The ancestral IncP replication system consisted of contiguous *oriV* and *trfA* segments as deduced from a comparison of the nucleotide sequences of diverse IncP plasmids

CHRISTOPHER A. SMITH,^{1†} MICHAEL PINKNEY,^{1‡} DONALD G. GUINEY² and CHRISTOPHER M. THOMAS^{1*}

¹School of Biological Sciences, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK

²Department of Medicine, University of California San Diego Medical Center, 225 Dickinson Street, San Diego, CA 92103, USA

(Received 25 January 1993; revised 5 April 1993; accepted 13 April 1993)

In most plasmids which have been studied to date the functions required for plasmid replication are clustered in a 2–3 kb region. However, in all known naturally occurring plasmids of the *Escherichia coli* incompatibility group P the essential replication functions, *oriV*, the vegetative replication origin and *trfA*, which encodes proteins essential to activate *oriV*, are separated by blocks of DNA consisting of either known genes conferring resistance to antimicrobial agents and/or putative transposable elements. Nucleotide sequence comparisons reported here reveal that these blocks of DNA have inserted at different points into a backbone of DNA common to IncP plasmids. The results indicate that in the common ancestor of present IncP plasmids *oriV* and *trfA* must have been contiguous, whilst a ρ -independent transcriptional terminator, now lost in IncP α plasmids, may have prevented *trfA* operon transcription from interfering with the activity of *oriV*.

Introduction

Dissection of naturally occurring conjugative plasmids such as members of the IncF family has generally shown that both replication and transfer genes can be isolated in contiguous blocks of related functions (Womble & Rownd, 1988). The basic replication apparatus of most plasmids can be isolated in a 2–3 kb segment (Couturier *et al.*, 1988). In contrast, the basic replication functions of plasmids belonging to the *Escherichia coli* incom-

patibility group P cannot be isolated on a small contiguous segment (Thomas *et al.*, 1980). This is partly because the replication functions are part of a large set of coordinately regulated genes which are potentially host-lethal if not accompanied by plasmid-encoded repressor functions (Figurski *et al.*, 1982; Thomas & Smith, 1987). However, it is also due to the fact that in all naturally occurring IncP plasmids so far studied, the basic replication functions, *oriV*, the vegetative replication origin and *trfA*, which encodes proteins essential to activate *oriV*, are separated by blocks of DNA which encode known phenotypic markers and/or putative transposable elements (Villaroel *et al.*, 1983; Smith & Thomas, 1987, 1989). It is therefore unclear whether there was ever an ancestral clustered IncP replication system analogous to that found in other plasmids. To investigate this we have analysed the DNA sequences between *oriV* and *trfA* for a number of different IncP plasmids.

Naturally occurring IncP plasmids have been divided into two subgroups: IncP α , consisting of the majority of known IncP plasmids; and IncP β , consisting of R751, R906, R772 and pJP4 (Yakobson & Guiney, 1983; Chikami *et al.*, 1985; Lanka *et al.*, 1985; Smith &

* Author for correspondence. Tel. (021) 414 5903; fax (021) 414 5925.

† Present address: Department of Anatomy, University of Birmingham Medical School, PO Box 363, Birmingham B15 2TT, UK.

‡ Present address: Northumberland Biologicals Ltd, South Nelson Industrial Estate, Cramlington, Northumberland NE23 9HL, UK.

Abbreviations: Ap, ampicillin; Hg, mercuric ion; Km, kanamycin; Pn, penicillin; Sm, streptomycin; Su, sulphonamide; Tp, trimethoprim.

The nucleotide sequence data reported in this paper have been submitted to GenBank and assigned the accession numbers L13264 (R751 *trfA*–Tn4321), X01751 (R751 Tn4321–*oriV*) and L13265 (R906 *trfA*–*oriV*).

Thomas, 1987). Heteroduplex analysis (Villaroel *et al.*, 1983) and restriction mapping (Currier & Morgan, 1981) have shown that all the IncP α plasmids differ primarily only by the insertion/deletion of transposable elements. Since all IncP α plasmids carry tetracycline resistance it seems likely that they all have *oriV* and *trfA* separated by the *tet* genes as found in the best-studied IncP α plasmids RK2/RP4/RP1/R68/R18 (Thomas & Smith, 1987). In contrast, R751 (Tp^R) (Jobanputra & Datta, 1974) and R906 (Pn^RSm^RSu^RHg^R) (Hedges *et al.*, 1977) do not confer tetracycline resistance and originally we hypothesized that they might carry contiguous *oriV* and *trfA* functions. Whilst this is not the case, *oriV* and *trfA*, still being widely spaced in both plasmids (Smith & Thomas, 1987), the putative transposable elements which separate them are different, suggesting the possibility that they arose by different insertion events at different sites. We have therefore determined and compared DNA sequences in this region of RK2, R751 and R906 to see if it is possible to deduce the structure of an ancestral IncP replication system. The general conclusions of the results described here have been referred to previously in a review (Smith & Thomas, 1989).

Methods

Bacterial strains, growth conditions and plasmids. *Escherichia coli* K12 strains MV10Nal^R (*thr leu thi tonB galK trpE5 nalA*) (from D. R. Helinski, Department of Biology, UCSD, La Jolla, CA 92093, USA) or JM83 (Vieira & Messing, 1982) were the bacterial hosts. L Broth (Kahn *et al.*, 1979) and L agar (L Broth solidified with 1.5%, w/v, agar) were used, supplemented where appropriate with antibiotics: benzyl penicillin at 300 µg ml⁻¹ in solid and 100 µg ml⁻¹ in liquid media for Pn^R and Ap^R; kanamycin sulphate at 50 µg ml⁻¹ for Km^R; trimethoprim at 100 µg ml⁻¹ for Tp^R; streptomycin sulphate at 30 µg ml⁻¹ for Sm^R. Previously described plasmids were R751 (Jobanputra & Datta, 1974), R906 (Hedges *et al.*, 1977), pCAS212 (Smith & Thomas, 1985) and pVII28.1 (Shingler & Thomas, 1984). pVII28.1 has an *EcoRI* site, introduced by transposon mutagenesis, 345 bp downstream of the *trfA* stop codon (Smith & Thomas, 1984) and was thus helpful for obtaining sequence downstream of *trfA*. Plasmids not described previously are as follows. pCAS182 consists of the *SphI* to *SalI* fragment of R751 from coordinate 16.9 to 14.45 kb (Fig. 1*b*), inserted between the *SphI* and *SalI* sites of pUC18 (Norrander *et al.*, 1983). pCAS184 is the *XhoI* to *SalI* fragment of R751 from coordinate 14.55 to 14.45 (Fig. 1*b*), inserted into the *SalI* site of pUC18. pCAS270 is the *PstI* to *HindIII* fragment of R751 from coordinate 14.5 to 13.0 (Fig. 1*b*), inserted between the *PstI* and *HindIII* sites of pUC18. pCAS258 and pCAS256 are *TaqI* fragments of R906, respectively comprising base pairs 107 to 297, 297 to 600 bp of the sequence in Fig. 3, inserted into the *Clal* site of pBR322 (Bolivar *et al.*, 1977). pCAS285 is the *SalI* fragment of R906 from coordinate 11.0 to 10.0 kb (Fig. 1*a*), inserted into the *SalI* site of pUC18. pCAS290 is the *PstI* to *SphI* fragment of R906 from coordinate 10.05 to 6.7 kb (Fig. 1*a*; Smith & Thomas, 1989), inserted between the *PstI* and *SphI* sites of pUC18. Plasmids pMIK2.1, 3.1, 5.1, 6.1 and 8.1 have a structure identical to pVII28.1 (Shingler & Thomas, 1984) except that the location of the *EcoRI* site introduced by transposon mutagenesis with Tn1723, which contains *EcoRI* sites 15 bp from each end, differs, being 905, 590, 1070,

400 and 140 bp downstream of the *trfA* gene, respectively. These plasmids had been isolated by V. Shingler at the same time as those described by Shingler & Thomas (1984) but were not characterized. pCT800 consists of the *BglII* to *XhoI* fragment of R751 (coordinates 12.0 to 14.7 kb, Fig. 1*b*) inserted between the *BamHI* and *SalI* sites of pBR322.

DNA manipulations and sequencing. Plasmid DNA was isolated by the method of Birnboim & Doly (1979) for small-scale preparation and by a modified version of this method for large-scale preparation (Smith & Thomas, 1983). Manipulation and analysis of DNA was essentially as described by Maniatis *et al.* (1982). DNA sequencing was carried out by the method of Maxam & Gilbert (1980) with minor modifications (Smith & Thomas, 1984) and by the chain termination method of Sanger *et al.* (1977) using a Sequenase 2.0 kit from USB. Computer manipulations of the DNA sequence were carried out using the University of Wisconsin GCG package (Devereux *et al.*, 1984).

Results and Discussion

Relationship between trfA and oriV in IncPβ plasmids R751 and R906

Our Southern blotting previously indicated that *oriV* and *trfA* are not adjacent in R751 and R906 (Smith & Thomas, 1987). However, comparison of the restriction maps of these regions of the two plasmids (Fig. 1) showed that a pair of *PstI* and *SalI* restriction sites may be conserved between the two plasmids and that in R751 they are adjacent to *trfA*, whilst in R906 they are adjacent to *oriV*. This suggested that the putative transposable elements between *oriV* and *trfA* may be inserted at different locations. We therefore determined the DNA sequence around the ends of these putative elements and compared the results of this sequencing for R751 and R906.

Fig. 1(*a*) shows that part of the sequences LH (left-hand side as shown in the Figure) and RH (right-hand side) from R906 can be aligned with a continuous sequence segment from R751 to create a 5 bp overlap between LH and RH. Between this duplicated sequence is a segment both of whose ends (in LH and RH) show homology to the ends of Tn501/Tn21-like transposons (Grinstead & Brown, 1984). We have noted previously that the restriction map in the Hg^R region is related to that of Tn501 and there are also similarities in the Ap^R region to transposon Tn2410 (Kratz *et al.*, 1983). Whilst this element is not identical to any known element, this family of transposons is sufficiently diverse as to make it quite likely that the entire 13 kb region is a single transposable element. Transposons of this family are known to generate 5 bp duplications on insertion into a new site (De La Cruz & Grinstead, 1982; Grinstead *et al.*, 1982). Since such a duplication is observed in R906 we propose that this duplication was caused by a single transposon insertion event. Therefore we can deduce the common IncP β backbone in this region simply by

Fig. 1. Restriction maps of the *trfA-oriV* regions of plasmids R906 (a, i) and R751 (b, i) showing the extent of the transposon-like elements in each plasmid (hatched blocks) and indicating the segments whose sequences are compared below (LH and RH). Restriction sites are shown as: *Bam*HI, B; *Bgl*II, Bg; *Clai*I, C; *Eco*RI, E; *Hind*III, H; *Pst*I, P; *Sal*I, S; *Sph*I, Sp; *Sst*II, Ss. (a, ii) Schematic alignment of LH and RH ends of the Hg^RSm^RPn^R element in R906 with R751 and Tn501/Tn21 sequences. (a, iii) Alignment of the actual sequences shown schematically in (a, ii). The sequences not published previously are derived from plasmids pCAS285 (R906 LH), pCAS182 and pCAS184 (R751). The R906 RH sequence comes from Smith & Thomas (1987). (b, ii) Schematic alignment of LH and RH ends of Tn4321 in R751 with R906 and Tn501/Tn21 elements. (b, iii) Alignment of the actual sequences shown schematically in (b, ii). The sequences not published previously are derived from pCAS212 (part of R751 RH), pCAS256 (R906) and pCAS270 (R751 LH). Part of R751 RH was published previously (Smith & Thomas, 1985).

Fig. 2. Comparison of sequences around the N-terminal end of the *trfA* gene of RK2, R751 and R906. For simplicity only RK2 and R751 are compared in the C-terminal region of *trfA* and for an extra 15 nucleotides to show the complete disappearance of homology after the *trfA* stop codon. R906 sequence is only included to show the almost complete conservation of sequences downstream of *trfA* including the putative transcriptional terminator. The RK2 sequence is derived from Smith & Thomas (1984), the R751 sequence (unpublished) is derived from pCAS182 and pCAS184, whilst the R906 sequence comes from Smith & Thomas (1987).

Fig. 1(b) shows that parts of LH and RH from R751 can be aligned with a continuous sequence from R906. However, this does not create an overlap between LH and RH but rather creates a gap of 24 bp. The sequences not homologous to R906 show homology to *Tn501* and *Tn21* suggesting the presence of one or more elements in the same family as the element which disrupts the R906 backbone. However, in contrast to the element in R906, *Tn4321* present in R751 does not encode any known phenotypic markers and is considerably shorter, being only 4.5 kb in length. The gap in the R751 sequence relative to the R906 sequence could be due to loss of IncP backbone DNA by R751 or gain of extra DNA by R906. Two facts suggest the former possibility. First, as shown below (Fig. 3) there is some homology between the R906 sequences present in the gap region and RK2 sequences and the spacing of highly homologous sequences is conserved between these two plasmids, suggesting that the R906 DNA resembles the original IncP backbone.

	<i>XhoI</i>	<i>R1</i>			
R751	TCTCGAGCAT	CGCCACTTCC	GGCGAGGCAG	TACCGCGCAC	CAGCTCGCGC
R906	TCTCGAGCAT	CGCCACTTCC	GGCGAGGCAG	TACCGCGCAC	CAGCCCGCGC
	<i>PstI</i>	<i>R2</i>			
R751	ACGTCGAGCT	GCAGCATCGT	CACCTCCGGC	GAGGCTGGCC	ACGTCCTGGC
R906	ACGTCGAGCT	GCAGCATCGT	CACCTCCGGC	GAGGCTGGCC	ACGTCCTGGT
	<i>SalI</i>	<i>R3</i>		<i>ApaI</i>	
R751	CACGGTCGAC	CATCGCCACA	TCCAACGATG	CCCCGAGGCG	CCCATCGCCA
R906	CACGGTCGAC	CATCGCCACC	TTTGACGATG	AACCGAGGCG	CCCATCGCCA
	<i>BglII</i>				
RK2		AGATC	TAGCGTGGAC	TCAAAGCTCT	CGGAAATGGC
R751	GTTCGACGA	TGACGGCCGC	CG.....CT	GGATCTCGCA	TGCGCACTTC
R906	GTTCGACGA	TGACGGCCGC	CGCGGCGCT	GGATCTCGCA	TGCGCGCTTC
RK2		AGATC	TAGCGTGGAC	TCAAAGCTCT	CGGAAATGGC
	<i>TrfA</i>				
RK2	TGCGGTGGG	AACTTTTCAT	TGACACTTGA	GGGGCAGCGC	A..GGGAAAT
R751	CGACGATGAG	ATCCACTCTT	TGACATTTGA	GGGGCGAAGC	TGCGGGGAT
R906	TGACGATGAG	ATCCACCCCT	TGACACTTGA	GGGGCGGAGC	TGCGGGGAT
RK2	TGCGGTGGG	AACTTTTCAT	TGACACTTGA	GGGGCAGCGC	A..GGGAAAT
RK2	TCTC.....	GTCTTGGC	AGAACCGGCT	ATGTCGTGCT	GGCATCGAG
R751	TCTC.....	GTCTTGGC	AGAACCGGCT	ATGTCGTGCT	GGCATCGAG
R906	TCTC.....	GTCTTGGC	AGAACCGGCT	ATGTCGTGCT	GGCATCGAG
RK2	TCTC.....	GTCTTGGC	AGAACCGGCT	ATGTCGTGCT	GGCATCGAG
RK2	CCTGCGCCCT	T..GGCTTGT	TGCGCCCTCT	CCGCGTGGCT	ACGGGGCTTC
R751	CGACGCGCC	ACGCGCGGTC	TGCGCCCTCT	CCGCGTGGCT	ACGGGGCTTC
R906	CGACGCGCA	GAGCGCGCTC	TGCGCCCTCT	CCGCGTGGCT	ACGGGGCTTC
RK2	CCTGCGCCCT	T..GGCTTGT	TGCGCCCTCT	CCGCGTGGCT	ACGGGGCTTC
	<i>HaeII</i>				
RK2	CAGCGCCTTT	CCGACGCTCA	CCGGGCTGGT	TGCCCTCGCC	GCTGGGCTGG
R751	CAGCGCCTTC	CCGACGCTCA	CCGGGCTGGT	TGCCCTCGCC	GCTGGGCTGG
R906	CAGCGCCTTC	CCGACGCTCA	CCGGGCTGGT	TGCCCTCGCC	GCTGGGCTGG
RK2	CAGCGCCTTT	CCGACGCTCA	CCGGGCTGGT	TGCCCTCGCC	GCTGGGCTGG
RK2	CGCGCGCTTA	TGGCCCTGCA	AACGCGCCAG	AAAGCCGCTC	GAAGCCGTGT
R751	GAGGCGCTGG	CGGCGCTTTC	TGGCCCTT..	..TCTCGG	GGCATAGAG.
R906	CGGCGCGCTA	CGGCGCTTTC	TGGCCCGT..	..TTCGCG	GGCTTCCTCG
RK2	CGCGCGCTTA	TGGCCCTGCA	AACGCGCCAG	AAAGCCGCTC	GAAGCCGTGT
	<i>SstII</i>				
RK2	GCGAGACAC	GCGGCGCGCG	GCCTTGTGGA	TACCTCGCGG	AAACTTGGC
R751	Position of Tn4321	..TTGTGGA	TCGTTTGGGG	ACAAAGCGCC	
R906	CCCTTGCCCC	ACCGGCGCGG	CCCTTGTGGA	CAGTTTCGGG	ACAAGTCGGC
RK2	GCGAGACAC	GCGGCGCGCG	GCCTTGTGGA	TACCTCGCGG	AAACTTGGC
	<i>TrfA</i>		<i>TrfA</i>		
RK2	CCTCACTGAC	AGATGAGGGG	CGGACGTTGA	CACCTGAGGG	GGCGACTCAC
R751	GGCTGTGAC	AGTTGAGGGG	CGTTTCTTGA	CACCTGGGGG	GCAGCGCACC
R906	CGAGGCGCGC	CGGCGCGCTC	CTGACGTTGA	CACCTTGAAGG	GCAGCGGGG.
RK2	CCTCACTGAC	AGATGAGGGG	CGGACGTTGA	CACCTGAGGG	GGCGACTCAC
	<i>TrfA</i>		<i>TrfA</i>		
RK2	CGGCGCGCGC	GTGACAGAT	GAGGGCGAGG	C..TCGATTTC	GACCGCGGAC
R751	CGGCGCGGAT	GCTGACAGAT	GAGGGCGCGG	CCTCGAATTC	GGCGCGC..
R906	CGGCGCGGAT	GTGACAGAT	GAGGGCGAGC	GCGCGATTTC	GGCGCTGG..
RK2	CGGCGCGCGC	GTGACAGAT	GAGGGCGAGG	C..TCGATTTC	GACCGCGGAC

Second, the absence of duplications on either side of the Tn4321 element suggest that the present DNA sequence was not the result of a simple insertion event. It seems likely therefore that the gap resulted either from two closely spaced insertion events of related transposable elements followed by a heterospecific resolution event or from the original element causing a deletion of DNA adjacent to the original site of insertion.

From the results of these sequence comparisons it is therefore possible to reconstruct the likely organization of the ancestral IncPβ plasmid as summarized in Fig. 5.

Relationship between *trfA* and *oriV* in IncPα and IncPβ plasmids

Since it appears possible to reconstruct an ancestral IncPβ plasmid it was of interest to determine the extent to which this organization is also conserved in IncPα plasmids. To establish the relationships between the IncP backbone sequences in the *trfA*–*oriV* region we first compared the sequences from R751 and R906 running towards *trfA* with those of RK2, which include *trfA*, and the sequences between *trfA* and *tetA*, which were determined using plasmids pVII28.1 and pMIK2.1, 3.1, 5.1, 6.1 and 8.1. The result (Fig. 2) shows that there is a high degree of homology between IncPα and IncPβ plasmids within the *trfA* ORF itself. Of the 34 amino acids shown in this figure, 30 (88%) are completely conserved. The four non-identical amino acids show one highly conservative substitution (Glu/Asp), whilst three substitutions are less conservative (Gly/Glu, Ser/Asn, His/Ser). This high degree of conservation is not surprising for a protein whose function is essential for plasmid survival and is consistent with the observation that both IncPα and β *trfA* genes can activate both α and β *oriV* regions (Smith & Thomas, 1985).

In contrast to this high degree of conservation, the IncPα and IncPβ sequences diverge immediately after the stop codon for the *trfA* ORF and no homology is observed in the *trfA*–*tetA* region. The IncPβ sequence downstream of the *trfA* ORF is highly conserved between R751 and R906 and contains a region with the characteristics of a *p*-independent transcriptional terminator. No such sequence is found downstream from the IncPα *trfA* ORF. The position of the IncPβ terminator would prevent *trfA* transcription from entering *oriV* and

Fig. 3. Comparison of sequences upstream of *oriV* in RK2, R751 and R906. The main features shown are the repeated sequences which bind TrfA (indicated by an arrow labelled TrfA). The RK2 sequence comes from Stalker *et al.* (1981) and Cross *et al.* (1986). The R751 sequence comes from pCAS270, pCT800 and Smith & Thomas (1985). The R906 sequence comes from pCAS285, 290, 254, 256 and 258.

possibly interfering with *oriV* function. It may therefore have been an essential element of a clustered *trfA/oriV* region segment. In *IncPα* plasmids the presence of *tetA* and *tetR* between *trfA* and *oriV* may have removed the pressure to retain this feature, since *tetR* is followed by a terminator which would ensure that little *trfA* transcription enters *oriV*. The need to protect *oriV* from such transcription is confirmed by the demonstration that the inability to clone the *trfA* operon of RK2 in the absence of *korA*, which represses the *trfA* promoter, is due to the strength of the *trfA* promoter (Ayres *et al.*, 1991).

Comparison of the *IncPα oriV* sequences with the *IncPβ oriV* backbone sequences (Fig. 3) indicated that significant homology continues until close to the *Bgl*II site upstream of *oriV* in RK2. Therefore the *IncPβ* plasmids also have the regions containing the single repeat and the group of three tandem repeats (iterons) to which TrfA protein binds (Pinkney *et al.*, 1988; Perri *et al.*, 1991). Whilst one of the iterons in the group of three has not been conserved in R906, this general conservation suggests that these sequences have some significance relative to *oriV* function even though they are not essential for replication. This is consistent with our finding that their presence can modulate plasmid copy number (Thomas *et al.*, 1984). There are minor divergences in the TrfA iterons which in RK2 have the consensus TGACA^{C/G}/A^TTGAGGGGC. The T at position 6 in repeat 1 of R751 is also found in repeat 9 of RK2. All of the repeats of R906 have CT at the variable positions, whereas in RK2 they alternate between CT and AG. Finally repeat 2 of R751 and repeat 3 of R906 has the sequence GGGGGG instead of GAGGGG in the second part of the iteron. The effect of this change on iteron function is not known.

It is not just the regions which contain the repeats to which TrfA binds that are conserved between all three plasmids. Some other regions, for example the region from base 340 to 400 in Fig. 3, show even more remarkable conservation. In addition, in the region between the *Xho*I and the site of the first TrfA-binding repeat the *IncPβ* plasmids contain a novel series of five direct repeats (Figs 3 and 4). A feature which may be of interest is that repeats 1–4 are immediately preceded by a hexanucleotide restriction site (*Xho*I, *Pst*I, *Sal*I and *Apa*I, respectively). Repeat 5 has the sequence GGATCTC, which is just one mismatch from a *Bam*HI site. The idea that this might be some sort of recombination region associated with multiple restriction systems encountered in diverse bacteria is intriguing.

From the above comparison we can conclude that in *IncPα* plasmids the *tet* genes were inserted into the backbone of a common ancestor and that possibly a series of events have resulted in the loss of some *IncP* sequences as well as of the putative transposable element

R751 R1	CATCGCCACTTCCGGCGAGG
R906 R1	CATCGCCACTTCCGGCGAGG
R751 R2	CATCGTCACCTTCCGGCGAGG
R906 R2	CATCGTCACCTTCCGGCGAGG
R751 R3	CATCGCCACATCCAACGATG
R906 R3	CATCGCCACCTTTGACGATG
R751 R4	CATCGCCAGTTCCGACGATG
R906 R4	CATCGCCACTTCCGACGATG
R751 R5	CATCGCCACTTCCGACGATG
R906 R5	CATCGCCGCTTCTGACGATG
Con	CATCGCCACTTCCGACGATG
	<<<<< < IVR > >>>>>

Fig. 4. Alignment of the new set of repeats identified downstream of *trfA* in R751 and R906. The arrowheads and IVR indicate the inverted repeat within this repeat unit. Con, consensus.

Fig. 5. Summary of proposed organization of ancestral *IncP* replicons. Unidirectional arrows represent TrfA binding sites. Bidirectional arrows indicate inverted repeats. The flat triangles in the region of the *Xho*I–*Pst*I–*Sal*I sites in R751 and R906 represent the repeats shown in Fig. 4. The hairpin followed by t to the left of the Figure is the terminator present downstream of *trfA* in R751 and R906.

sequences responsible for the original insertion of the *tet* genes. However, our results do not help to establish the relationship between the RK2 *tet* genes and those in Tn1721 to which they are closely related (Waters *et al.*, 1983).

Fig. 5 summarizes our conclusions about the organization of the minimal replication system of *IncP* plasmids. It is intriguing to note that the region between *trfA* and *oriV* has been subject to insertion/deletions on a number of separate occasions during evolution of *IncP* plasmids. The concentration of such events in just a few regions may be the consequence of much of the plasmid being occupied by a series of closely packed operons involved in either plasmid conjugative transfer or maintenance and therefore essential for normal plasmid survival.

This work was supported by MRC project grants G8224213CB, G8309838CB and G8819550CB awarded to C.M.T. Plasmids pCAS258 and 256 were constructed by Jane Adamson and Suzanne Ashpole during an undergraduate project in this laboratory.

References

- AYRES, E., SAADI, S., SCHREINER, H. C., THOMSON, V. J. & FIGURSKI, D. H. (1991). Differentiation of lethal and non-lethal, *kor*-regulated functions in the *kilB* region of broad host-range plasmid RK2. *Plasmid* **25**, 53–63.
- BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1523.
- BOLIVAR, F., RODRIGUEZ, R. L., GREEN, P. J., BETLACH, M. V., HEYNECKER, H. L., BOYER, H. W., CROSA, J. W. & FALKOW, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95–113.
- CHIKAMI, G. K., GUINEY, D. G., SCHMIDHAUSER, T. J. & HELINSKI, D. R. (1985). Comparison of ten IncP plasmids: homology in the regions involved in plasmid replication. *Journal of Bacteriology* **162**, 656–660.
- COUTURIER, M., BEX, F., BERGQUIST, P. L. & MAAS, W. K. (1988). Identification and classification of bacterial plasmids. *Microbiological Reviews* **52**, 375–395.
- CROSS, M. A., WARNE, S. R. & THOMAS, C. M. (1986). Analysis of the vegetative replication origin of broad host range plasmid RK2 by transposon mutagenesis. *Plasmid* **15**, 132–146.
- CURRIER, T. C. & MORGAN, M. K. (1981). Restriction endonuclease analyses of the incompatibility group P-1 plasmids RK2, RP1, RP4, R68 and R68.45. *Current Microbiology* **5**, 323–327.
- DE LA CRUZ, F. & GRINSTED, J. (1982). Genetic and molecular characterization of Tn21, a multiple resistance transposon from R100.1. *Journal of Bacteriology* **151**, 222–228.
- DEVEREUX, J. P., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- FIGURSKI, D. H., POHLMAN, R. F., BECHHOFFER, D. H., PRINCE, A. S. & KELTON, C. A. (1982). The broad-host-range plasmid RK2 encodes multiple *kil* genes potentially lethal to *Escherichia coli* host cells. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 1935–1939.
- GRINSTED, J. E. & BROWN, N. L. (1984). A Tn21 terminal sequence within Tn501: complementation of *tnpA* gene function and transposon evolution. *Molecular and General Genetics* **197**, 497–502.
- GRINSTED, J., DE LA CRUZ, F., ALTENBUCHNER, J. & SCHMITT, R. (1982). Complementation analysis of transposition of *tnpA* mutants of Tn3, Tn21, Tn501 and Tn1721. *Plasmid* **8**, 276–286.
- HEDGES, R. W., MATTHEW, M., SMITH, D. I., CRESSWELL, J. M. & JACOB, A. E. (1977). Properties of a transposon conferring resistance to penicillins and streptomycin. *Gene* **1**, 241–253.
- JOBANPUTRA, R. S. & DATTA, N. (1974). Trimethoprim R-factors in Enterobacteria from clinical specimens. *Journal of Medical Microbiology* **7**, 169–177.
- KAHN, M., KOLTER, R., THOMAS, C., FIGURSKI, D., MEYER, R., REMAUT, E. & HELINSKI, D. R. (1979). Plasmid cloning vehicles derived from plasmids ColE1, F, R6K and RK2. *Methods in Enzymology* **68**, 268–280.
- KRATZ, J., SCHMIDT, F. & WIEDEMANN, B. (1983). Characterization of Tn2411 and Tn2410, two transposons derived from R-plasmid R1767 and related to Tn2603 and Tn21. *Journal of Bacteriology* **155**, 1333–1342.
- LANKA, E., FÜRST, J. P., YAKOBSON, E. & GUINEY, D. G. (1985). Conserved regions at the DNA primase locus of the IncP α and IncP β plasmids. *Plasmid* **14**, 217–223.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- MAXAM, A. & GILBERT, W. (1980). Sequencing end-labelled DNA with base-specific cleavage. *Methods in Enzymology* **65**, 449–560.
- NORRANDER, J., KEMPE, T. & MESSING, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **27**, 101–106.
- PERRI, S., HELINSKI, D. R. & TOUKDARIAN, A. (1991). Interactions of plasmid-encoded replication initiation proteins with the origin of DNA replication in the broad host range plasmid RK2. *Journal of Biological Chemistry* **266**, 12536–12543.
- PINKNEY, M., DIAZ, R., LANKA, E. & THOMAS, C. M. (1988). Replication of mini RK2 plasmid in extracts of *Escherichia coli* requires plasmid-encoded protein TrfA and host-encoded proteins DnaA, B, G, DNA gyrase and DNA polymerase III. *Journal of Molecular Biology* **204**, 927–938.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SHINGLER, V. & THOMAS, C. M. (1984). Analysis of the *trfA* region of broad host range plasmid RK2 by transposon mutagenesis and identification of polypeptide products. *Journal of Molecular Biology* **175**, 229–250.
- SMITH, C. A. & THOMAS, C. M. (1984). Nucleotide sequence of the *trfA* gene of the broad host range plasmid RK2. *Journal of Molecular Biology* **175**, 251–262.
- SMITH, C. A. & THOMAS, C. M. (1985). Comparison of the nucleotide sequences of the vegetative replication origins of broad host range IncP plasmids R751 and RK2 reveals conserved features of probable functional importance. *Nucleic Acids Research* **13**, 557–572.
- SMITH, C. A. & THOMAS, C. M. (1987). Comparison of the organization of the genomes of phenotypically diverse plasmids of incompatibility group P: members of the IncP β subgroup are closely related. *Molecular and General Genetics* **206**, 419–427.
- SMITH, C. A. & THOMAS, C. M. (1989). Relationships and evolution of IncP plasmids. In *Promiscuous Plasmids of Gram-Negative Bacteria*, pp. 57–77. Edited by C. M. Thomas. London: Academic Press.
- STALKER, D. M., THOMAS, C. M. & HELINSKI, D. R. (1981). Nucleotide sequence of the origin of replication of the broad host range plasmid RK2. *Molecular and General Genetics* **181**, 8–12.
- THOMAS, C. M. & SMITH, C. A. (1987). Incompatibility group P plasmids: genetics, evolution and use in genetic manipulation. *Annual Review of Microbiology* **41**, 77–101.
- THOMAS, C. M., CROSS, M. A., HUSSAIN, A. A. K. & SMITH, C. A. (1984). Analysis of the copy number control elements in the region of the vegetative replication origin of broad host range plasmid RK2. *EMBO Journal* **3**, 57–63.
- THOMAS, C. M., MEYER, R. & HELINSKI, D. R. (1980). Regions of the broad host range plasmid RK2 which are essential for replication and maintenance. *Journal of Bacteriology* **141**, 213–222.
- VIEIRA, J. & MESSING, J. (1982). The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.
- VILLAROL, R., HEDGES, R. W., MAENHAUT, R., LEEMANS, J., ENGLER, G., VAN MONTAGU, M. & SCHELL, J. (1983). Heteroduplex analysis of P-plasmid evolution: the role of insertion and deletion of transposable elements. *Molecular and General Genetics* **189**, 390–399.
- WATERS, S. H., GRINSTED, J. E., ROGOWSKY, P., ALTENBUCHNER, J. & SCHMITT, R. (1983). The tetracycline resistance determinants of RP1 and Tn1721: nucleotide sequence analysis. *Nucleic Acids Research* **11**, 6089–6105.
- WOMBLE, D. D. & ROWND, R. H. (1988). Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. *Microbiological Reviews* **52**, 433–451.
- YAKOBSON, E. & GUINEY, D. (1983). Homology in the transfer origins of broad-host-range IncP plasmids: definition of two subgroups of P-plasmids. *Molecular and General Genetics* **192**, 436–438.