed). Oxford: Blackwell, pp. 447-479.

Heitman, J., Treisman, J., Davis, N. G., and Russel, M. (1989) Cassettes of the fl intergenic region. *Nucl Acids Res* 17: 4413.

Hénaut, A., and Danchin, A. (1996) Analysis and predictions from *Escherichia coli* sequences, or *E. coli* In silico. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. 2nd edn. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 2047-2066

Horiuchi, T., and Komano, T. (1998) Mutational analysis of plasmid R64 thin pilus prepilin: the entire prepilin sequence is required for processing by type IV prepilin peptidase. *J Bacteriol* 180: 4613-4620.

Howland, C. J., and Wilkins, B. M. (1988) Direction of conjugative transfer of IncI1 plasmid Collb-P9. *J Bacteriol* 170: 4958-4959.

Howland, C. J., Rees, E. E. D., Barth, P. T., and Wilkins, B. M. (1989) The *ssb* gene of plasmid Collb-P9. *J Bacteriol* 171: 2466-2473.

Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A., Brubaker, R. R., and Garcia, E. (1998) Structural organisation of virulence-associated plasmids of *Yersinia pestis*. *J Bacteriol* 180: 5192-5202.

Jones, A. L. (1991) The leading region of IncI1 plasmid Collb-P9. Ph.D. thesis, University of Leicester, Leicester UK.

Jones, A. L., Barth, P. T., and Wilkins, B. M. (1992) Zygotic induction of plasmid *ssb* and *psiB* genes following conjugative transfer of IncI1 plasmid Collb-P9. *Mol Microbiol* 6: 605-613.

Kelleher, J. E., Daniel, A. S., and Murray, N. E. (1991) Mutations that confer *de novo* activity upon a maintenance methyltransferase. *J Mol Biol* 221: 431-440.

Kohring, G-W., and Mayer, F. (1987) In situ distribution of EcoRI methylase and restriction endonuclease in cells of *Escherichia coli* Bs 5. *FEBS Letters* 216: 207-210.

Kim, S-R., and Komano, T. (1997) The plasmid R64 thin pilus identified as a type IV pilus. *J Bacteriol* 179: 3594-3603.

Kobayashi, I., Naito, T., Handa, N., and Kusano, K. (1995) Selfish genes – restriction-modification systems force cells to maintain them. *J Cell Biochem* S19A: 110.

Komano, T., Toyoshima, A., Morita, K., and Nisioka, T. (1988) Cloning and nucleotide sequence of the *oriT* region of the IncI1 plasmid R64. *J Bacteriol* 170: 4385-4387.

- Komano, T., Funayama, N., Kim, S. R., and Nisioka, T. (1990) Transfer region of IncI1 plasmid R64 and role of shufflon in R64 transfer. *J Bacteriol* **172**: 2230-2235.
- Komano, T., Kim, S. R., Yoshid, T., and Nisioka, T. (1994) DNA rearrangements of the shufflon determines recipient specificity in liquid matings of IncI1 plasmid R64. *J Mol Biol* **243**: 6-9.
- Korona, R., Korona, B., and Levin, B. R. (1993) Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *J Gen Microbiol* **139**: 1283-1290.
- Korona, R., and Levin, B. R. (1993) Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution* **47**: 556-575.
- Kotova, V. Y., Zavilgelskil, G. B., and Belogurov, A. A. (1988) Weakening of type I restriction in the presence of plasmids of the IncI group - General characteristics and cloning of *ard* gene. *J Mol Biol* **22**: 227-232.
- Krause, M., Roudier, C., Fierer, J., Harwood, J., and Guiney, D. (1991) Molecular analysis of the virulence locus of the *Salmonella dublin* plasmid pSDL2. *Mol Microbiol* **5**: 307-316.
- Krüger, D. H., and Bickle, T. A. (1983) Bacteriophage survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their host. *Microbiol Rev* **47**: 345-360.
- Kulakauskas, S., Lubys, A., and Ehrlich, S. D. (1995) DNA restriction-modification systems mediate plasmid maintenance. *J Bacteriol* **177**: 3451-3454.
- Kusano, K., Naito, T., Handa, N., and Kobayashi, I. (1995) Restriction-modification systems as genomic parasites in competition for specific sequences. *Proc Natl Acad Sci USA* **92**: 11095-11099.
- Lanka, E., and Barth, P. T. (1981) Plasmid RP4 specifies a deoxyribonucleic acid primase involved in its conjugal transfer and maintenance. *J Bacteriol* **148**: 769-781.
- Lanka, E., and Wilkins, B. M. (1995) DNA processing reactions in bacterial conjugation. *Ann Rev Biochem* **64**: 141-169.
- Lederberg, J., and Tatum, E. L. (1946) Gene recombination in *E. coli*. *Nature* **158**: 558.
- Lessl, M., Balzer, D., Weyrauch, K., and Lanka, E. (1993) The mating-pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J Bacteriol* **175**: 6415-6425.
- Loh, S., Cram, D., and Skurray, R. (1989) Nucleotide sequence of the leading region adjacent to the origin of transfer on plasmid F and its conservation among conjugative plasmids. *Mol Gen Genet* **219**: 177-186.
- Lunnen, K. D., Barsomian, J. M., Camp, R. R., Card, C. O., Chen, S. Z., Croft, R., Looney, M. C., Meda, M. M., Moran, L. S., Nwankwo, D. O., Slatko, B. E., Van Cott, E. M., and Wilson, G. G. (1988) Cloning type II restriction and modification genes. *Gene* **74**: 25-32.
- Luria, S. E., and Suit, J. L. (1987) Colicins and Col plasmids. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Ingraham, J., Low, K. B., Magasanik, B., Schaechter, M., Umberger, H. E., and Neidhart, F. C. (eds). Washington: ASM Press, pp. 1615-1624.

- Makovets, S., Titheradge, A. J. B., and Murray, N. E.** (1998) ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol Microbiol* **28**: 25-35.
- Manning, P. A., Morelli, G., and Achtman, M.** (1981) *TraG* protein of the F sex factor of *Escherichia coli* K-12 and its role in conjugation. *Proc Natl Acad Sci USA* **78**: 7487-7491.
- Marsh, J. L., Erfle, M., and Wykes, E. J.** (1984) The pIC plasmid and phage vectors with versatile sites for recombinant selection by insertional inactivation. *Gene* **32**: 481-485.
- Masai, H., and Arai, K.** (1997) *FrpO*: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication. *Cell* **89**: 897-907.
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H.** (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911-940.
- Mathews, C. K., and van Holde, K. E.** (1990) In *Biochemistry*. Redwood City: The Benjamin / Cummings Publishing Company.
- Matic, I., Taddei, F., and Radman, M.** (1996) Genetic barriers among bacteria. *Trends Microbiol* **4**: 69-73.
- Matson, S. W., and Morton, B. S.** (1991) *Escherichia coli* DNA helicase I catalyses a site and strand-specific nicking reaction at the F plasmid *oriT*. *J Biol Chem* **266**: 109-114.
- Mazodier, P., and Davis, J.** (1991) Gene transfer between distantly related bacteria. *Annu Rev Genet* **25**: 147-171.
- McClarín, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J., and Rosenberg, J. M.** (1986) The structure of the DNA-*EcoRI* endonuclease complex at 3Å resolution. *Science* **234**: 1526-1541.
- Merryweather, A., Rees, C. E. D., Smith, N. M., and Wilkins, B. M.** (1986) Role of *sog* polypeptides specified by plasmid ColIb-P9 and their transfer between conjugating bacteria. *EMBO J* **5**: 3007-3012.
- Meyer, R., and Laine, P. S.** (1990) The single-stranded DNA binding protein of *Escherichia coli*. *Microbiol Rev* **54**: 342-380.
- Meynell, E., and Datta, N.** (1967) Mutant drug resistant factors of high transmissibility. *Nature* **259**: 455-458.
- Milkman, R.** (1997) Recombination and population structure in *Escherichia coli*. *Genetics* **146**: 745-750.
- Minkley, E. G., and Willetts, N. S.** (1984) Overproduction, purification and characterisation of the F *traT* protein. *Mol Gen Genet* **196**: 225-235.
- Murai, N., Kamata, H., Nagashima, Y., Yagisawa, H., and Hirata, H.** (1995) A novel insertion sequence (IS)-like element of the thermophilic bacterium PS3 promotes expression of the alanine carrier protein-encoding gene. *Gene* **163**: 103-107.
- Naito, T., Kusano, K., and Kobayashi, I.** (1995) Selfish behaviour of restriction-modification systems. *Science* **267**: 897-899.
- Newman, A. K., Rubin, R. A., Kim, S. H., and Modrich, P.** (1981) DNA sequences of structural genes for *EcoRI* DNA restriction and modification enzymes. *J Biol Chem* **256**: 2131-2139.

Nishigaki, K., Kaneko, Y., Wakuda, H., Husimi, Y., and Tanaka, T. (1985) Type II restriction endonucleases cleave single-stranded DNAs in general. *Nucl Acids Res* 13: 5747-5760.

Nomura, N., Masai, H., Inuzuka, M., Miyazaki, C., Ohtsubo, E., Itoh, T., Sasamoto, S., Matsui, M., Ishizaki, R., and Arai, K. (1991) Identification of eleven single-stranded initiation sequences (*ssi*) for priming of DNA replication in the F, R6K, R100 and ColE2 plasmids. *Gene* 108: 15-22.

Ochman, H., and Lawrence, J. G. (1996) Phylogenetics and the amelioration of bacterial genomes. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. 2nd edn. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 2627-2637.

O'Neill, M., Chen, A., and Murray, N. E. (1997) The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: They do not resist loss and are readily replaced by alleles conferring different specificities. *Proc Natl Acad Sci USA* 94: 14596-14601.

Ou, J. T., and Anderson, T. F. (1970) Role of pili in bacterial conjugation. *J Bacteriol* 102: 648-654.

Panicker, M. M., and Minkley, E. G. Jr. (1985) DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. *J Bacteriol* 162: 584-590.

Pansegrau, W., Ziegelin, G., and Lanka, E. (1990a) Covalent association of the *traI* gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation Nick site. *J Biol Chem* 265: 10637-10644.

Pansegrau, W., Balzer, D., Krufft, V., Lurz, R., and Lanka, E. (1990b) *In vitro* assembly of relaxosomes at the transfer origin of plasmid RP4. *Proc Natl Acad Sci USA* 87: 6555-6559.

Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., Helinski, D. R., Schwab, H., Stanisich, V. A., and Thomas, C. M. (1994a) Complete nucleotide sequence of Birmingham IncP α plasmids: compilation and comparative analysis of sequence data. *J Mol Biol* 239: 623-663.

Pansegrau, W., Schroder, W., and Lanka, E. (1994b) Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (TraI) of conjugative plasmid RP4. *J Biol Chem* 269: 2782-2789.

Pansegrau, W., and Lanka, E. (1991) Common sequence motifs in DNA relaxases and nick regions from a variety of DNA transfer systems. *Nucl Acids Res* 19: 3455.

Pansegrau, W., and Lanka, E. (1996) Mechanism of initiation and termination reactions in conjugative DNA processing - Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncP alpha plasmid RP4. *J Biol Chem* 271: 13068-13076.

Pearson, W. R., and Lipman, D. J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85: 2444-2448.

Prakash-Cheng, A., Chung, S. S., and Ryu, J. (1993) The expression and regulation of *hsd_k* genes after conjugal transfer. *Mol Gen Genet* 241: 491-496.

Prakash-Cheng, A., and Ryu, J. (1993) Delayed expression of *in vivo* restriction activity following conjugal transfer of *Escherichia coli* *hsd_K* (restriction-modification) genes. *J Bacteriol* 175: 4905-4906.

- Prentki, P., and Krisch, H. M.** (1984) *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **19**: 303-313
- Rashtchian, A., Crooks, J. H., and Levy, S. B.** (1983) *TraJ* independence in expression of *traT* on F. *J Bacteriol* **154**: 1009-1012.
- Rastorguev, S. M., Zavilgelsky, G. B., and Tchurikov, N. A.** (1998) IncI1 plasmid R64 encodes the ArsR protein that alleviates type I restriction. *FEBS Letters* **426**: 21-23.
- Read, T. D.** (1993) Evasion of restriction systems by plasmid ColIb-P9 during bacterial conjugation. Ph.D thesis, University of Leicester, Leicester UK.
- Read, T. D., Thomas, A. T., and Wilkins, B. M.** (1992) Evasion of type I and type II restriction systems by IncI1 plasmid ColIb-P9 during transfer by bacterial conjugation. *Mol Microbiol* **6**: 1933-1941.
- Redaschi, N., and Bickle, T. A.** (1996) DNA restriction and modification systems. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. 2nd edn. Neidhardt, F. C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 773-781.
- Rees, C. E. D.** (1986) A genetic analysis of the transfer genes of the IncI1 plasmid ColIb-P9. Ph.D thesis, University of Leicester, Leicester UK.
- Rees, C. E. D., Bradley, D. E., and Wilkins, B. M.** (1987) Organisation and regulation of the conjugation genes of IncI1 plasmid ColIb-P9. *Plasmid* **18**: 223-236.
- Rees, C. E. D., and Wilkins, B. M.** (1989) Transfer of *tra* proteins into the recipient cell during bacterial conjugation mediated by plasmid ColIb-P9. *J Bacteriol* **171**: 3152-3157.
- Rees, C. E. D., and Wilkins, B. M.** (1990) Protein transfer into the recipient cell during bacterial conjugation: studies with F and RP4. *Mol Microbiol* **4**: 1199-1205.
- Roberts, R. J., and Macelis, D.** (1994) REBASE – restriction enzymes and methylases. *Nucl Acids Res* **22**: 3628-3639.
- Roongta, V. A., Jones, C. R., and Gorenstein, D. G.** (1990) Effect of distortions in the deoxyribose phosphate backbone conformation of duplex oligodeoxyribo-nucleotide dodecamers containing GT, GG, GA, AC and GU base-pair mismatches on P-31 NMR spectra. *Biochemistry* **29**: 5245-5258.
- Roscoe, R. A.** (1996) Regulation of leading region genes on IncI1 plasmid ColIb-P9. Ph.D. thesis. University of Leicester, Leicester, UK.
- Russell, C. B., Thaler, D. S., and Dahlquist, F. W.** (1989) Chromosomal transformation of *Escherichia coli recD* strains with linearised plasmids. *J Bacteriol* **171**: 2609- 2613.
- Saadi, S., Maas, W. K., Hill, D. F., and Bergquist, P. L.** (1987) Nucleotide sequence analysis of RepFIC, a basic replicon present in IncFI plasmids P307 and F, and its relation to the RepA replicon of IncFII plasmids. *J Bacteriol* **169**: 1836-1846.
- Saiki, R. K.** (1989) The design and optimisation of the PCR. In *PCR Technology: Principles and Applications for DNA Amplification*. Erlich, H. A. (ed). New York / London / Tokyo / Melbourne / Hong Kong: M Stockton Press, pp. 7-16.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. M.** (1989) *Molecular Cloning: A Laboratory Manual*- 2nd edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

- Schandel, K. A., Muller, M. M., and Webster, R. E. (1992) Localization of TraC, a protein involved in assembly of the F conjugative pilus. *J Bacteriol* **174**: 3800-3806.
- Shen, P., and Huang, H. V. (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441-457.
- Sia, E. A., Kuehner, D. M., and Figurski, D. H. (1996) Mechanism of retrotransfer in conjugation – prior transfer of the plasmid is required. *J Bacteriol* **178**: 1457-1464.
- Silverman, P. M. (1987) The structural basis of prokaryotic DNA transfer. In *Bacterial outer membranes as model systems*. Inouye, M. (ed). New York: John Wiley and Sons, pp. 277-309.
- Simonsen, L. (1990) Dynamics of plasmid transfer on surfaces. *J Gen Microbiol* **136**: 1001-1007.
- Skurray, R. A., and Reeves, P. (1973) Characterisation of lethal zygotis associated with conjugation in *Escherichia coli* K-12. *J Bacteriol* **113**: 58-70.
- Smith, H. R., Humphreys, G. O., Wilshaw, G. A., and Anderson, E. S. (1976) Characterisation of plasmids coding for the restriction endonuclease EcoRI. *Mol Gen Genet* **143**: 319-325.
- Soberon, X., Covarrubias, L., and Bolivar, F. (1980) Construction and characterisation of new cloning vehicles .IV. Deletion derivatives of pBR322 and pBR325. *Gene* **9**: 287-305.
- Stahl, F. W., Kobayashi, I., Thaler, D., and Stahl, M. M. (1986) Direction of travel of RecBC recombinase through bacteriophage lambda DNA. *Genetics* **113**: 215-227.
- Studier, F. W., and Bandyopadhyay, F. W. (1988) Model for how type I enzymes select cleavage sites in DNA. *Proc Natl Acad Sci USA* **85**: 4677-4681.
- Summers, D. K. (1996) Plasmid dissemination. In *The Biology of Plasmids*. Summers, D. K. (ed). Oxford: Blackwell Science Ltd, pp. 92-110.
- Suri, B., and Bickle, T. A. (1985) EcoA: the first member of a new family of type I restriction modification systems – gene organisation and enzymatic activities. *J Mol Biol* **186**: 77-85.
- Szybalski, W., Kim, S. C., Hasan, N., and Podhajska, A. J. (1991) Class-IIS restriction enzymes – a review. *Gene* **100**: 13-26.
- Tao, T., Patel, J., Firman, K., and Kneale, G. (1991) A family of regulatory genes associated with type II restriction-modification systems. *J Bacteriol* **173**: 1367-1375.
- Tardif, G., and Grant, R. B. (1983) Transfer of plasmids from *Escherichia coli* to *Pseudomonas aeruginosa* - characterisation of a *Pseudomonas aeruginosa* mutant with enhanced recipient ability for enterobacterial plasmids. *Antimicrobiol Agents Chemo* **24**: 201-208.
- Thorsted, P. B., MacCartney, D. P., Akhtar, P., Haines, A. S., Ali, N., Davidson, P., Stafford, T., Pocklington, M. J., Pansegrau, W., Wilkins, B. M., Lanka, E., and Thomas, C. M. (1998) Complete sequence of the IncP β plasmid R751: implications for evolution and organisation of the IncP backbone. *J Mol Biol* **282**: 969-990.
- Titheradge, A. J. B., Ternent, D., and Murray, N. E. (1996) A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol Microbiol* **22**: 437-447.

- Traxler, B. A., and Minkley, E. G. Jr.** (1987) Revised genetic map of the distal end of the F transfer operon: implications for DNA helicase I, nicking at *oriT*, and conjugal DNA transport. *J Bacteriol* **169**: 3251-3259.
- Uemura, H., and Mizobuchi, K.** (1982) Genetic and physical characterisation of the ColIb plasmid using ColIb-R222 hybrids. *Mol Gen Genet* **185**: 1-12.
- Vapnek, D., and Rupp, W. D.** (1971) Identification of individual sex-factor DNA strands and their replication during conjugation in thermosensitive DNA mutants of *Escherichia coli*. *J Mol Biol* **60**: 413-424.
- Viera, J., and Messing, J.** (1982) The pUC plasmids, and M13mp&-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259-268.
- Walia, S. K., and Duckworth, D. H.** (1986) The relationship of the delta factor and KcolIb to the ColIb plasmid. *J Gen Microbiol* **132**: 3261-3268.
- Walker, G.** (1987) The SOS response of *Escherichia coli*. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. 2nd edn. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 1346-1357.
- Wang, J. Y., and Syvanen, M.** (1992) DNA twist as a transcriptional sensor for environmental changes. *Mol Microbiol* **6**: 1861-1866.
- Waters, V. L., and Guiney, D. G.** (1993) Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. *Mol Microbiol* **9**: 1123-1130.
- Wickner, S., and Maurizi, M. R.** (1999) Here's the hook: similar substrate binding sites in the chaperone domains of Clp and Lon. *Proc Natl Acad Sci USA* **96**: 8318-8320.
- Willetts, N. S., and Skurray, R. A.** (1987) Structure and function of the F factor and mechanism of conjugation. In *Escherichia coli and Salmonella typhimurium*. Neidhardt, F.C. (ed.-in-chief). Washington, DC: American Society for Microbiology Press, pp. 692-696.
- Willetts, N., and Wilkins, B. M.** (1984) Processing of plasmid DNA during bacterial conjugation. *Microbiol Rev* **48**: 24-41.
- Williams, K. R., Spicer, E. K. A. L. R., and Guggenheimer, R. A.** (1983) Limited proteolysis studies on the *Escherichia coli* single-stranded binding protein. *J Biol Chem* **258**: 3346-3355.
- Wilkins, B. M.** (1975) Partial suppression of the phenotype of *Escherichia coli* K-12 *dnaG* mutants by some I-like conjugative plasmids. *J Bacteriol* **122**: 899-904.
- Wilkins, B. M.** (1995) Gene transfer by bacterial conjugation: diversity of systems and functional specializations. In *Population Genetics of Bacteria*. Baumberg, S., Young, J. P. W., Saunders, S. R., and Wellington, E. M. H. (eds) Cambridge: Cambridge University Press, pp. 59-88.
- Wilkins, B. M., and Bates, S.** (1997) Gene transfer by bacterial conjugation: Establishment of the immigrant plasmid in the recipient cell. In *Molecular Microbiology*. NATO ASI series Vol. H 103. Busby, S. J. W., Thomas, C. M., and Brown, N. L. (eds). Berlin Heidelberg: Springer-Verlag, pp. 43-64.
- Wilkins, B. M., and Lanka, E.** (1993) DNA processing and replication during plasmid transfer between Gram-negative bacteria. In *Bacterial Conjugation*. Clewell, D.B. (ed). New York: Plenum Press, pp. 105-136.

- Wilkins, B. M., Chilly, P. M., Thomas, A. T., and Pocklington, M. J. (1996) Distribution of restriction enzyme recognition sequences on broad host range plasmid RP4: molecular and evolutionary implications. *J Mol Biol* **258**: 447-456.
- Wilson, G. G. (1991) Organisation of restriction-modification systems. *Nucl Acids Res* **19**: 2539-2566.
- Wilson, G. G., and Murray, N. E. (1991) Restriction and modification systems. *Ann Rev Genet* **25**: 585-627.
- Winans, S. C., and Walker, G. C. (1983) Genetic localisation and characterisation of a pKM101-coded endonuclease. *J Bacteriol* **154**: 1117-1125.
- Winans, S. C., Elledge, S. J., Krueger, J. H., and Walker, G. C. (1985) Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J Bacteriol* **161**: 1219-1221.
- Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.
- Yuan, R., and Hamilton, D. L. (1984) Type I and type III restriction-modification enzymes. In *DNA methylation*. Razin, A., Cedar, H., and Riggs, A. D. (eds). New York: Springer-Verlag, pp. 11-37.
- Zechner, E. L., de la Cruz, F., Eisenbrandt, R., Grahn, A. M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C. M., Wilkins, B. M., and Zatyka, M. (1999) Conjugative DNA transfer processes. In *The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread*. Thomas, C. M. (ed). Amsterdam: Harwood Academic Publishers, Chapter 3.