Microbiology (2000), **146**, 305–313 Printed in Great Britain

Analysis of the internal replication region of a mycobacterial linear plasmid

Mathieu Picardeau, Corinne Le Dantec and Véronique Vincent

Author for correspondence: Mathieu Picardeau. Tel: +33 1 45 68 83 60. Fax: +33 1 40 61 31 18.
e-mail: mpicard@pasteur.fr

Laboratoire de Référence des Mycobactéries, Institut Pasteur, 75724 Paris Cedex 15, France

Linear plasmids have previously been identified by the authors in mycobacteria, the telomeres of which have terminal inverted repeats and covalently attached proteins. In this study, the replication of these unusual molecules was investigated by studying a 25 kb linear plasmid from the slow-growing species *Mycobacterium celatum* called pCLP. An internal region of pCLP responsible for replication in *Mycobacterium smegmatis* was identified. The nucleotide sequence of the minimum replication region of pCLP, which was 2·8 kb long, contained a putative replication gene, *rep*, and a putative origin of replication consisting of an 18 bp direct repeat and an AT-rich region. A short section of the pCLP replication region was also found to have sequence identity with the replication regions of mycobacterial circular plasmids, suggesting that these linear and circular plasmids are related. It was found that pCLP replicated in *Mycobacterium bovis* BCG and was compatible in *M. smegmatis* with pAL5000- and pJAZ38-derived plasmids from *Mycobacterium fortuitum*, which belong to two different compatibility groups. Thus, this new *Escherichia coli*–mycobacteria shuttle vector may be used in both slow- and fast-growing mycobacteria and in co-transformation experiments with other mycobacterial vectors.

Keywords: mycobacteria, linear plasmid, shuttle vector, replication

INTRODUCTION

Chromosomal DNA and many plasmids are linear in the genus *Streptomyces* (Hinnebusch & Tilly, 1993). Linear plasmids have recently been described in the other actinomycetes *Planobispora*, *Rhodococcus* and *Mycobacterium* (Crespi *et al.*, 1992; Dabrock *et al.*, 1994; Kalkus *et al.*, 1993; Kebeler *et al.*, 1996; Kosono *et al.*, 1997; Picardeau & Vincent, 1997; Polo *et al.*, 1998). All these linear replicons belong to a class of genetic elements called invertrons (Hinnebusch & Tilly, 1993; Sakaguchi, 1990) which have terminal inverted repeats with 5′ ends covalently linked to a terminal protein. The invertron-like structure is also found in some bacteriophages, adenoviruses and mitochondrial linear plasmids (Hinnebusch & Tilly, 1993; Sakaguchi, 1990). The unusual structure of these linear elements raises several questions about their replication mechanism. Previous studies have demonstrated that the replication of linear plasmids and of linear chromosomes in *Streptomyces* spp. proceeds bidirectionally from a central origin towards the telomeres (Chang & Cohen, 1994; Chang *et al.*, 1996; Fischer *et al.*, 1998; Musialowski *et al.*, 1994; Shiffman & Cohen, 1992; Zakrzewska-Czerwinska & Schrempf, 1992). This internally initiated replication leaves single-strand gaps at the 3′ ends (Chang & Cohen, 1994). The 3′ strand overhang then folds back, due to the presence of multiple terminal palindromes, resulting in a DNA duplex, thereby providing a recognition site for the terminal protein, which serves as a primer to complete DNA synthesis at the telomeres (Qin & Cohen, 1998).

Several circular plasmids have been detected in mycobacteria living in the environment or as opportunistic pathogens, but never in the *Mycobacterium tuberculosis* complex. The replication origins of circular plasmids from *Mycobacterium avium*, *Mycobacterium fortuitum* and *Mycobacterium scrofulaceum* have been described,

Abbreviations: Ap, ampicillin; Hg, hygromycin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin.

The GenBank accession number for the nucleotide sequence and putative ORFs of the replication origin region of pCLP determined in this work is AF144883.

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics*	Source or reference
Strains		
E. coli DH5α	–	Hanahan (1983)
M. smegmatis mc^{2}155	Competent for electroporation	Snapper *et al*. (1990)
M. smegmatis EP10	mc^{2}155, Km^r	Ainsa *et al*. (1997)
M. bovis BCG	BCG Pasteur strain	CNRM†
Plasmids		
pPV8	Km^r derivative of pUC19	V. Pelicic, Institut Pasteur, Paris, France
pB4	Km^r derivative of pAL5000 from *M. fortuitum*	Ranes *et al*. (1990)
pJAZ42	pJAZ40 plus Sm/Sp resistance cassette	Gavigan *et al*. (1997)
pCLP	25 kb linear plasmid from in *M. celatum* 4	Picardeau & Vincent (1998)
pMPV7	10 kb *Xba*I fragment of pCLP in pBS, Ap^r	Picardeau & Vincent (1998)
pMPV3	14·5 kb *Xba*I fragment of pCLP in pBS, Ap^r	Picardeau & Vincent (1998)
pCL4A	*Kpn*I fragment A of pCLP in pPV8	This study
pCL4B	*Kpn*I–*Xba*I fragment B of pCLP in pPV8	This study
pCL4C	*Kpn*I–*Xba*I fragment C of pCLP in pPV8	This study
pCL4D	*Kpn*I fragment D of pCLP in pPV8	This study
pBSh-D	*Kpn*I fragment D of pCLP in pBS, Hg^r	This study
pCL4E	*Kpn*I fragment E of pCLP in pPV8	This study
pCL4F	*Kpn*I–*Xho*I subclone of pCL4D in pPV8	This study
pCL4G	*Xho*I–*Kpn*I subclone of pCL4D in pPV8	This study
pCL5	*Sal*I–*Kpn*I subclone of pCL4D in pPV8	This study
pCRa	Poa/Poc-amplified product in pPV8	This study
pCRb	Pob/Poc-amplified product in pPV8	This study

* Locations of plasmid fragments are shown in Fig. 1, except for pCL5, pCRa and pCRb, which are given in Fig. 5.

† CNRM, Centre National de Référence des Mycobactéries, Institut Pasteur, Paris, France.

and this has facilitated the construction of shuttle vectors capable of replication in both mycobacteria and *Escherichia coli* (Beggs *et al*., 1995; Gavigan *et al*., 1997; Qin *et al*., 1994; Ranes *et al*., 1990). However, the genetic system for studying mycobacteria is still limited. Some of the *E. coli*–mycobacteria shuttle vectors cannot replicate in the fast-growing *Mycobacterium smegmatis* strain mc^{2}155 which is commonly used as a cloning host for studying mycobacterial genes (Snapper *et al*., 1990). Moreover, pAL5000-derived plasmids are the only class that replicate in both slow- and fast-growing species.

We have recently characterized linear plasmids in the slow-growing species *M. avium*, *Mycobacterium xenopi*, *Mycobacterium branderi* and *Mycobacterium celatum* (Picardeau & Vincent, 1997, 1998), and in the fast-growing species *M. fortuitum* (C. Le Dantec, M. Picardeau & V. Vincent, unpublished results). The 25 kb plasmid pCLP from the opportunistic pathogen *M. celatum* has been cloned and its telomeres sequenced. The telomeres have structural features in common with other actinomycete linear plasmids (Picardeau & Vincent, 1998). In this study, we cloned and characterized the internal region of pCLP, which is able to replicate in a circular form in *M. smegmatis* mc^{2}155, and in the slow-growing species *Mycobacterium bovis* BCG. The nucleotide sequence of the replication region includes two ORFs encoding putative replication and partition proteins, and a short region with a sequence similar to that of the origin regions of mycobacterial circular plasmids. Our vector was compatible with plasmids derived from the circular plasmids pAL5000 and pJAZ38 from *M. fortuitum*.

METHODS

Bacterial strains, plasmids and culture conditions. The *M. celatum* strain 4 used in this study is a clinical isolate kindly provided by W. R. Butler, which we have previously shown to contain two linear replicons, one of about 25 kb designated pCLP, and another of 320 kb (Picardeau & Vincent, 1998). Cultures of mycobacteria were grown in 7H9 Middlebrook liquid medium at 37 °C with antibotics added to the media as required. The bacterial strains and plasmids used in this study are listed in Table 1.

Cloning and electrotransformation. The recombinant plasmids pMPV7 and pMPV3 used in this study correspond to the insertion of the two *Xba*I fragments of pCLP (the extremities of which are blunt-ended) between the *Xba*I and *Hinc*II restriction sites of pBluescript II KS(+/−) (Stratagene) (Picardeau & Vincent, 1998). The 10 and 14·5 kb inserts of pMPV7 and pMPV3, respectively, correspond to the entire length of the linear plasmid pCLP (Picardeau & Vincent, 1998). These recombinant plasmids were digested with *Bst*EII, *Kpn*I and *Xho*I to construct a restriction endonuclease map of

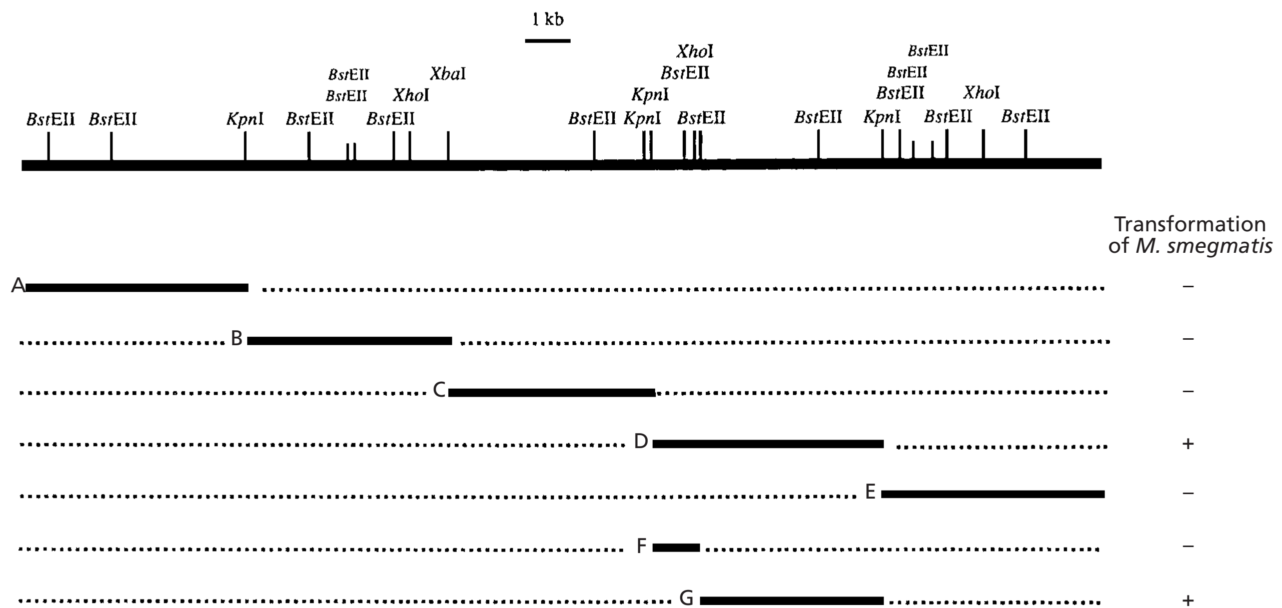

Fig. 1. Determination of the minimum region required for replication of pCLP in *M. smegmatis*. A restriction endonuclease map of pCLP is shown at the top of diagram. The approximate locations of the restriction sites are indicated. DNA fragments tested for autonomous replication are indicated by solid bars. Plasmids used for transformation are indicated in Table 1. The ability (+) or inability (−) of the clone to replicate in *M. smegmatis* mc^{2}155 is shown on the right.

pCLP (Fig. 1). Seven restriction fragments from 1 to 6 kb in size were subcloned into a kanamycin (Km)-resistant pUC19 derivative called pPV8 (Table 1). We also performed PCR assays using primers Poa (5′-ACC AAT GAG CAG TAA GCA GC-3′), Pob (5′-GCA GCA GCG ACA AAG ATG GG-3′) and Poc (5′-CAT CGG GCT GCG GGA AAC CC-3′). The PCR mixture and the amplification reactions were performed as previously described (Picardeau & Vincent, 1998). The products amplified using primers Poa/Poc and Pob/Poc (see Fig. 5) were inserted into the pGEM-T Easy vector (Promega) and the resulting inserts were released by *Eco*RI digestion and inserted into pPV8.

Plasmid constructs were introduced into *E. coli* DH5α by electroporation (Gene Pulser unit; Bio-Rad), selected on solid Luria-Bertani medium supplemented with 20 μg Km ml^{-1}, 2 mM IPTG and 0·004% X-Gal, and plasmids were recovered using a Qiagen Midi kit. Electrocompetent cells of *M. smegmatis* and *M. bovis* BCG were prepared as previously described (Pelicic *et al.*, 1996). Briefly, *M. smegmatis* and *M. bovis* BCG cultures were grown to exponential phase; pellets were washed three times in 10% glycerol and resuspended in 10% glycerol. The competent cells were electroporated in the presence of 2 μg vector DNA by using a Gene Pulser (Bio-Rad) and then transferred to 5 ml 7H9 Middlebrook liquid medium, in which they were incubated for 4 h (*M. smegmatis*) or 24 h (*M. bovis* BCG) at 37 °C before plating.

Plasmids in *M. smegmatis* and *M. bovis* BCG transformants were detected by Southern analysis and electroporation of *E. coli* competent cells with total genomic DNA of *M. smegmatis* or *M. bovis* BCG transformants.

DNA extraction and Southern analysis. Mycobacterial strains were grown in 5 ml Middlebrook 7H9 broth containing 1 mg D-cycloserine ml^{-1} and incubated overnight at 37 °C. Cells were heated for 20 min at 80 °C, centrifuged and resuspended in 250 μl 25% (w/v) sucrose/50 mM Tris pH 8/50 mM EDTA containing 500 μg lysozyme ml^{-1}. Incubation was continued for an extra night. Aliquots (250 μl) of a solution consisting of 100 mM Tris, 1% (w/v) SDS and 400 μg proteinase K ml^{-1} were added and the mixtures incubated for 4 h at 55 °C. DNA was extracted with phenol/chloroform (50:50, v/v), precipitated with absolute ethanol and dissolved in TE (10 mM Tris, 1 mM EDTA, pH 8).

Two micrograms of genomic DNA was digested with *Sac*I or *Pvu*II, subjected to electrophoresis overnight in a 1% agarose gel and transferred onto nylon membranes (Hybond-N+; Amersham). Membranes were hybridized overnight at 65 °C in Rapid hybridization buffer (Amersham) with pCL4D radiolabelled with [α-^{32}P]dCTP using a commercial kit (Megaprime; Amersham). The membranes were then washed as previously described (Picardeau & Vincent, 1997).

Stability of pCL4D and compatibility studies. *M. smegmatis* cells carrying one of the Km-resistant plasmids pCL4D or pB4 were grown in 5 ml Middlebrook 7H9 broth with no antibiotic selection pressure for 24 h at 37 °C. The culture was then diluted 1:100 and the bacteria grown in fresh antibiotic-free medium for a further 24 h; this procedure was repeated three times. After each dilution, the cells were plated on agar plates with and without Km and the proportion of resistant cells was taken as a measure of the number of cells carrying the plasmid. The same procedure was used for compatibility studies. To determine the number of plasmid-carrying cells, *M. smegmatis* mc^{2}155 transformants were grown on Middlebrook 7H10 agar plates containing 20 μg Km ml^{-1} and 20 μg streptomycin (Sm) ml^{-1} (for co-transformation with pJAZ42 plus pCL4D) or Km and hygromycin (Hg) (50 μg ml^{-1}) (for co-transformation with pB4 plus pBSh-D) for 4 d at 37 °C. Each experiment was repeated twice.

Plasmid copy number determination. The relative copy number of pCL4D in *M. smegmatis* mc^{2}155 was determined by single-cell resistance to Km as previously described

(Gavigan *et al.*, 1997; Stolt & Stoker, 1996). The minimum concentration of Km necessary to inhibit cell growth was taken as the single-cell resistance. In this experiment, *M. smegmatis* strain EP10 (Ainsa *et al.*, 1997) was used as a control because it contains only one copy of the Km resistance gene on its chromosome. *M. smegmatis* strain EP10 arose from the insertion of the gene from Tn903 confering resistance to Km into the *M. smegmatis* mc^{2}155 chromosome (Ainsa *et al.*, 1997).

Sequencing experiments. We sequenced double-stranded plasmid DNA by the dideoxy chain-termination method (Sanger *et al.*, 1977) using a *Taq* DyeDeoxy terminator cycle sequencing kit (Applied Biosystems), a model 9600 GenAmp PCR system (Perkin-Elmer) and a model 373 stretch DNA analysis system (Applied Biosystems). We used universal and reverse primers and a DNA walking strategy along the linear plasmid as an insert. Nucleotide sequences were analysed using the GCG software package (University of Wisconsin, Madison, WI, USA) and we searched for sequence similarities using the BLAST algorithm (Altschul *et al.*, 1997).

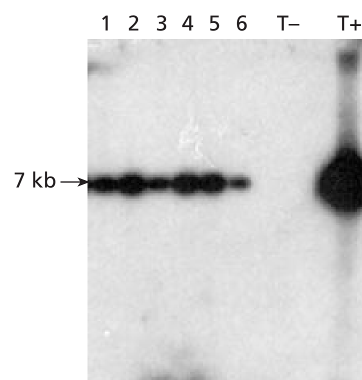

Fig. 2. Southern blot analysis of digested DNA of *M. smegmatis* mc^{2}155 transformed with pCL4D. DNA in all lanes was digested with *Sac*I and probed with pCL4D. Lanes: 1–6, individual colonies of *M. smegmatis* transformants; T−, untransformed *M. smegmatis* mc^{2}155; T+, original plasmid pCL4D. The arrow indicates the expected size for the linear pCL4D construct.

RESULTS

Identification of the internal region of the linear plasmid required for replication

We first constructed an endonuclease restriction map of pCLP by single and double digestion with the restriction enzymes *Bst*EII, *Kpn*I and *Xho*I (Fig. 1). A set of recombinant plasmids was constructed by inserting restriction fragments covering the complete sequence of pCLP into a pUC19 derivative containing the Km gene from Tn903 as a selective marker. These recombinant plasmids were used to transform *M. smegmatis*, to test their ability to replicate in mycobacteria. Transformation was also performed with pB4, which carries a known mycobacterial origin of replication from the *M. fortuitum* circular plasmid pAL5000 (Ranes *et al.*, 1990), as a control. The plasmid containing fragment D, designated pCL4D, replicated in *M. smegmatis*, but plasmids containing fragments A, B, C and E did not (Fig. 1). The efficiency of transformation of *M. smegmatis* with pCL4D was 2×10^4 bacteria (μg DNA)$^{-1}$. We then subcloned fragment D, using *Xho*I, to determine the minimum replication region. Deletion of the first kilobase of fragment D did not prevent plasmid replication (Fig. 1). Therefore, the region necessary for autonomous replication in *M. smegmatis* was located in a 4 kb fragment extending from the *Xho*I restriction site to the *Kpn*I site (Fig. 1). The banding patterns obtained by probing Southern blots of *M. smegmatis* transformant colonies with pCL4D demonstrated that pCL4D was neither integrated into the *M. smegmatis* chromosome nor in a linear form, but was instead present as an autonomously replicating circular plasmid (Fig. 2).

We tested whether the replication region of pCLP could replicate in slow-growing species by transforming *M. bovis* BCG with pCL4D, which resulted in around 10^2 transformants (μg DNA)$^{-1}$. Again, Southern blots of digested DNA from *M. bovis* BCG transformant colonies demonstrated that pCL4D was not integrated into the chromosome (Fig. 3a). Southern blots of non-digested DNA from *M. bovis* BCG transformant colonies revealed patterns that could correspond to the relaxed and supercoiled DNA (which migrates more rapidly than the relaxed counterpart) of pCL4D (Fig. 3b). DNA was also extracted from transformants and used to transform *E. coli*, which resulted in *E. coli* transformant colonies from which pCL4D-derived plasmids could be isolated (Fig. 3c), further indicating that pCL4D was present in *M. bovis* BCG as an autonomously replicating circular plasmid. However, plasmids from almost all the *M. bovis* BCG transformants displayed structural modifications, which included insertion and deletions (Fig. 3). For example, the recombinant plasmid extracted from a *M. bovis* BCG clone and used to transform *E. coli* resulted in a deletion of almost 3 kb (Fig. 3c).

Characteristics of the *E. coli*–mycobacteria shuttle vector pCL4D

The single-cell resistance to Km (Gavigan *et al.*, 1997) of *M. smegmatis* mc^{2}155 transformants carrying pCL4D gave a relative copy number for pCL4D of 1 (data not shown). After 4 d growth in an antibiotic-free medium, which corresponds to more than 40 generations, 85 % of *M. smegmatis* transformant cells contained pCL4D. This plasmid was therefore about as stable as pB4, the stability of which was also assessed in the absence of selection pressure (Fig. 4a). We tested whether our pCL4D construct was compatible with a pJAZ38-derived replicon and a pAL5000-derived replicon in *M. smegmatis*. We co-transformed *M. smegmatis* with pJAZ42 (a pJAZ38-derived plasmid expressing Sm resistance; Gavigan *et al.*, 1997) and pCL4D. *M. smegmatis* transformants [5×10^2 transformants (μg DNA)$^{-1}$] expressing both Km and Sm resistance were obtained. Every day, the *M. smegmatis* transformants were diluted in antibiotic-free medium, cultured for a

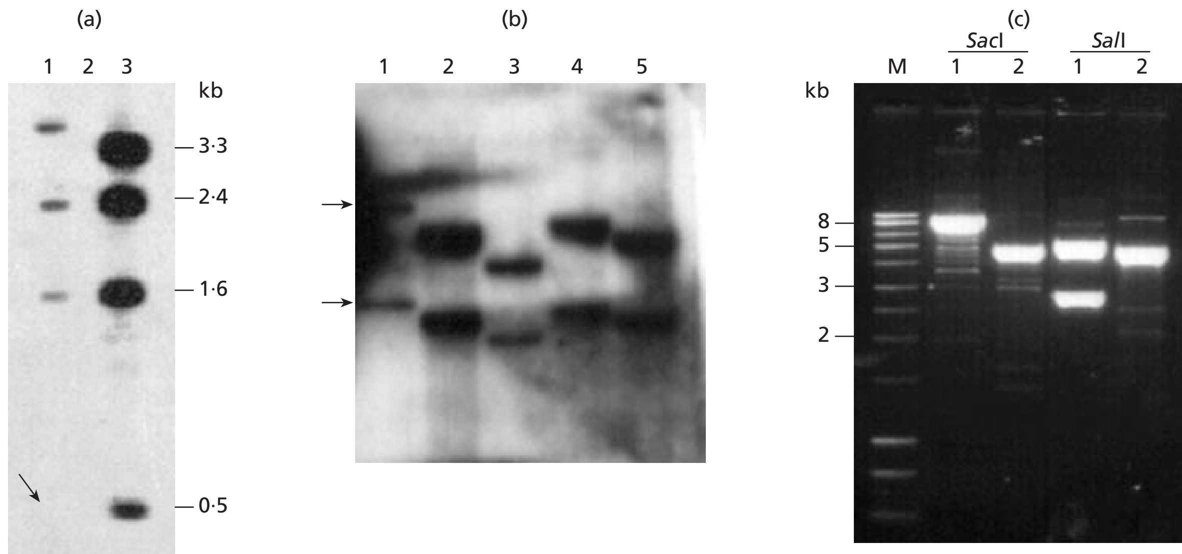

Fig. 3. Analysis of plasmids recovered from *M. bovis* BCG transformed with pCL4D. (a) Southern blot analysis of *Pvu*II-digested DNA of *M. bovis* BCG transformed with pCL4D. DNA in all lanes was digested with *Pvu*II and probed with pCL4D. Lanes: 1, *M. bovis* BCG transformant; 2, untransformed *M. bovis* BCG; 3, original plasmid pCL4D (expected sizes are indicated on the right). The weak band indicated by a small arrow is better visualized after a longer exposure. (b) Southern blot of non-digested DNA of *M. bovis* BCG transformants probed with pCL4D. Lanes 1–5, individual colonies of *M. bovis* BCG transformants. The arrows indicate the relaxed (top of the gel) and supercoiled (bottom of the gel) DNA of the pCL4D circular molecule. (c) Restriction of plasmid extracted from *E. coli* transformed with total DNA from one *M. bovis* BCG transformant. Lanes: 1, original plasmid pCL4D; 2, recombinant plasmid recovered from *E. coli*. DNA was digested with *Sac*I or *Sal*I. Lane M, molecular mass marker; sizes of the fragments are indicated on the left.

few generations and then tested for antibiotic resistance. Our results indicated that these two plasmids coexisted in *M. smegmatis* and were stable for at least 40 generations (Fig. 4b). In a similar experiment, we co-transformed *M. smegmatis* with a pCLP-derived plasmid and pB4 (containing the complete sequence of pAL5000 and expressing Km resistance; Ranes *et al.*, 1990). Fragment D of pCL4D was inserted into a pUC derivative vector containing a Hg resistance cassette (Table 1). We obtained 7×10^2 *M. smegmatis* transformants (μg DNA)$^{-1}$ expressing both Km and Hg resistance, even after 4 d growth (data not shown).

Sequencing of the replication region of pCLP and identification of a putative origin of replication

We determined the complete nucleotide sequence of the 4 kb fragment G of pCLP, which conferred autonomous replication in *M. smegmatis* (Fig. 1). The genetic organization of the origin of replication region of pCLP is shown schematically in Fig. 5. The nucleotide sequence contains two possible ORFs encoding proteins of 214 and 350 amino acids (Fig. 5). The deduced amino acid sequence of the first ORF was up to 40 % identical to homologues of the ParA partition protein of a wide variety of bacteria, including *Pseudomonas* spp. and *Agrobacterium tumefaciens*. The second putative ORF encodes a putative protein the carboxy-terminal region of which is about 25 % identical to the Rep protein of a circular plasmid from a *Rhodococcus* sp. and to the circular plasmid pLR7 from *M. avium*. The region of sequence identity between the putative Rep proteins encoded by pLR7 and pCLP includes the previously identified helix–turn–helix motif (Beggs *et al.*, 1995). Analysis of the replication region of pCLP also identified a 67 bp region more than 80 % identical to the nucleotide sequences of the replication regions of the mycobacterial circular plasmids pLR7, pJAZ38 and pMSC262; it includes a 20 bp sequence that is absolutely identical (except for 1 bp in pJAZ38) (Fig. 5). In all of these plasmids, the conserved region was located about 50 bp upstream from the start codon of the putative *rep* genes. No sequence similarity was detected between the replication region of pCLP and pAL5000 from *M. fortuitum*, which has been completely sequenced (Rauzier *et al.*, 1988). We found no consensus sequences typical of the origin of replication of bacterial circular plasmids (Del Solar *et al.*, 1998). However, the most striking feature of the replication region was the presence of an 18 bp sequence, 5′-TCC GAA ACC CGC TTA GCG-3′, as a direct repeat (Fig. 5). The overall GC content of the replication region was 66 %, similar to that of mycobacterial chromosomal DNA, but a 40 bp region with a GC content of 37 % was found in the vicinity of the 18 bp direct repeats, which may facilitate strand separation at the initiation site of replication. Sequence analysis of the pCLP replication region led to the identification of a *Sal*I restriction site located upstream from the candidate origin of replication (Fig. 5). This restriction site was used to construct a plasmid, pCL5, containing only the candidate origin of replication and the putative *rep* gene (Table 1, Fig. 5). This

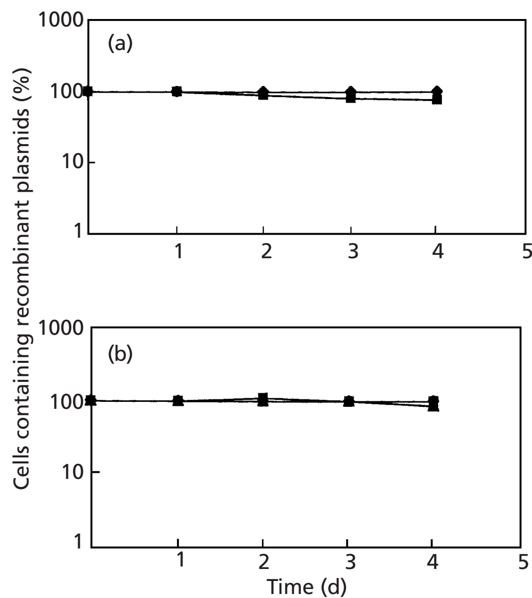

Fig. 4. Stability and compatibility studies for the pCLP-derivative pCL4D in *M. smegmatis* mc^{2}155. (a) Stability of pCL4D (■) versus that of the pAL5000-derivative pB4 (◆). Cells carrying the constructs were grown without selection. (b) Compatibility of pCL4D with pJAZ42. Cells carrying the two constructs were grown without selection (◆) or with Km (■), Sm (▲), or Km plus Sm (○).

construct replicated in *M. smegmatis* and *M. bovis* BCG, suggesting that the region containing the putative *parA* gene is not necessary for replication. We defined the location of the origin of replication of pCLP by constructing plasmids with (plasmid pCRa) and without (plasmid pCRb) the candidate origin of replication (Table 1, Fig. 5). *M. smegmatis* transformants were obtained with pCRa, which contains the minimum replication region of pCL5 (Fig. 5). However, no Km-resistant transformants were obtained with pCRb, which contains the region conserved in other mycobacterial replicons and the putative *rep* gene (Fig. 5). Therefore, the deleted fragment, which contains a non-coding sequence including the 18 bp direct repeat and the AT-rich region, may be the origin of replication of pCLP.

DISCUSSION

We have previously demonstrated that mycobacterial linear plasmids have an invertron-like structure similar to that of other replicons in actinomycetes (Picardeau & Vincent, 1998). In this study, we identified the origin of replication of pCLP, which is located in a region about one third of the way along the sequence. The replication region of this linear plasmid had several features in common with those of circular plasmids. Firstly, pCL4D, which contains the internal origin of replication of pCLP, was stably maintained in mycobacteria as a circular plasmid. Previous studies of linear chromosomes and linear plasmids in *Streptomyces* spp., which also have a terminal protein covalently bound to each 5′ end, have shown that these linear DNA structures initiate replication from a centrally located origin and replicate after circularization (Chang & Cohen, 1994; Fischer *et al*., 1998; Musialowski *et al*., 1994; Zakrzewska-Czerwinska & Schrempf, 1992). In contrast, the linear replicons with an invertron-like structure found in organisms other than actinomycetes, such as the *Bacillus subtilis* phage *ϕ*29, initiate the replication of full-length DNA strands at the telomeres by a protein-primed strand-displacement mechanism (Salas, 1991). The similar terminal structure of the linear plasmids in actinomycetes suggests that, like the linear replicons of *Streptomyces* spp., linear plasmids in mycobacteria may have another replication mechanism, in addition to the internal replication origin, for filling in the ends of lagging-strand DNA (Qin & Cohen, 1998). The termini of pCLP contain several palindromes (Picardeau & Vincent, 1998) that may be involved in such a mechanism. The structural features of the replication region of pCLP are similar to those of circular plasmids. Indeed, homologues of bacterial circular DNA partition and replication proteins were found in the replication region of pCLP. We identified a possible ParA homologue, which appears not to be essential for the replication of pCLP. The ParA and ParB proteins encoded by the P1 plasmid are involved in the partitioning of stable low-copy-number plasmids. In the P1 partition cycle, ParB binds to an AT-rich sequence, *parS*, to form a complex that interacts with ParA, stimulating the ATPase activity of the Par A (Davis *et al*., 1992). All three elements, *parA*/*parB*/*parS*, are absolutely essential for plasmid stability. The low copy number of pCL4D in *M. smegmatis*, and the loss of pCL4D by 15 % of *M. smegmatis* transformant cells after 40 generations, may be due to the presence of an incomplete partitioning system, consisting of only the ParA homologue. ParA homologues have also been found in the candidate origins of replication of *Borrelia burgdorferi* linear and circular plasmids (M. Picardeau, J. R. Lobry & B. Hinnebusch, unpublished results). The other putative ORF in the replication region of pCLP encodes a putative protein that is a homologue of the Rep protein of the circular plasmids of other actinomycetes. Subcloning experiments with pCL4D suggest that this ORF is necessary for replication. Rep-like proteins have in common a helix–turn–helix motif, typical of DNA-binding proteins (Brennan & Matthews, 1989), and may therefore be involved in the initiation of replication. A region was identified that was almost completely conserved, upstream from the putative *rep* genes of the linear plasmid pCLP and mycobacterial circular plasmids pLR7 from *M. avium*, pJAZ38 from *M. fortuitum* and pMSC262 from *M. scrofulaceum*. As pCL4D and pJAZ38, both containing this conserved region, are compatible in *M. smegmatis*, this sequence is presumably not involved in the incompatibility process. In bacterial circular plasmids, the origins of replication usually contain directly repeated sequences that act as binding sites for the Rep protein and AT-rich regions, facilitating strand separation (Del Solar *et al*., 1998).

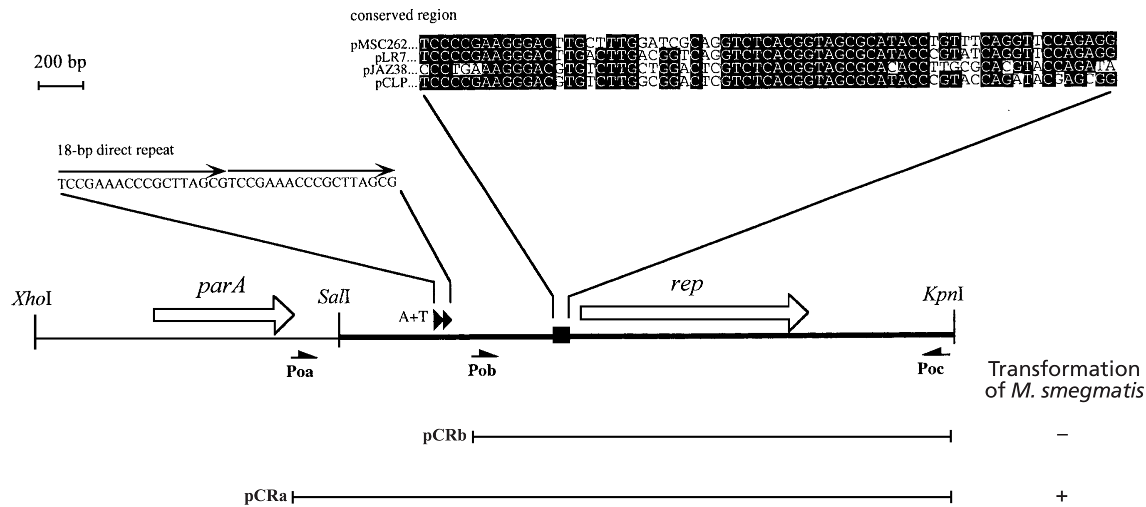

Fig. 5. Genetic organization of the replication region of pCLP. The putative genes *par*A and *rep* and their transcription orientation are indicated by open arrows. Arrowheads indicate the 18 bp repeat sequences. 'A+T' indicates the AT-rich segment of the replication region. The line in bold type corresponds to the minimum replication region of pCLP (pCL5). The location of primers Poa, Pob and Poc are indicated. The amplified product Poa/Poc (pCRa) facilitated plasmid replication in *M. smegmatis*, but the amplified product Pob/Poc (pCRb) prevented plasmid replication.

The region conserved in pCLP, pLR7, pJAZ38 and pMSC262 contains no repeats or AT-rich regions. Its location, upstream from the putative *rep* gene, suggests that this region may have control properties. A candidate origin of replication was found between the putative *parA* and *rep* genes of pCLP; it contains direct repeats, which may act as a Rep-binding site, and in its vicinity there are AT-rich segments. The replication regions of the linear plasmids pSCL and pSLA2 of *Streptomyces* spp. also contain long juxtaposed direct repeats (Chang *et al.*, 1996). Cloning experiments in which the candidate origin of replication of pCLP was not used, but the conserved region upstream from the putative *rep* gene was, abolished plasmid replication, further suggesting that the region containing the 18 bp direct repeat and the AT-rich segments is the origin of replication and is required with the Rep protein for plasmid replication.

The replication region of the mycobacterial linear plasmid is more similar to that of mycobacterial circular plasmids than to that of other linear replicons of *Streptomyces* spp. This suggests that the linear plasmids in actinomycetes may not have a common ancestor, but that they may have arisen from several independent events, such as genetic exchanges with bacteriophages or fungi (Hinnebusch & Tilly, 1993), resulting in genetic entities with similar terminal structures.

Several of our results suggest that pCL4D may be useful as a shuttle vector for genetic studies in mycobacteria. (i) We obtained a high efficiency of transformation with pCL4D in *M. smegmatis* and *M. bovis* BCG. However, DNA rearrangements were observed in plasmids extracted from *M. bovis* BCG transformants and never in *M. smegmatis*. Although the restriction patterns of these plasmids showed that they underwent rearrangements, pCL4D-derived plasmids are still able to replicate in mycobacteria. This phenomenon of plasmid structural instability has previously been observed for pAL5000-derived plasmids expressing heterologous DNA in *M. bovis* BCG (Haeseleer, 1994). Restriction analysis and sequencing of the rearranged plasmids will allow determination of the mechanisms of these structural modifications. This may help in constructing pCL4D-derived plasmids with more stable structures in *M. bovis* BCG. (ii) We found that this pCLP-derived plasmid can co-exist in *M. smegmatis* with pJAZ38- and pAL5000-derived plasmids, which have been shown to be compatible (Gavigan *et al.*, 1997). This finding suggests that the three plasmids belong to three different compatibility groups. (iii) Previous studies have shown that pMSC262 from *M. scrofulaceum* and pLR7 from *M. avium* cannot transform the hypertransformable mutant strain, *M. smegmatis* mc^{2}155 (Beggs *et al.*, 1995; Qin *et al.*, 1994). Therefore, our plasmid, which replicated in both slow- and fast-growing species, may be useful for the introduction of genes into various species of mycobacteria. Thus, for example, pCL4D may be used in experiments requiring more than one plasmid together with plasmids derived from pAL5000, the most studied mycobacterial plasmid which has been extensively used for the construction of vectors.

In conclusion, as previously suggested, linear plasmids in actinomycetes may be evolutionary intermediates between circular plasmids and linear phage replicons (Shiffman & Cohen, 1992). Future studies of linear plasmids in actinomycetes should focus on the characterization of their terminal proteins and investigation of their function. These terminal proteins have not been characterized in bacteria and their function is unknown; they may be part of the replication machinery and/or

they may protect the extremities from exonuclease activity. Intergeneric transfer of linear plasmids has never been demonstrated in mycobacteria, but may be possible between mycobacteria, *Streptomyces* spp. and *Rhodococcus* spp., in which linear plasmids have also been found (Picardeau & Vincent, 1998). Indeed, in *Streptomyces* spp., almost all the extensively studied linear replicons have been found to be conjugative plasmids. Further studies are required to define the role of these linear replicons in the spread of genes in the natural habitats of actinomycetes.

ACKNOWLEDGEMENTS

We thank C. Guilhot for plasmid pB4, C. Martin for plasmid pJAZ42 and strain EP10, and V. Pelicic for plasmid pPV8. We would also like to thank J. Rauzier for help with sequencing. M.P. is a recipient of a fellowship from Fondation Roux (Institut Pasteur). This work is part of the doctoral thesis in Microbiology of C.L.D.

REFERENCES

Ainsa, J. A., Perez, E., Pelicic, V., Berthet, F. X., Gicquel, B. & Martin, C. (1997). Aminoglycoside 2′-*N*-acetyltransferase genes are universally present in mycobacteria: characterization of the *aac*(2′)-*lc* gene from *Mycobacterium tuberculosis* and the *aac*(2′)-*ld* gene from *Mycobacterium smegmatis*. *Mol Microbiol* **24**, 431–441.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Beggs, M. L., Crawford, J. T. & Eisenach, K. D. (1995). Isolation and sequencing of the replication region of *Mycobacterium avium* plasmid pLR7. *J Bacteriol* **17**, 4836–4840.

Brennan, R. G. & Matthews, B. W. (1989). The helix–turn–helix DNA binding motif. *J Biol Chem* **264**, 1903–1906.

Chang, P. C. & Cohen, S. N. (1994). Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science* **265**, 952–954.

Chang, P. C., Kim, E. S. & Cohen, S. N. (1996). *Streptomyces* linear plasmids that contain a phage-like, centrally located, replication origin. *Mol Microbiol* **22**, 789–800.

Crespi, M., Messens, E., Caplan, A. B., van Montagu, M. & Desomer, J. (1992). Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. *EMBO J* **11**, 795–804.

Dabrock, B., Kebeler, M., Averhoff, B. & Gottschalk, G. (1994). Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. *Appl Environ Microbiol* **60**, 853–860.

Davis, M. A., Martin, K. A. & Austin, S. J. (1992). Biochemical activities of the ParA partition protein of the P1 plasmid. *Mol Microbiol* **6**, 1141–1147.

Del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**, 434–464.

Fischer, G., Holl, A. C., Volff, J. N., Vandewiele, D., Decaris, B. & Leblond, P. (1998). Replication of the linear chromosomal DNA from the centrally located *oriC* of *Streptomyces ambofaciens* revealed by PFGE gene dosage analysis. *Res Microbiol* **149**, 203–210.

Gavigan, J. A., Ainsa, J. A., Perez, E., Otal, I. & Martin., C. (1997). Isolation by genetic labeling of a new mycobacterial plasmid, pJAZ38, from *Mycobacterium fortuitum*. *J Bacteriol* **179**, 4115–4122.

Haeseleer, F. (1994). Structural instability of recombinant plasmids in mycobacteria. *Res Microbiol* **145**, 683–687.

Hanahan, S. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Hinnebusch, J. & Tilly, K. (1993). Linear plasmids and chromosomes in bacteria. *Mol Microbiol* **10**, 917–922.

Kalkus, J., Dorrie, C., Fischer, D., Reh, M. & Schegel, H. G. (1993). The giant linear plasmid pHG207 from *Rhodococcus* sp. encoding hydrogen autotrophy: characterization of the plasmid and its termini. *J Gen Microbiol* **139**, 2055–2065.

Kesseler, M., Dabbs, E. R., Averhoff, B. & Gottschalk, G. (1996). Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. *Microbiology* **142**, 3241–3251.

Kosono, S., Maeda, M., Fuji, F., Arai, H. & Kudo, T. (1997). Three of the seven *bphC* genes of *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem, are located on an indigenous plasmid associated with biphenyl degradation. *Appl Environ Microbiol* **63**, 3282–3285.

Musialowski, M. S., Flett, F., Scott, G. B., Hobbs, G., Smith, C. & Oliver, S. G. (1994). Functional evidence that the prinicipal DNA replication origin of the *Streptomyces coelicolor* chromosome is close to the *dnaA–gyrB* region. *J Bacteriol* **176**, 5123–5125.

Pelicic, V., Reyrat, J. M. & Gicquel, B. (1996). Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. *Mol Microbiol* **20**, 919–925.

Picardeau, M. & Vincent, V. (1997). Characterization of large linear plasmids in mycobacteria. *J Bacteriol* **179**, 2753–2756.

Picardeau, M. & Vincent, V. (1998). Mycobacterial linear plasmids have an invertron-like structure related to other linear replicons in actinomycetes. *Microbiology* **144**, 1981–1988.

Polo, S., Guerini, O., Sosio, M. & Deho, G. (1998). Identification of two linear plasmids in the actinomycete *Planobispora rosea*. *Microbiology* **144**, 2819–2825.

Qin, M., Taniguchi, H. & Mizuguchi, Y. (1994). Analysis of the replication region of a mycobacterial plasmid, pMSC262. *J Bacteriol* **176**, 419–425.

Qin, Z. & Cohen, S. N. (1998). Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. *Mol Microbiol* **28**, 893–903.

Ranes, M. G., Rauzier, J., Lagranderie, M., Gheorghiu, M. & Gicquel, B. (1990). Functional analysis of pAL5000, a plasmid from *Mycobacterium fortuitum*: construction of a 'mini' *Mycobacterium–Escherichia coli* shuttle vector. *J Bacteriol* **172**, 2793–2797.

Rauzier, J., Moniz-Pereira, J. & Gicquel-Sanzey, B. (1988). Complete nucleotide sequence of pAL5000, a plasmid from *Mycobacterium fortuitum*. *Gene* **71**, 315–321.

Sakaguchi, K. (1990). Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adenotype viruses. *Microbiol Rev* **54**, 66–74.

Salas, M. (1991). Protein-priming of DNA replication. *Annu Rev Biochem* **60**, 39–71.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing

with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.

Shiffman, D. & Cohen, S. N. (1992). Reconstruction of a *Streptomyces* linear replicon from separately cloned DNA fragments: existence of a cryptic origin of circular replication within the linear plasmid. *Proc Natl Acad Sci USA* **89**, 6129–6133.

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R., Jr (1990). Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**, 1911–1919.

Stolt, P. & Stoker, N. G. (1996). Functional definition of regions necessary for replication and incompatibility in the *Mycobacterium fortuitum* plasmid pAL5000. *Microbiology* **142**, 2795–2802.

Zakrzewska-Czerwinska, J. & Schrempf, H. (1992). Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. *J Bacteriol* **147**, 2688–2693.

Received 9 August 1999; revised 1 November 1999; accepted 3 November 1999.