Protein-DNA Interactions in the *ori* Region of the *Mycobacterium fortuitum* Plasmid pAL5000

PELLE STOLT AND NEIL G. STOKER*

Bacterial Molecular Genetics Unit, Department of Clinical Sciences, London School of Hygiene & Tropical Medicine, London WC1E 7HT, England

Received 17 June 1996/Accepted 30 September 1996

Plasmid pAL5000 from *Mycobacterium fortuitum* encodes two proteins necessary for replication: RepA (307 amino acid residues) and RepB (119 residues). A single RNA species encoding these proteins was characterized, and its 5' end was defined. The proteins were expressed as maltose-binding protein fusions in *Escherichia coli*. The RepB protein was shown in vitro to bind specifically to a previously defined 435-bp region of pAL5000 containing the origin of replication (*ori*). The precise RepB binding sites were defined by DNase I footprinting experiments. RepB binds to two motifs in the *ori* region: a high-affinity site within its own promoter region, implying autoregulation of its expression, and a low-affinity site further upstream, presumably the origin of replication itself. The binding to the latter motif seems to occur on one DNA strand only. The high-affinity binding site contains several palindromic sequences. Gel retardation assays were performed with the different binding sites as templates, and the binding constant to each site was estimated from protein titrations. This is the first molecular dissection of mycobacterial DNA-binding proteins and their interactions with their targets.

The mechanisms governing plasmid replication and maintenance in mycobacteria are little understood. Plasmids have been described in several species, e.g., *Mycobacterium fortuitum*, *Mycobacterium scrofulaceum*, *Mycobacterium chelonae*, and *Mycobacterium avium* (4, 8, 17, 18, 23, 29). No plasmids have been found in the pathogen *Mycobacterium tuberculosis*, but it is possible to transform *M. tuberculosis* with plasmids from related mycobacteria, showing that the lack of plasmids is not due to an intrinsic inability to support their replication.

The replicon most used in genetic manipulation of mycobacteria is the 4,837-bp *M. fortuitum* plasmid pAL5000 (19) (see Fig. 1) (GenBank accession number M23557). This replicon is the basis of several shuttle plasmids and cosmids used to manipulate DNA in a wide variety of mycobacteria, including the fast-growing *Mycobacterium smegmatis* and the slow growers *Mycobacterium bovis* BCG and *M. tuberculosis* (9, 31, 36, 41). We have previously shown that the origin of replication (*ori*) of this plasmid lies within a 435-bp segment and comprises a region with a high incidence of repeated sequences (39). In addition, it has been shown that two open reading frames (ORFs) are necessary for a functional replicon (22, 39). The products of these ORFs, RepA and RepB, will act in *trans* to activate the *ori* if this region is present on another plasmid (39).

The *ori* region is also able to confer incompatibility to otherwise unrelated replicons. Whether the same sequence motifs within the region are responsible for the *inc* and *ori* functions or whether *inc* is physically separated from *ori* as in plasmid P1 (1, 5, 7) is at present unclear.

The RepA protein (307 amino acid [aa]) has a high similarity to replication proteins from ColE2-related plasmids (12, 39). RepB is a small protein (119 aa) which has recently been shown to have similarity to an ORF product encoded by the *Bifidobacterium longum* plasmid pMB1 (34). The *repA* and *repB* genes overlap by 1 bp, suggesting that they are transcribed as an operon.

The observations that the *ori* region has several directly repeated sequence motifs, that the incompatibility effect is orientation dependent, and that the region can be activated in *trans* (39) make this region a prime candidate for protein-binding sites related to the replication of pAL5000.

This paper presents a closer study of RepB and the sites of its interaction with *ori*. We determined the transcription start point for the RNA encoding both replication proteins and expressed them as fusion proteins in *Escherichia coli*. Using gel retardation assays and DNase I footprinting techniques we could show that RepB binds with different affinity to two sites in the *ori* region, one of which is its own promoter. Thus, RepB has a role in the regulation of its own expression as well as in the initiation of replication of pAL5000. RepA, despite its similarities to replication proteins from other plasmids, did not bind DNA in any of our assays.

MATERIALS AND METHODS

Bacterial strains and media. *E. coli* DH5 α (11) was used for routine manipulations of plasmid DNA. For the expression of RepA and RepB as fusion proteins to maltose-binding protein (MBP), *E. coli* JM109 (45) was used. *E. coli* cells were grown in TY medium (16 g of tryptone, 10 g of yeast extract, 5 g of NaCl per liter) with the addition of kanamycin or ampicillin (50 μ g/ml each) or chloramphenicol (40 μ g/ml) as appropriate. *M. smegmatis* mc²155 (37) was used throughout and was grown in Lemco medium (Difco) or on Lemco agar plates.

Transcription analysis. For reverse transcription-PCR, 2 μ g of RNA from either wild-type *M. smegmatis* or cells carrying the pAL5000-derived shuttle vector pYUB12 (36) were precipitated with 1 pmol of oligonucleotide and taken up in a total volume of 20 μ l of murine leukemia virus reverse transcriptase buffer (United States Biochemical [USB]). Two hundred units of murine leukemia virus reverse transcriptase was added, and the reaction was allowed to proceed for 20 min. The reaction mixture was treated with RNase I, and 1/10 of the product was used in subsequent PCRs. Oligonucleotides employed were ORF1D (5' TGGCGCGATTGCCACGATC 3'), ORF1E (5' GCGCGATATCGAGCCGAGAAC 3'), ORF1B2 (5' TCTACTAGGATCCGTGAGCCACGTCG 3'), OYUB7 (5' AGATGCAGCCCCGAAATG 3'), and OGS1 (5' TCTCAGCGGCCGGAACGT 3'). The products were analyzed on agarose gels. S1 nuclease protection assays were carried out as described previously (40) using S1 nuclease from Promega. The oligonucleotides used in generating the probes were ORF1C (5' TCAGATATTTGCTGAGGGG 3'), ORF1D, OYUB5 (5' TTGTGGTCGCTGTGCGC 3'), and OYUB9 (5' GTTGGGTTCGCCTCGATGTA 3'). Ten micrograms of RNA was used in each S1 nuclease assay.

Expression of the RepA and RepB proteins. The products of pAL5000 ORF1 (RepA) and ORF2 (RepB) were expressed in *E. coli* as MBP fusion proteins. The DNA was amplified in PCRs using the *Pyrococcus furiosus* (Pfu) DNA

* Corresponding author. Mailing address: Bacterial Molecular Genetics Unit, Department of Clinical Sciences, London School of Hygiene & Tropical Medicine, Keppel St., London WC1E 7HT, England. Phone: 44-(0)171-927 2425. Fax: 44-(0)171-637 4314.

FIG. 1. Schematic map of pAL5000 as the plasmid DNA is present in the shuttle vector pYUB12 (36), used in most manipulations described here. The ORFs are shaded; ORF4 is split by the cloning procedure. The arrow indicates the *rep* transcription unit. The positions of oligonucleotides used in the transcription studies are shown as arrowheads. The numbering in the scale bar refers to the pAL5000 sequence in GenBank (accession number M23557). The origin of replication (*ori*) is marked by a dashed line. pUH52, the smallest replicon defined to date to be functional in *M. smegmatis* (39), is shown above the sequence.

polymerase (Stratagene). The oligonucleotides used in the reactions to generate the fragments inserted in the pMalEII vector (New England Biolabs) were ORF1B2 (5' TCTACTAGGATCCGTGAGCCACGTCG 3'), ORF1B (5' GCTGCCGTTCTAGAGGTCGCTGTAG 3'), ORF2E (5' ATGGAGGCATTGCTATGAGCGACGG 3'), and ORF2B (5' CGACACCGGATCCCCAATTGCGTTA 3'). The recombinant DNA was sequenced with the Sequenase kit (United States Biochemicals). Expression and affinity purification of the recombinant proteins over an amylose matrix were performed according to the manufacturer's instructions. In the DNA-binding experiments, the recombinant protein was used either as a purified fusion protein or after removal of the MBP moiety by treatment overnight at room temperature with factor X protease (New England Biolabs; 1 U/50 µg of recombinant protein).

DNA-binding assays. Gel retardation assays were carried out using the 435-bp *ori* fragment comprising the pAL5000 origin of replication (39). The fragment, purified from a pUC18 clone, was radioactively labelled with [α - 32 P]dATP (Amersham) and the Klenow fragment of DNA polymerase I (Promega). Recombinant RepA-MBP fusion or RepB-MBP fusion (1 to 2 µg) was mixed with 10 ng of DNA in binding buffer (100 mM Tris-HCl [pH 8.0], 10 mM dithiothreitol, 20 mM MgCl₂, 1 µg of salmon sperm DNA per reaction) and incubated at 37°C for 15 min. The complexes were separated by electrophoresis on a 6% nondenaturing polyacrylamide gel and visualized by exposing the dried gel to X-ray film.

To generate templates for gel retardation assays specific for the high-affinity binding site (H site) and the low-affinity binding site (L site), PCRs were performed. For H-site templates the pair OKDH (5' CGTGTCCGACCATACACCGGTGATTAA 3') and OFP1 (5' GAGCAGATCGTCGCTTGCCA 3') was used; the pair OKDL (5' GCGGGGGAGTGTGCAGTTGTGG 3') and OKDB (5' ACGAGCTCCAAGTCAGATAT 3') was employed in generating L-site templates. The forward oligonucleotide (OKDH and OKDL) was labelled with [γ - 32 P]ATP and T4 PNK polynucleotide kinase (Promega). The autoradiograms were scanned with a Sharp JX-330P scanner and analyzed with Phoretix 1D Version 2.01 software (NonLinear Dynamics Ltd.).

DNase I footprinting was carried out essentially as described earlier (21). Templates were generated in PCRs with oligonucleotide primers 32 P labelled at the 5' end with [γ - 32 P]ATP and T4 PNK. Oligonucleotides OGS1, ORF1D, OYUB9, and OFP1 were employed. RepB protein in increasing dilutions (270 to 0.27 nM) was incubated with approximately 50 ng (0.25 pmol) of the labelled template for 10 min before the addition of DNase I. The mixture was incubated for 10 min at room temperature, and the products were analyzed on a 6% denaturing polyacrylamide gel. As a size ladder, M13mp18 was sequenced with the -40 universal primer with the Sequenase kit (USB).

RESULTS

Determination of the 5' end of the *repA* and *repB* mRNA. It has previously been shown that the regions encoding RepA and RepB are necessary for the replication of pAL5000 (22, 39). However, no transcription studies have been performed, and the start point for RepA has only been inferred from DNA sequence inspections (32). Indeed, the possibility that RepA and RepB are not two distinct protein species but are instead

one protein generated by ribosomal slipping had not been ruled out.

Evidence for the presence of two proteins was gathered by coelectroporation of *M. smegmatis* with the two plasmids pUH36 (39) and pYUB285 (22). Neither of these can replicate on its own; pUH36 lacks the *repB* gene, and pYUB285 has a deletion in *repA*. In cotransformation experiments, however, both plasmids replicated in the transformants. This strongly supports the existence of two distinct protein species which can both act in *trans*.

Prior to expressing the RepA protein, the start point of transcription for the *repA* mRNA was determined. Since mycobacteria are G+C-rich organisms, primer extension techniques were not used, as the reverse transcriptase is inclined to break off transcription prematurely at stable secondary structures (33, 38). We therefore used S1 nuclease protection techniques, which are less susceptible to these problems.

To narrow down the area for the fine-detail mapping, we first performed reverse transcription-PCR using oligonucleotide ORF1D for the reverse transcription reaction and oligonucleotides ORF1E, OGS1, and ORF1B2 for the subsequent PCR. The regions where these primers map on the pAL5000 sequence are shown schematically in Fig. 1. With this method (not shown), we found no bands using oligonucleotides ORF1E and OGS1 but found a signal with ORF1B2, indicating that the start point of transcription is downstream of bp 4420 in the pAL5000 sequence. We also obtained evidence that there is a common transcription unit for *repA* and *repB*, using the oligonucleotide OYUB7 for the reverse transcription reaction and ORF1B2 for the PCR (not shown).

Performing S1 protection assays, we were able to detect one single signal attributable to a 5' end of an mRNA. This was found in the experiment in which oligonucleotide ORF1D was used to generate the probe. The experiment (Fig. 2) defined the start point of the *rep* mRNA at the nucleotide C number 4632 in the pAL5000 sequence (Fig. 3). This is 85 bp downstream of the start point for the RepA ORF suggested earlier (32). No signals attributable to 5' ends were detected with probes derived from oligonucleotide OYUB5, OTF1C, or OYUB9, indicating that the single 5' end is not a processing site from an RNA species starting further upstream.

FIG. 2. S1 nuclease protection assay defining the 5' end of the *rep* RNA. The sequence is M13mp18 sequenced with the -40 universal primer (USB). The probe was generated with oligonucleotide ORF1D (Fig. 1). Lane 1, RNA from *M. smegmatis* cells carrying pYUB12; lane 2, RNA from *M. smegmatis* cells without plasmid. The arrow marks the 251-bp fragment protected from nuclease cleavage.

Expression of RepA and RepB in *E. coli* as MBP fusion proteins. From the transcription analysis described above it was clear that the start point of translation for RepA could not be upstream of base 4632. From comparison with related Rep proteins (12, 39), the most probable start point would be on the GTG codon at position 4792. There is a putative GAGG Shine-Dalgarno motif upstream of this start, at positions 4566 to 4569 (Fig. 3). The predicted size of the product would be 28 kDa, the same size as the protein most closely related to RepA, the Rep protein from the *Corynebacterium glutamicum* plasmid pXZ10142 (GenBank accession number X72691).

However, another possibility would be the GTG at position 4633, just 1 base from the start of the mRNA. A protein translated from this codon would have a size of 34 kDa, which is indeed close to many of the related *E. coli* replication proteins. To make certain that we did not express a truncated protein, we decided to express the larger ORF between bases 4633 and 722 as RepA and the ORF between bases 719 and 1078 as RepB. The predicted masses of the mycobacterial part of the fusion proteins would be 34 and 13.5 kDa, respectively.

The apparent masses of the purified fusion proteins on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were 80 and 55 kDa (not shown), which conformed to expectation, since the MBP moiety is 42.7 kDa. Digesting the RepA-MBP fusion protein with factor X gave rise to several bands on SDS-PAGE (not shown), in spite of the fact that

FIG. 3. DNA sequence of the pAL500 *ori* region. The start point of the *rep* mRNA is shown as an arrowhead. Oligonucleotide sequences are underlined. The binding sites for RepB are in boldfaced italics. Palindromic sequences within the high-affinity protein-binding site are boxed. Arrows above the sequence mark repeats. The numbering of bases is that from the pAL500 sequence (32). The start codon for ORF1 (arrowhead) is the first GTG encountered and the one used in expressing the product in *E. coli*; the more probable start for RepA, from the data presented here and in reference 39, is given as well.

FIG. 4. Gel retardation assay using the RepB-MBP fusion protein with the ^{32}P -labelled 435-bp *ori* fragment (10 ng) as a template. Lane 1, template only, with no protein; lane 2, RepB-MBP (2 μg); lane 3, competitive inhibition of binding through the addition of 250 ng of unlabelled *ori* DNA. All binding assays contained 1 μg of nonspecific DNA.

there should be no factor X recognition site within RepA. Hence, all manipulations with RepA were carried out with the intact fusion protein. No such problems were encountered with RepB.

DNA-binding assays. To test the recombinant proteins for DNA-binding activity, we used as a probe the 435-bp *ori* segment described previously (39). This fragment was end labelled with ^{32}P and used in gel retardation assays. In this region there are two pairs of directly repeated sequence motifs (Fig. 3). One of these pairs (bp 4430 to 4435 and 4441 to 4446) has a periodicity of 11 bp, and thus, the two motifs would be on the same side of the DNA helix, which made them potential sites of binding by one or both the replication proteins.

Unexpectedly, the RepA fusion protein, despite its similarity to other replication proteins, had no DNA-binding ability at all under a range of conditions, including varied pH (pH 5 to 9) and in the presence or absence of ATP or different cations (Mg^{2+} , Ca^{2+} , Zn^{2+} , Mn^{2+} , and K^{+}). We also tested RepA for possible nicking or topoisomerase-like activities, but half-hour incubations of supercoiled plasmid DNA with RepA had no effect on the topology of the plasmid (not shown).

In contrast, RepB specifically bound to the *ori* region (Fig. 4). It was possible to show specificity of binding, as we could block out the bound DNA with unlabelled *ori* DNA (Fig. 4, lane 3), but the addition of several micrograms of nonspecific DNA had no influence on the binding. In all binding experiments using the 435-bp *ori* region as a template, we observed one band which seemed to arise from nonspecifically bound template. This nonspecific band remained in the same position when unlabelled *ori* DNA was added.

Boiling the RepB sample before adding the template abolished binding. Mixing RepB with RepA before or after the addition of template had no detectable effect on the DNA-binding properties (not shown).

DNase I footprinting of RepB on the pAL5000 origin of replication. To determine the exact sequence where the RepB protein binds to *ori*, we performed DNase I footprinting experiments. The templates used were generated in PCRs with oligonucleotide OGS1, together with either OYUB9 or OFP1 (Fig. 3), in which one of the oligonucleotides was labelled with ^{32}P , yielding templates specifically labelled on one strand. The experiments were performed either with the RepB-MBP fusion protein or after the MBP moiety had been removed by treatment with factor X.

The results of these footprinting experiments are shown in

Fig. 5. There are two regions where bases are protected from DNase I cleavage by bound RepB protein. The protected regions correspond to the binding sites shown in Fig. 3. The footprinting gels suggest that RepB binds with different affinities to the two different regions. It is notable that the protein does not bind to the repeated sequences suspected from the earlier functional studies to be binding sites (39). However, in the H site, there are palindromic or repeated sequences (boxed or marked with arrows in Fig. 3) which could be recognition sequences for RepB. The protected regions of both binding sites are staggered, but there is no sequence similarity between the H and L sites.

The H site between bases 4590 and 4623 is only 9 bp upstream of the start point for *rep* transcription (Fig. 3). This implicates RepB in regulation of its own expression.

In the case of the L site between bp 4534 and 4564, the RepB-MBP fusion protein showed a footprint with protected bases on both sides of the helix, but the RepB protein alone, used after factor X digestion, showed a much stronger binding to the DNA strand corresponding to the sense strand (compare lanes 1 and 2 in Fig. 5a with lanes 7 and 8 in Fig. 5b). From these experiments we conclude that the binding of RepB to this site is predominantly to one side of the DNA helix. No such influence of the MBP moiety on binding was observed with the H site (not shown).

As a negative control, we repeated the footprinting experiments with purified *E. coli* MBP and also with purified *E. coli* LexA protein. Neither bound to any of the pAL5000 *ori* templates employed (not shown).

Determination of binding constants to the H site and the L site. Based on the data from the DNase I footprinting experiments, new probes for gel retardation assays specific for the H site and L site were generated as described in Materials and Methods. Both probes bound RepB with different affinities (Fig. 6), corroborating the footprinting data.

If the protein concentration is much higher than the concentration of template, the concentration at which half of the template DNA is unbound is an approximation of the binding constant, $K_d(\text{DNA})$, for the protein (3). Serial dilutions of RepB were used in gel retardation assays with the H site and L site, and the fraction of free template was determined by scanning the autoradiograms. From the graphs in Fig. 7, approximate K_d of 25 and 300 nM were obtained for the H site and L site, respectively.

DISCUSSION

This work is a study of the two replication proteins, RepA and RepB, from the *M. fortuitum* plasmid pAL5000. We have mapped the 5' end of the single RNA which encodes both proteins and expressed the proteins in *E. coli* as MBP fusions. RepB was found to bind to two sites in the plasmid origin of replication, which were defined by DNase I footprinting techniques, and the binding constants were approximated by gel retardation assays. This is to our knowledge the first detailed dissection of protein-DNA interactions in mycobacteria.

The promoter region for the *rep* transcript does not resemble classic promoters from *E. coli* or archaeobacteria (10), nor is there much similarity to other mycobacterial promoters described to date (summarized in, e.g., reference 16). Clearly there is a great need for the characterization of more promoters from mycobacteria, especially since several reported 5' ends of mycobacterial mRNA have been deduced from primer extension experiments. As discussed above, this technique is prone to generating false signals when applied to RNA from

FIG. 5. DNase I footprinting assays of RepB on the *ori* region of pAL5000. (a) RepB binding to the sense strand (after factor X cleavage to remove the MBP moiety). The template was a PCR fragment generated using ^{32}P -5'-labelled oligonucleotide OGS1 and unlabelled oligonucleotide ORF1D. The DNA sequence is M13mp18 sequenced with the -40 universal primer (USB) as a size marker. RepB concentrations: lane 1, 270 nM; lane 2, 27 nM; lane 3, 2.7 nM; lane 4, 0.27 nM; lane 5, no protein. (b) Binding to the antisense strand. The template was generated from ^{32}P -labelled oligonucleotide OFP1 and unlabelled OGS1. The DNA sequence is M13mp18 sequenced with the -40 universal primer (USB) as a size marker. Lanes 1 to 6 represent footprints from untreated RepB-MBP fusion protein; lanes 7 to 12 were generated with factor X-treated protein. RepB concentrations: lanes 1 and 7, 270 nM; lanes 2 and 8, 27 nM; lanes 3 and 9, 2.7 nM; lanes 4 and 10, 0.27 nM; lanes 5 and 11, no protein. Lanes 6 and 12, untreated template.

FIG. 6. Gel retardation assays showing RepB specifically interacting with the L site (a) and the H site (b). The arrows show free DNA template. Protein concentrations: lane 1, 2600 nM; lane 2, 1280 nM; lane 3, 640 nM; lane 4, 320 nM; lane 5, 160 nM; lane 6, 80 nM; lane 7, 40 nM; lane 8, 20 nM; lane 9, 10 nM; lane 10, no protein.

G+C-rich organisms and should be corroborated by other techniques.

We believe that the 5' signal that we found represents a true transcription start point for several reasons. First, S1 mapping using probes derived from oligonucleotides mapping further

FIG. 7. Binding curves for RepB to the H site and L site showing the fraction of free template as a function of the protein concentration. The curves are averages from eight (for the H site) and seven (for the L site) independent measurements.

FIG. 8. Comparison of the pAL5000 *rep* promoter region with similar regions in other organisms. The motif common to pAL5000 and the area upstream of the *L. lactis* primase gene is marked by stars with the palindromic motif boxed. The 8-bp palindromic motif shared between pAL5000 and the *Corynebacterium* plasmid pEP2 is boldfaced and underlined. Arrows show inverted repeats which are present in pAL5000 and pEP2 but which are not identical in the two sequences. Start points for transcription are shown as arrows above the boldfaced start nucleotides. The RepB binding area shown comprises the entire protected area on both strands seen in DNase I footprinting experiments.

upstream showed no bands, indicating that the only detected signal was not from a processing event. Second, there are similarities in the RepB-binding region to the promoter for the *repA* gene, encoding the replication protein of the *Corynebacterium* wide-host-range plasmid pEP2 (47). A comparison of the two regions (Fig. 8) shows that the 8-bp palindrome CACCGGTG can be found 14 bp upstream of the start point of the pEP2 *repA* transcript. Although only a short motif and thus the identity could be random events, there are several other similarities between these two regions, e.g., a set of inverted repeats flanking the palindrome and a higher A+T content in the sequences immediately preceding and following the palindrome. Shuttle constructs based on the pEP2 sequence are able to replicate in mycobacteria (30), which supports the notion that these similarities are indeed essential parts of a regulatory structure.

Third, there is one 11-bp stretch in the *rep* promoter sequence which is also present upstream of the ORF encoding a DNA primase, DnaE, in *Lactococcus lactis* (2) (Fig. 8). This AT-rich motif could be a promoter element. It comprises an 8-bp palindrome, as does the motif common to pAL5000 *rep* and pEP2 *repA*.

Judging from the weakness of the S1 signal, the pAL5000 *rep* transcript is produced at a low level. The footprinting experiments described below indicate that the *rep* promoter is tightly regulated, supporting these observations.

Genetic evidence has demonstrated that both RepA and RepB are necessary for replication (22, 39). If the Rep proteins act to trigger replication of the plasmid, the number of RepA and RepB molecules in the cells would affect the efficiency of replication. The low copy numbers of pAL5000-derived replicons (39) could be a consequence of the low level of replication proteins.

The specific binding of the small (119-aa) RepB protein to the *ori* region was surprising, since on the basis of similarities to related proteins, we had expected RepA to bind to *ori*. The only significant similarity with proteins in the GenBank database is to the recently published but undefined product of the ORF2 from the *B. longum* plasmid pMB1 (31) (Fig. 9). The putative ORF2 product would be 106 aa, comparable to the 119-aa RepB. The organization of this plasmid seems very

reminiscent of pAL5000, with two replication proteins. The larger ORF1 of pMB1 is indeed similar to RepA. No closer study of the replication of this plasmid has been undertaken so far, but on the basis of our work we would expect the pMB1 ORF2 product to have DNA-binding properties.

The DNase I footprinting experiments described here show that RepB binds specifically to two sites in the *ori* region of the plasmid (Fig. 3). We had hypothesized that the replication protein might bind to a set of repeated sequences upstream of the actual binding site (shown by arrows in Fig. 3), but this turned out not to be the case. The function of these repeats is still unknown. It may be that they are responsible for the incompatibility function, since an extra copy of the *ori* region introduced into the plasmid enhances incompatibility, but only when in the same orientation as the wild-type *ori* (39). Such an arrangement of an *ori*, necessary for replication, and a separate set of repeated sequences, responsible for incompatibility but not essential for replication, has been described for plasmid P1 (1, 6, 7). Whether the incompatibility-conferring part of pAL5000 is necessary for replication is still unknown.

The two distinct binding sites for RepB, here called the H site and the L site, seem to have different roles in the replication of the plasmid. The H site partially occludes the putative promoter structures, suggesting that RepB autoregulates its expression. The binding to this box takes place on both DNA strands. In the middle of this binding site is the palindrome CACCGGTG, which could be a recognition motif for RepB. The other overlapping palindrome, GATTAATC (Fig. 3), which we suggested above is part of the promoter structure, would be partially covered by bound protein.

The binding of RepB to its promoter has a higher affinity than the binding to the second region, the L site, at positions 4534 to 4564. Here, the protein seems to bind more strongly, or exclusively, to one side of the DNA helix. The L site is probably too far away from the promoter region to have any influence on the regulation of *rep* expression, and we assume that this site is the *ori* itself. This site is different from that predicted from sequence comparisons (32).

The binding constants, 25 nM for the H site and 300 nM for the L site, are of course rough approximations from experiments in vitro under possibly suboptimal conditions. Neverthe-

RepB	QPTVRKKRRVTAEGARITGLSERHVRLVAQERSEWLAEQAARRERIRAYH 67
ORF2	RTTLRKKRPVSARELAAYGVSTRTIQSWVAMKREDWIDEQAAMREAVRSYH 54
RepB	DDEGHSWPQTAKHFLHLDTVKRLGYRARKERAAEQEAAQKAHNEADNPPLF 119
ORF2	DDEGHTWPQTAEHFNMSQGAVRQRCYRARKER. EDEAAEKS KHLPGETPLF 104

FIG. 9. Sequence alignment of RepB to the ORF2 product of plasmid pMB1 from *L. lactis* (31) using the Bestfit algorithm from the Wisconsin Genetics Computer Group package. Lines show identity; double dots show similarity; single dots show conservative substitutions.

less, the relative values from the gel retardation studies agree with what is seen in the footprinting experiments. The findings suggest a model for the triggering of replication: the H sites on plasmid molecules in the cell titrate RepB, which will not bind to *ori* as long as there are free H sites available. Ultimately, all those sites will be occupied and RepB will be able to bind to the low-affinity *ori* box, triggering replication and thus generating a new set of unoccupied binding sites. Since RepB seems to autoregulate its synthesis, the moment all H sites are occupied there will be competition between shutdown of protein synthesis and triggering of replication by binding to *ori*.

This model may be an oversimplification, as other plasmids from different bacteria have more-complex regulatory circuits (15, 25–27). The expression of RepB may well be more intricately regulated. Nevertheless, the current growth in molecular techniques applicable to mycobacteria, such as in vitro transcription systems with the *M. smegmatis* RNA polymerase (20, 28), should make it possible to test some of the implications of these hypotheses.

The role of RepA in the biology of pAL5000 remains to be established. The protein is necessary for replication. The ColE2 Rep to which it is similar is a plasmid-specific primase which binds to double-stranded DNA and synthesizes a unique RNA primer (43, 44). ColE2-type plasmids need host-encoded proteins, e.g., DNA polymerase I for the initial DNA synthesis and DNA polymerase III for completing the synthesis (13, 14, 42, 46). There are indications that pAL5000-based replicons have a similar requirement for polymerase I (24). Since, in contrast to pAL5000, none of the plasmids of the ColE2 class has a requirement for two Rep proteins, the parallels should not be drawn too far.

The inactivity of RepA in our assays is puzzling, and the easiest explanation is that the MBP fusion renders the protein inactive, as is often seen with this system (35). Digesting the MBP-RepA protein with factor X protease gave rise to several bands on an SDS-PAGE gel, where only two were expected, indicating nonspecific cleavage of the fusion protein or possibly instability of the cleaved RepA.

Other explanations could be that the assay conditions were not optimal or, since all assays were carried out in vitro, that RepA needs host proteins for its activity, possibly interacting with the replication machinery of the host. Expressing the protein in a different system would be a way of testing these explanations.

ACKNOWLEDGMENTS

This work was carried out as part of the Glaxo-Wellcome Action TB initiative. P.S. also thanks Glaxo-Wellcome for additional support allowing the DNase I footprinting experiments to be performed at the South African Institute for Medical Research in Johannesburg.

Special thanks are due to S. Anderson for help in setting up the DNase I footprinting assay and providing *E. coli* LexA protein; she, V. Mizrahi, S. S. Dawes, and A. DeMeyer are thanked for stimulating discussions. R. McAdam generously provided plasmid pYUB285. K. Duncan's critical comments on the manuscript are gratefully acknowledged.

REFERENCES

1. Abeles, A. L., L. D. Reaves, and S. J. Austin. 1989. Protein-DNA interactions in regulation of plasmid P1 replication. *J. Bacteriol.* **171**:43–52.
2. Araya, Y., N. Ishibashi, S. Shimamura, K. Tanaka, and H. Takahashi. 1993. Genetic and molecular analysis of the *rpoD* gene from *Lactococcus lactis*. *Biosci. Biotechnol. Biochem.* **57**:88–92.
3. Astatke, M., N. D. F. Grindley, and C. M. Joyce. 1995. Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **270**:1945–1954.
4. Beggs, M. L., J. T. Crawford, and K. D. Eisenach. 1995. Isolation and sequencing of the replication region of *Mycobacterium avium* plasmid pLR7. *J. Bacteriol.* **177**:4836–4840.
5. Chatteraj, D., K. Cordes, and A. Abeles. 1984. Plasmid P1 replication: negative control by repeated DNA sequences. *Proc. Natl. Acad. Sci. USA* **81**:6456–6460.
6. Chatteraj, D. K., R. J. Mason, and S. H. Wickner. 1988. Mini-P1 plasmid replication—the autoregulation-sequestration paradox. *Cell* **52**:551–557.
7. Chatteraj, D. K., K. M. Snyder, and A. L. Abeles. 1985. P1 plasmid replication: multiple functions of RepA protein at the origin. *Proc. Natl. Acad. Sci. USA* **82**:2588–2592.
8. Crawford, J. T., M. D. Cave, and J. H. Bates. 1981. Evidence for plasmid mediated restriction modification in *Mycobacterium avium* intracellular. *J. Gen. Microbiol.* **127**:333–338.
9. De Smet, K. A. L., S. Jamil, and N. G. Stoker. 1993. Tropist3: a cosmid vector for simplified mapping of both G + C-rich and A + T-rich genomic DNA. *Gene* **136**:215–219.
10. Hain, J., W. D. Reiter, U. Hudepohl, and W. Zillig. 1992. Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res.* **20**:5423–5428.
11. Hanahan, D., J. Jessee, and F. R. Bloom. 1991. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* **204**:63–114.
12. Hiraga, S. I., T. Sugiyama, and T. Itoh. 1994. Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* **176**:7233–7243.
13. Itoh, T., and T. Horii. 1989. Replication of ColE2 and ColE3 plasmids: in vitro replication dependent on plasmid-coded proteins. *Mol. Gen. Genet.* **219**:249–255.
14. Kingsbury, D. T., and D. R. Helinski. 1970. DNA polymerase as a requirement for the maintenance of the bacterial plasmid colicinogenic factor E1. *Biochem. Biophys. Res. Commun.* **41**:1538–1544.
15. Kittell, B. L., and D. R. Helinski. 1991. Itron inhibition of plasmid RK2 replication in vitro: evidence for intermolecular coupling of replication origins as a mechanism for RK2 replication control. *Proc. Natl. Acad. Sci. USA* **88**:1389–1393.
16. Kremer, L., A. Baulard, J. Estaquier, O. Poulain-Godefroy, and C. Locht. 1995. Green fluorescent protein as a new expression marker in mycobacteria. *Mol. Microbiol.* **17**:913–922.
17. Labidi, A., C. Dauguet, K. Seng Goh, and H. L. David. 1984. Plasmid profiles of *Mycobacterium fortuitum* complex isolates. *Curr. Microbiol.* **11**:235–240.
18. Labidi, A., H. L. David, and D. Roulland-Dussoix. 1985. Cloning and expression of mycobacterial plasmid DNA in *Escherichia coli*. *FEMS Microbiol. Lett.* **30**:221–225.
19. Labidi, A., H. L. David, and D. Roulland-Dussoix. 1985. Restriction endonuclease mapping and cloning of *Mycobacterium fortuitum* var. *fortuitum* plasmid pAL5000. *Ann. Inst. Pasteur Microbiol.* **136**:209–215.
20. Levin, M. E., and G. F. Hatfull. 1993. *Mycobacterium smegmatis* RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. *Mol. Microbiol.* **8**:277–285.
21. Little, J. W., D. W. Mount, and C. R. Yanisch-Perron. 1981. Purified *lexA* protein is a repressor of the *recA* and *lexA* genes. *Proc. Natl. Acad. Sci. USA* **78**:4199–4203.
22. McAdam, R. A., T. R. Weisbrod, J. Martin, J. D. Scuderi, A. M. Brown, J. D. Cirillo, B. R. Bloom, and W. R. Jacobs, Jr. 1995. In vivo growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* **63**:1004–1012.
23. Meissner, P. S., and J. O. Falkinham III. 1986. Plasmid DNA profiles as epidemiological markers for clinical and environmental isolates of *Mycobacterium avium*, *Mycobacterium intracellulare*, and *Mycobacterium scrofulaceum*. *J. Infect. Dis.* **153**:325–331.
24. Mizrahi, V. Personal communication.
25. Nordström, K. 1985. Control of plasmid replication: theoretical considerations and practical solutions, p. 189–214. In D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender (ed.), *Plasmids in bacteria*. Plenum Press, New York.
26. Nordström, K. 1990. Control of plasmid replication—how do DNA iterons set the replication frequency? *Cell* **63**:1121–1124.
27. Novick, R. P. 1989. Staphylococcal plasmids and their replication. *Annu. Rev. Microbiol.* **43**:537–565.
28. Predich, M., L. Doukhan, G. Nair, and I. Smith. 1995. Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. *Mol. Microbiol.* **15**:355–366.
29. Qin, M., H. Taniguchi, and Y. Mizuguchi. 1994. Analysis of the replication region of a mycobacterial plasmid, pMSC262. *J. Bacteriol.* **176**:419–425.
30. Radford, A. J., and A. L. M. Hodgson. 1991. Construction and characterization of a *Mycobacterium-Escherichia coli* shuttle vector. *Plasmid* **25**:149–153.
31. Ranes, M. G., J. Rauzier, M. Lagranderie, M. Gheorghiu, and B. Gicquel. 1990. Functional analysis of pAL5000, a plasmid from *Mycobacterium fortuitum*: construction of a “mini” mycobacterium-*Escherichia coli* shuttle vector. *J. Bacteriol.* **172**:2793–2797.
32. Rauzier, J., J. Moniz-Pereira, and B. Gicquel-Sanzey. 1988. Complete nucleotide sequence of pAL5000, a plasmid from *Mycobacterium fortuitum*. *Gene* **71**:315–321.

33. **Reiter, W.-D.** 1988. Genexpression im Archaeobacterium *Sulfolobus*: Untersuchungen anhand des virusartigen Partikels SSV1 und der ribosomalen RNA-Gene. Ph.D. thesis. Ludwigs-Maximilians-Universität, Munich.
34. **Rossi, M., P. Brigidi, A. Gonzalez Vara y Rodriguez, and D. Matteuzzi.** 1996. Characterization of the plasmid pMB1 from *Bifidobacterium longum* and its use for shuttle vector construction. *Res. Microbiol.* **147**:133–143.
35. **Sayers, J.** Personal communication.
36. **Snapper, S. B., L. Lugosi, A. Jekkel, R. E. Melton, T. Kieser, B. R. Bloom, and W. R. Jacobs, Jr.** 1988. Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc. Natl. Acad. Sci. USA* **85**:6987–6991.
37. **Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**:1911–1919.
38. **Stolt, P.** Unpublished observations.
39. **Stolt, P., and N. G. Stoker.** 1996. Functional definition of regions necessary for replication and incompatibility in the *Mycobacterium fortuitum* plasmid pAL5000. *Microbiology* **142**:2795–2802.
40. **Stolt, P., and W. Zillig.** 1993. Antisense RNA mediates transcriptional processing in an archaeobacterium, indicating a novel kind of RNase activity. *Mol. Microbiol.* **7**:875–882.
41. **Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs, Jr., and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. *Nature (London)* **351**:456–460.
42. **Tacon, W., and D. Sherratt.** 1976. ColE plasmid replication in DNA polymerase I-deficient strains of *Escherichia coli*. *Mol. Gen. Genet.* **147**:331–335.
43. **Takechi, S., and T. Itoh.** 1995. Initiation of unidirectional ColE2 DNA replication by a unique priming mechanism. *Nucleic Acids Res.* **23**:4196–4201.
44. **Takechi, S., H. Matsui, and T. Itoh.** 1995. Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. *EMBO J.* **14**:5141–5147.
45. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**:3–11.
46. **Yasueda, H., T. Horii, and T. Itoh.** 1989. Structural and functional organization of ColE2 and ColE3 replicons. *Mol. Gen. Genet.* **215**:209–216.
47. **Zhang, Y., J. Praszker, A. Hodgson, and A. J. Pittard.** 1994. Molecular analysis and characterization of a broad-host-range plasmid, pEP2. *J. Bacteriol.* **176**:5718–5728.