Expression of leading region genes on IncI1 plasmid ColIb-P9: genetic evidence for single-stranded DNA transcription

Steven Bates,1† Richard A. Roscoe,1 Nicola J. Althorpe,1 William J. Brammar2 and Brian M. Wilkins1

Author for correspondence: Brian Wilkins. Tel: +44 116 252 3432. Fax: +44 116 252 5101.
e-mail: bmw1@le.ac.uk

INTRODUCTION

The colicinogenic plasmid ColIb-P9 (93 kb) and closely related R64 plasmid represent the IncI1 group of enterobacterial conjugative plasmids. I1 plasmids possess a large transfer region (~ 50 kb): this encodes two types of conjugal pilus, which comprise the system for the formation of mating contacts (see Rees et al., 1987; Kim & Komano, 1997) and a DNA processing mechanism. DNA transfer is initiated by a strand-specific cleavage at the unique nic site in the origin-of-transfer region (oriT; Furuya & Komano, 1991; Lanka & Wilkins, 1995). Specific cleavage at nic potentiates transfer of a unique plasmid strand, called the T-strand, to the recipient cell (Vapnek et al., 1971). Strand transfer is inferred to occur with a 5′ to 3′ polarity (Wilkins & Lanka, 1993).

ColIb oriT flanks one end of the I1 transfer region and is orientated such that the transfer genes are the last to enter the recipient cell (Howland & Wilkins, 1988). The first portion of the plasmid to be transferred is called the leading region. This sector of ColIb carries some inessential genes that aid establishment of the entrant DNA in the new host cell. These genes are: ssb, encoding a ssDNA-binding protein of the type required for bacterial DNA replication (Howland et al., 1989); psiB, determining a protein that inhibits induction of the bacterial SOS stress response in conjugation (Howland et al., 1989); and ardA, which functions specifically in the conjugatively active recipient cell to alleviate type I restriction of the immigrant plasmid (Read et al., 1992; Althorpe et al., 1999). Highly conserved homologues of these genes are found in different combinations in the leading regions of enterobacterial plasmids of other Inc groups (Golub et al., 1988; Chilley & Wilkins, 1995). Measurement of transcript levels and the accumulation of gene products indicate that ardA, psiB and ssb genes function as ‘early’ genes that are expressed rapidly and transiently in the conjugatively infected cell (Bagdasarian et al., 1992; Jones et al., 1992; Althorpe et al., 1999). This mode of gene expression is called zygotic induction.

We report here tests indicating that zygotic induction is
regulated by a process which is independent of a trans-
acting repressor but dependent on the transcriptional
orientation of the inducible gene on the T-strand. To
identify motifs responsible for this mode of gene
expression, the nucleotide sequence of 11.7 kb of the
ColIb leading region was determined. The sequence
contains three dispersed repeats indicative of single
strand promoters. We propose that these signals initiate
rapid transcription of genes on the incoming T-strand to
facilitate establishment of the incoming plasmid in the
new host cell.

METHODS
Plasmids. Colllb-P9 derivatives and recombinant plasmids
used are described in Table 1. E and S fragments refer to EcoRI
and SalI restriction enzyme cleavage products, respec-
tively; fragments are numbered according to rank order of
size. Recombinant pRR2-1 contained, in addition to E1, the
E16 fragment which is apparently necessary for stable cloning
of E1. pRR8 was constructed using a strategy similar to that
previously described (Jones et al. 1987). Oligonucleotides for
sequencing primers were generated using a Perkin Elmer
Biosystems (PE Biosystems) 394 synthesizer. Sequencing
reactions used standard dye terminator kits from PE
Biosystems and were analysed with a PE Biosystems 377
automated DNA sequencer. Both machines were operated by
the PNAI service of Leicester University. Sequence data
were analysed using the Wisconsin package version 10.0,
Genetics Computer Group (GGC), Madison, Wisconsin, USA.
The complete sequence was compiled using the GEL set of
programs. FASTA and BLAST searches were carried out through
the European Bioinformatics Institute site on the web
(http://www.ebi.ac.uk). The entire sequence of Colllb-P9 has
recently been made available as EMBL accession number
AB021078.

RESULTS AND DISCUSSION
A repressor model for psIB control

One hypothesis to explain zygotic induction is that the
conjugatively activated genes are regulated by a plasmid-
encoded, trans-acting repressor which accumulates by
de novo synthesis in the conjugatively infected cell to
repress transcription. A prediction of this hypothesis is that
psIB induction should be prevented if Colllb transfers to a recipient cell already carrying a copy of the
plasmid. Such a test requires inactivation of the plasmid
entry exclusion function that inhibits transfer between
cells harbouring the same plasmid. For this purpose, an

<table>
<thead>
<tr>
<th>Table 1. Plasmids used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plasmid</strong></td>
</tr>
<tr>
<td>pLG257</td>
</tr>
<tr>
<td>pLG290</td>
</tr>
<tr>
<td>pLG2009</td>
</tr>
<tr>
<td>pLG2038</td>
</tr>
<tr>
<td>pLG2062</td>
</tr>
<tr>
<td>pRR1</td>
</tr>
<tr>
<td>pRR2-1</td>
</tr>
<tr>
<td>pRR2-4</td>
</tr>
<tr>
<td>pRR2-12</td>
</tr>
<tr>
<td>pRR2-14</td>
</tr>
<tr>
<td>pRR8</td>
</tr>
</tbody>
</table>

* E1 is the largest EcoRI fragment of Colllb (Fig. 3). S4 is the SalI fragment internal to E1; S3 is located to the left of S4 in the orientation
shown in Fig. 3. Cm, chloramphenicol; Km, kanamycin.
Expression of leading region genes

exclusion (exc) mutant of ColIIb (pRR8) was generated by replacing the central portion of the exc locus with a TeC marker (see Methods). The exclusion phenotype of pRR8 was examined by measuring its effect on transconjugant production by a pBR322-based ColIIb oriT recombinant (pLG2009), mobilized from donors by ColIIbdrd-1 (pLG273). Such an assay of pLG2009, rather than ColIIb-containing transconjugants avoids incompatibility effects exerted by pRR8 on the establishment of an incoming I1 plasmid. The exclusion index of pRR8, defined as the ratio of pLG2009 transconjugants formed in the plasmid-free recipient strain relative to the homogenic strain carrying pRR8, was 1:4. Evidently pRR8 confers a slight but significant barrier to ColIIb-directed plasmid transfer for which allowance must be made.

The repressor hypothesis was tested by measuring the effect that carriage of pRR8 exerts on zygotic induction of a psiB::lacZ operon fusion transferred on a ColIIb plasmid (Fig. 1a). Transfer of the lacZ fusion resulted in a burst of β-galactosidase accumulation shown previously to occur in the transconjugant cell (Jones et al., 1992). In such a conjugation, transfer is initiated asynchronously in the cell population typically to give a burst of transconjugant production between 7 and 20 min by which time about 60% of the input recipient cells have acquired the ColIIbdrd plasmid. Subsequently, the rate of increase of transconjugant production diminishes. Within this framework, transfer of a single ColIIb molecule is thought to take less than 3 min and to be associated with an even briefer burst of psiB transcription. Whereas psiB transcripts are unstable, β-galactosidase is a metabolically stable protein (Althorpe et al., 1999).

Results obtained when a ColIIbdrd plasmid (pLG2038) carrying the psiB::lacZ fusion was used to test the repressor hypothesis are shown in Fig. 1(a). β-Galactosidase specific activity was observed to increase substantially irrespective of whether the recipient strain contained the exclusion mutant of ColIIb (pRR8) or was plasmid-free. When allowance is made for the different transfer frequencies – determined by pLG2009 transconjugant yields – it is estimated that 60% of the normal level of zygotic induction occurred in the recipients harbouring the pre-existing copy of ColIIb.

In a parallel experiment (Fig. 1b), the psiB::lacZ fusion was transferred on an oriT-recombinant plasmid (pRR2-12) which was mobilized by a non-transmissible ColIIb plasmid (pLG2062) in the donor. The latter plasmid determines all of the transfer functions but is immobile due to deletion of the mic site. Specific activity of β-galactosidase was found to increase linearly in both the plasmid-free and pRR8-containing recipient cells to give approximately threefold higher values than those attained in Fig. 1(a). The increases are attributed to enhanced transfer of the psiB::lacZ recombinant plasmid (pRR2-12) rather than to its higher copy number relative to ColIIb. The favoured explanation is that multiple rounds of pRR2-12 transfer occur when the plasmid is mobilized by a non-transmissible ColIIb plasmid, possibly because the conjugation cycle is not terminated until the recipient has acquired a copy of the conjugative plasmid. This interpretation is suggested by data in Fig. 2, showing that when pRR2-12 was mobilized by a transmissible ColIIb plasmid, the amount of zygotic induction per transconjugant cell was similar to that determined when the psiB::lacZ fusion was

...
located on ColIb itself (Fig. 1a). Taken together, the results in Fig. 1 contradict the repressor hypothesis by showing that a recipient cell carrying a pre-existing copy of ColIb can support one or more cycles of zygotic induction.

**Zygotic induction depends on gene orientation on the leading region**

The consistent orientation of ardA, psiB and ssb on different plasmids suggests that zygotic induction requires the genes to be arranged in a specific direction in the leading region (Chilley & Wilkins, 1995). To test this possibility, the 14.5 kb SalI (S4) fragment containing psiB::lacZ was inverted in the E1 fragment (Fig. 3) and the effect of the rearrangement on expression of the fusion gene was determined. To simplify the constructions, the inversion was made in a pBR328-based recombinant containing E1, rather than in ColIb itself. Mobilization tests (Fig. 2) showed that zygotic induction of β-galactosidase synthesis was only observed with the construct (pRR2-12) containing psiB in the natural orientation relative to oriT. While the psiB gene on the inversion plasmid (pRR2-14) was expressed in vegetative cells from an uncharacterized promoter, no enhanced expression occurred during conjugative transfer. Thus zygotic induction of psiB is dependent on the orientation of the gene in the leading region.

Inversion of the S4 fragment disrupts the downstream orf5 gene (Fig. 3), raising the hypothesis that zygotic induction requires carriage of a functional orf5 gene in cis. This possibility can be rejected because the lacZ–kanamycin resistance cassette used to construct the psiB::lacZ operon fusion contains a bidirectionally active transcriptional terminator downstream of the lacZ component (Kokotek & Lotz, 1989). Since orf5 is thought to be part of the same transcriptional unit as psiB (see next section), the terminator should prevent orf5 transcription on both pRR2-12 and pRR2-14. Another possible explanation of the orientation-dependent induction of psiB is that inversion of S4 uncouples the gene from a promoter outside the fragment. If so, the hypothetical promoter would be at least 7-7 kb upstream of psiB (Fig. 3). A more intriguing possibility is that psiB can only undergo zygotic induction when orientated in a specific direction relative to oriT.

**Orientation of ORFs in the ColIb leading region**

To establish whether all ORFs in the leading region have the same orientation, the sequence of the first 11712 nt in the ColIb leading region was determined. The sequence starts at the nic site which was located previously by sequence alignment with known features of the R64 oriT region (Althorpe et al., 1999). Ten ORFs were identified (Table 2; Fig. 3): three correspond to ardA, psiB and ssb, while a fourth is a homologue of the psiA gene of F-like plasmids (Loh et al., 1990). Six other candidate genes (orf1–6, orf7–6) were identified on the basis of a codon preference bias similar to that in known leading region genes and a separation of at least 4 nt between the initiator codon and the putative Shine–Dalgarno sequence. Importantly, all 10 ORFs have the same transcriptional orientation.

Properties of orf5 and orf6 are relevant to this study. The orf5 locus is homologous to the single protein-encoding gene present in a heterogeneous group of unusual insertion elements that lack the terminal repeat structures associated with classic insertion sequences. The elements are typified by IS891 from the cyanobacterium Anabaena sp. (Bancroft & Wolk, 1989) and IS1253 from Dichelobacter nodosus (Billington et al., 1996). Thus orf5 may encode the transposase of a functional IS element. As described for IS1253, orf5 is flanked by an imperfect direct repeat (coordinates 5943–5972 and 7231–7261) containing an internal region of dyad symmetry, which may be the recognition sequence for the transposase. The G + C content of orf5 is low relative to other parts of the leading region, suggesting that the locus is a relatively recent insertion (Fig. 3).

Orf6 is predicted to be a DNA-binding protein since it contains a helix–turn–helix (residues 186–205) motif of the AraC family of regulatory proteins which is indicative of sequence-specific DNA-binding proteins. The protein also contains a potential ATP/GTP binding-site motif A (P-loop; residues 178–185). Apparently F-like plasmids also carry an orf6 homologue, as indicated by strong sequence similarity of orf6 to the sequence...
Expression of leading region genes

**Fig. 3.** Maps and G+C profile of the ColIb leading region. (a) Outline map of the EcoRI E1 fragment (20–4 kb) showing the internal SalI S4 fragment. The direction of transfer from nic is from left to right. (b) Location and orientation of ORFs and ssi sites in the leading portion of the T-strand. All ORFs (open arrows) are transcribed in the leftward direction. The ssi sites (filled arrows) are presumptive promoters for leftward transcription of the T-strand. Sites t1 and t2 are putative transcriptional terminators. Restriction cleavage sites are: B, BglII; C, Clal; E, EcoRI; P, PstI; S, SalI. (c) G+C profile of the leading region as determined in a moving window of 500 nt and a shift increment of 25 nt. The horizontal line indicates the mean value of 55–7 mol% G+C for the 11–7 kb leading region. Coordinate 1 is the nucleotide 3' of the specific cleavage site at nic.

**Table 2.** Known and potential genes, and putative transcription signals, in the ColIb-P9 leading region

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates from nic</th>
<th>Start codon</th>
<th>Amino acid residues</th>
<th>Mass (kDa)</th>
<th>Motifs*</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>552–520</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ρ-independent terminator</td>
</tr>
<tr>
<td>orf1</td>
<td>1409–561</td>
<td>AUG</td>
<td>282</td>
<td>32.2</td>
<td>–</td>
<td>Unknown: methylase?</td>
</tr>
<tr>
<td>ssi1</td>
<td>1787–1459</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Single strand promoter</td>
</tr>
<tr>
<td>t2</td>
<td>2494–2466</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ρ-independent terminator</td>
</tr>
<tr>
<td>orf2</td>
<td>3444–2503</td>
<td>GUG</td>
<td>313</td>
<td>35.8</td>
<td>–</td>
<td>yadD homologue</td>
</tr>
<tr>
<td>orf3</td>
<td>3775–3509</td>
<td>GUG</td>
<td>88</td>
<td>10.5</td>
<td>–</td>
<td>Unknown</td>
</tr>
<tr>
<td>orf4</td>
<td>4301–3867</td>
<td>AUG</td>
<td>144</td>
<td>16.5</td>
<td>–</td>
<td>Unknown</td>
</tr>
<tr>
<td>ssi2</td>
<td>4678–4353</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Single strand promoter</td>
</tr>
<tr>
<td>ardA</td>
<td>5530–5030</td>
<td>AUG</td>
<td>166</td>
<td>19.2</td>
<td>–</td>
<td>Antirestriction protein</td>
</tr>
<tr>
<td>orf5</td>
<td>7153–5978</td>
<td>AUG</td>
<td>391</td>
<td>43.9</td>
<td>–</td>
<td>Putative transposase</td>
</tr>
<tr>
<td>psiA</td>
<td>8002–7283</td>
<td>AUG</td>
<td>239</td>
<td>27.7</td>
<td>–</td>
<td>Unknown</td>
</tr>
<tr>
<td>psiB</td>
<td>8433–7999</td>
<td>AUG</td>
<td>144</td>
<td>15.9</td>
<td>–</td>
<td>Inhibitor of SOS induction</td>
</tr>
<tr>
<td>orf6</td>
<td>10527–8488</td>
<td>AUG</td>
<td>679</td>
<td>74.3</td>
<td>P-loop; helix–turn–helix motif; coiled coil</td>
<td>DNA binding?</td>
</tr>
<tr>
<td>ssi3</td>
<td>11314–10787</td>
<td>AUG</td>
<td>175</td>
<td>19.2</td>
<td>ssDNA binding</td>
<td>ssDNA-binding protein</td>
</tr>
<tr>
<td>ssi3</td>
<td>11711–11381</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Single strand promoter</td>
</tr>
</tbody>
</table>

* Motifs were detected using the GCG program MOTIFS searching against the PROSITE database. Predicted helix–turn–helix motifs and coiled coils were detected using HTTHSCAN and COILSCAN, respectively.

upstream of psiB on F and R6-5 (Dutreix et al., 1988). Thus it is proposed that psiA-psiB-orf6-ssb comprise a conserved module on ColIb and F-like plasmids. The module is inferred to be expressed as a single transcriptional unit on both plasmids, as indicated by activation of the PsiB phenotype by insertions in the upstream ssi3 gene (Dutreix et al., 1988; Jones et al., 1992) and the overlap between the start codon of psiA and the stop codon of psiB (Loh et al., 1990; see also Table 2).

**Three copies of a putative single strand promoter**

The leading region is remarkable in containing three repeats of a 328 bp element homologous to a novel promoter in ssDNA called Frpo (Masai & Arai, 1997).
The repeats (ssi1, ssi2 and ssi3; Table 2, Fig. 3) are located in the same orientation approximately 55 nt upstream of orf1, orf4 and ssb; they share \( \geq 80\% \) sequence identity and display \( \geq 59\% \) identity to the central 150 nt portion of Frpo shown in Fig. 4.

Frpo was isolated from the F leading region as a single-stranded initiation sequence (ssiD) active in priming DNA synthesis on the viral strand of phage M13 (Nomura et al., 1991). The promoter is formed in ssDNA as a specific 328 bp secondary structure containing a number of mispaired regions plus a subterminal single-stranded loop (Fig. 4). Transcription from Frpo was stimulated highly by SSB, which is thought to bind the loops and facilitate binding of RNA polymerase through protein–protein interactions on the template DNA. The promoter was only active in ssDNA where it directed \( E. \) coli RNA polymerase–\( \sigma^{70} \) to start transcription from a specific site within the stem of the secondary structure (Masai & Arai, 1997). Frpo maps upstream of the third ORF (orf95) in the F leading region at a location containing an \textit{in vivo} transcriptional start site (Loh et al., 1989).

The ColIb ssi elements have predicted secondary structures similar to that of Frpo (Fig. 4). The structure shown for Frpo is based on sensitivity to a single-strand-specific nuclease (Masai & Arai, 1997); a thermodynamically favoured variant predicted by the mfold program (Mathews et al., 1999) is shown as Frpo* (Fig. 4). The secondary structures for ssi1, 2 and 3 are similar thermodynamically favoured predictions of the program. We therefore postulate that ssi1, 2 and 3 are also promoters active in ssDNA. The putative –10 regions contain multiple mismatches. As shown with synthetic heteroduplexes, mismatches in the \( \textit{fi}10 \) region may aid promoter function by facilitating melting of the duplex and formation of the open complex necessary for transcription. Strand bias exists in that \( E. \) coli RNA polymerase–\( \sigma^{70} \) primarily recognizes bases in the non-template strand of the \( \textit{fi}10 \) region, with \( \textit{fi}7T, \textit{fi}11A \) and \( \textit{fi}12T \) being the three most conserved residues (Roberts & Roberts, 1996). Two of these three bases are present in the non-template strand of the putative –10 region of Frpo and ColIb ssi2 and ssi3.

Promoter strength is also affected by the spacing between the –35 and –10 hexamers. While a separation of 17 bp in duplex DNA is optimal, changes of 3 bp can be tolerated – albeit with some loss of function – possibly as a result of DNA deformation induced by the holoenzyme (Dombroski et al., 1996). The spacing between the putative –35 and –10 regions of the ssi elements and Frpo ranges from 18 to 20 nt. Such elongated spacings might be required to compensate for the mispaired nucleotides within them.

Frpo and the ColIb ssi elements are orientated in the respective leading regions such that the promoter structure is formed in the T-strand. This feature leads to...
the appealing model that zygotic induction of leading region genes is due to the activity of Frrp-like sequences in directing transcription of the transferring plasmid strand. The model predicts that a psiB inversion will be non-inducible, as observed in Fig. 2, because the T-strand will contain the antisense strand of the gene. Termination of transcription may occur immediately downstream of orfI and orf2 at sequences t1 and t2 (Table 2, Fig. 3), which have the characteristics of p-independent transcription terminators (Platt, 1986). Termination at t1 will prevent transcription across orfT and potential interference with reactions involved in the termination of a round of DNA transfer at the oriT region and circularization of the transferring plasmid strand (see Langa & Wilkins, 1995).

The ssi promoters will be silenced by synthesis of the complementary DNA strand. There is no need to invoke a role for ColIb ssi elements in initiating complementary strand synthesis, since this mode of DNA synthesis is normally initiated by RNA primers made by a dedicated plasmid-encoded DNA primase. This enzyme is transmitted selectively from the donor to the recipient cell during bacterial conjugation (Rees & Wilkins, 1989).

Transcription of the incoming T-strand has biological appeal. It allows the AroA antirestriction protein to accumulate in the newly infected cell before the T-strand is converted into duplex DNA, the substrate of type-I restriction enzymes. It allows rapid expression of Psib as a function that prevents induction of the bacterial SOS response by the presumptive trigger of transferring ssDNA and it gives elevated levels of SSB protein, which may prevent drainage of the cellular counterpart and potentiate protein–DNA interactions peculiar to conjugation. The mechanism will also cause a burst of Orf5 transposase production in the new cell, which is expected to initiate transposition of the IS element and allow its spread to the genomes of bacteria outside the host range of ColIb. Finally, the concept that the plasmid leading region contains ‘early’ genes that are expressed transiently in the conjugatively infected cell to promote establishment of the incoming genetic element puts plasmids on a mechanistic parallel with bacterial viruses which are known to have evolved a variety of strategies for phasing gene expression during cellular infection.

ACKNOWLEDGEMENTS

This work was supported by MRC grant G9321196MB. R.A.R. held a BBSRC research studentship; N.J.A held a MRC research studentship.

REFERENCES


Received 2 June 1999; revised 28 July 1999; accepted 29 July 1999.