396–402 Nucleic Acids Research, 1999, Vol. 27, No. 2 © 1999 Oxford University Press

Identification of promoter elements in mycobacteria: mutational analysis of a highly symmetric dual promoter directing the expression of replication genes of the *Mycobacterium* plasmid pAL5000

Pelle Stolt*, Qiuwang Zhang+ and Stefan Ehlers

Division of Molecular Infection Biology, Research Centre Borstel, Parkallee 22, D-23845 Borstel, Germany

Received November 6, 1998; Revised and Accepted November 23, 1998

ABSTRACT

The 120 bp origin of replication (*ori*) for the *Mycobacterium* plasmid pAL5000 has been shown to comprise the binding sites for the replication protein RepB as well as the start site of transcription for the *repA* and *repB* genes, encoding the replication proteins RepA and RepB. In this work it is demonstrated that a third gene product, Rap, is involved in replication in addition to the previously described proteins. *Mycobacterium smegmatis* cells transformed with replicons carrying the *rap* gene recover markedly faster upon electroporation than those transformed with the minimal replicon, which lacks *rap*. The *rap* gene, oppositely orientated to *repA/B*, was shown to be transcribed from a promoter orientated back-to-back to and overlapping the *repA/B* promoter. As a consequence of the extensive dyad symmetry in this region the two promoters share several elements, most of which are situated inside the high-affinity RepB-binding motif in the *ori*. Transcription of *rap* runs through the low-affinity RepB-binding site, which is part of the *ori* and necessary for replication. Both promoters were shown to be repressed by RepB. These divergent promoters were studied through site-specific mutagenesis in a *xylE* reporter gene assay. The analysis furnished evidence supporting the existence of a distal as well as a proximal element in mycobacterial promoters.

INTRODUCTION

The plasmid pAL5000 from *Mycobacterium fortuitum* (1; GenBank accession no. M23557) is the best characterised mycobacterial replicon. Constructs based on this plasmid are widely used in *Eschericha coli–Mycobacterium* shuttle vectors, which are able to replicate in a wide variety of mycobacteria, including the pathogen *Mycobacterium tuberculosis* (2–6).

In addition to its use as a vector for genetic manipulation, pAL5000 has proved a good system for the studies of gene expression and protein–DNA interactions (7–9), about which little is known in mycobacteria. It was previously shown that the minimal functional replicon of pAL5000 comprises a *cis*-acting origin (*ori*) and two genes, *repA* and *repB*, coding for replication proteins. The *ori* region is characterised by a high incidence of repeated sequences. The *repA* and *repB* genes overlap by 1 bp and are transcribed as a single RNA species.

Both RepA (277 amino acids) and RepB (119 amino acids) are necessary for replication. RepA has proved difficult to purify as a recombinant protein in *E.coli* and its role in pAL5000 replication remains unknown. RepB binds with different affinities to two sites in the *ori* region, separated by 57 bp. Binding to the low-affinity site (L-site) is to one side of the DNA helix only. The exact role of binding to the L-site has not been established, but the site is necessary for replication. Binding to the high-affinity site (H-site) is to both sides of the DNA helix; here RepB seems to bind co-operatively in two copies, though whether as a monomer or dimer is unknown. The H-site overlaps with the promoter region for the *repA/B* cotranscript and thus RepB acts as an autoregulator of its expression. The affinity of RepB for the H-site is some 10-fold higher than that for the L-site.

While there are no obvious symmetry motifs in the L-site, the H-site has an intricate structure (Fig. 1). It consists of two interlocking 8 bp palindromes, one of which is (G+C)-rich and one (A+T)-rich. In addition there is a 5 bp inversely repeated motif (GACCA). The (G+C)-rich palindrome (the GC-box) has been shown in a mutational study to be essential for RepB binding and also for replication. Several of the bases upstream of the GC-box likewise are important for a functional replicon. It was possible through a single base pair change to abolish protein binding to one of the H-site motifs whilst retaining binding to the second; this mutation produced a non-functional replicon.

Most changes to the (A+T)-rich palindrome (the AT-box) have no effect on RepB binding nor on replication (9). Since the AT-box has no role in RepB binding, a hypothesis was that this palindrome has a function as a promoter element. Not much is known about promoters in mycobacteria and there have been few thorough mutational studies of promoter elements; with a few

*To whom correspondence should be addressed. Tel: +49 4537 188486; Fax: +49 4537 188686; Email: pstolt@altavista.net

+On leave from the Department of Microbiology and Immunology, Guangzhou Medical College, Guangzhou, 510182, Peoples' Republic of China

Figure 1. DNA sequence of the *ori* region of pAL5000. The low-affinity binding site for RepB is underlined; the high-affinity binding site is boxed. The GC-box and AT-box motifs are boxed. Sequences corresponding to some of the oligonucleotides used in the analyses are represented by arrows pointing in the 5′–3′ direction of the oligonucleotide. Arrowheads mark start points for transcripts and open reading frames, respectively. For the region around the H-site, both strands of the DNA sequence are shown.

exceptions (10–13) most work has been sequence comparisons of areas upstream of transcription start points.

The current study shows that the *repA/B* promoter is in reality an unusually overlapping dual promoter, directing the synthesis of two divergently transcribed mRNA species. The second mRNA, starting within the H-site and transcribed through the L-site in the opposite direction to *repA/B*, is shown to encode a protein (termed Rap) involved in replication. This dual promoter was subjected to a detailed mutational study, which gathered evidence for the presence of two promoter elements: a proximal box and a distal box. Single base pair mutations in these elements could abolish promoter function.

Thus the 102 bp *ori* region of pAL5000 is a biologically very active stretch of DNA, acted upon by RepB, two RNA polymerase complexes, the DNA replication machinery and possibly RepA and the *rap* gene product as well.

MATERIALS AND METHODS

Materials

The *E.coli* strain used for all DNA manipulation was NM554 (14) which was grown in LB medium (10 g tryptone, 5 g yeast extract, 5 g NaCl/l) with antibiotic if necessary [final concentration 50 µg/ml kanamycin (Km), 50 µg/ml ampicillin (Ap) or 40 µg/ml chloramphenicol (Cm)]. *Mycobacterium smegmatis* strain mc^{2}155 (15) was grown in LB or Lemco medium or on Middlebrook 7H10 plates with the appropriate antibiotic [30 µg/ml Km or 100 µg/ml hygromycin (Hy)].

DNA extraction and manipulation

Plasmid DNA was isolated from *E.coli* using Qiaspin miniprep columns from Qiagen. Mycobacterial plasmid DNA was isolated through electroduction (16) into *E.coli* cells followed by standard preparation methods. DNA manipulation was done using standard techniques (17).

The construction of most of the replicons used has been described previously (7,9). Replicons with mutated promoter motifs were made as those previously described (9) using the same oligonucleotides as in the *xylE* assay. pDQ6, pDQ13 and pDQ86 were constructed using the oligonucleotides OΔinc (5′-CTGGTTGGATCCGGTGGTTGGGGGTG-3′), OΔL (5′-CAGCGAGATATCTGACTTGGAGCT-3′) and HLEV (5′-GCGGGGGAGTGTGCAGTTGTGGGGTGGCCCCTCAGCGAGATATC-3′), respectively.

The integrating construct pDQ263, expressing *rap* from the +14 up-mutation of the *rap* promoter, was constructed by ligating the PCR product from ORF1D/OKD+14 directly into pCR-Script Amp SK+ (Stratagene) creating pDQ249, then digesting with *Eco*RV/*Cla*I and adding the *rap* gene as a PCR product (from the oligonucleotides ORF1A and the kinased OKDB; the PCR product was digested with *Cla*I) and moving the construct as a filled-in *Xba*I cassette to pDQ56 cut with *Dra*I.

To test for *rap*-mediated enhancement of replication in *trans*, pDQ265 was made by introducing the hygromycin resistance gene from pIJ963 (18) into pUH43 (7) as a *Bam*HI/*Bgl*II fragment ligated to the *Bam*HI site.

RNA preparation and transcription analysis

For reverse transcription–polymerase chain reactions (RT–PCR), 2 µg RNA from either wild-type *M.smegmatis*, prepared as described previously (8), or cells carrying the pAL5000-derived shuttle vector pUH4 (7), were precipitated with 1 pmol of the oligonucleotide ORF1A (5′-CACCAGCTCTAGAAGTTCGGGCGCTG-3′; bp 4315–4340) and taken up in a total volume of 20 µl buffer for murine-leukemia virus–reverse transcriptase (MLV–RT; United States Biochemical) containing 1 mM dNTPs. 200 U of MLV-RT were added and the reaction allowed to proceed at 37°C for 20 min. The reaction mixture was treated with RNase I and 1/10 of the product used in subsequent PCR reactions (30 s at 95°C, 30 s at 56°C, 30 s at 72°C for 25 cycles). Oligonucleotides employed were ORF1C (nt 4577–4558), OKDB (nt 4588–4570), OHR (nt 4627–4608) and OFP1 (nt 4705–4686). The products were analysed on agarose gels.

Primer extension was carried out as described (19) using 5 µg RNA and 0.2 pmol 5′-labelled oligonucleotide.

Table 1. List of the constructs used to define the *rap* gene

Construct	Base pairs	Encompassed features	Days typically needed for *M.smegmatis* to form 0.5-mm colonies
pUH61	3875-1093	*repA; repB*; H-site; L-site; *rap*; + 229 bp	3
pUH83	4050-1093	*repA; repB*; H-site; L-site; *rap* + 52 bp	3
pUH96	4165-1093	*repA; repB*; H-site; L-site; first 95 codons of *rap*	4
pUH52	4327-1093	*repA; repB*; H-site; L-site; first 42 codons of *rap*	5
pDQ6	4474-1093	*repA; repB*; H-site; L-site + 60 bp	5
pDQ81	4532-1093	*repA; repB*; H-site; L-site + 3 bp	5
pDQ13	4563-1093	*repA; repB*; H-site	-

The base pair numbering is according to the published pAL5000 sequence (1).

Electroporation

Competent *M.smegmatis* mc^{2}155 cells were prepared as described by Snapper *et al*. (15). Transformation with 0.1–1 µg DNA was performed using 100 µl aliquots of cells.

Construction of *xylE* fusions to mutated promoters

The construct pDQ56 carries the *xylE* gene from the plasmid pRCX3 (20) which was a kind gift from V. Deretic. The *xylE* gene was excised from pRCX3 on a *Kpn*I fragment and ligated to the *Kpn*I-digested integrating vector pMV306 (4); the resulting construct was called pDQ56.

The mutated *repA/B* and *rap* promoter sequences were synthesised in PCR reactions using kinased forward oligonucleotides based on the wild-type sequence 5′-CGTGTCGGACCATACACCGGTGATTAA-3′ (Oligonucleotide OKDH; 8) but with mutations corresponding to those shown in Figure 3. For mutations downstream of the end of OKDH, the oligonucleotide carried additional sequences at the 3′ end corresponding to the pAL5000 sequence. A reverse oligonucleotide, ORF1D (5′-TGGCGCGATTGCCCACGATC-3′; 7) was used. The PCR products were digested with *Sma*I and ligated to the *Dra*I-digested pDQ56. The orientation and sequence of the inserts were verified by DNA sequencing, performed on an automated DNA sequencer (ABI Prism 377; Applied Biosystems, Foster City, Ca) using the BigDye RR Terminator Cycle sequencing kit (Applied Biosystems) according to the manufacturer's instructions. The oligonucleotide OXyleR (5′-CATGTCCAGTACACGCAGCT-3′), which is complementary to part of the *xylE* gene, was used for sequencing.

Cells were grown in 5 ml medium (LB + Km) and harvested in late log phase by centrifugation. The cells were washed in 500 µl phosphate puffer (50 mM potassium phosphate pH 7.5), taken up in 100 µl of the same buffer and lysed in a bead beater. The typical yield was 300–400 mg protein. After centrifugation in a microcentrifuge to clear the lysates, enzymatic activity was determined in a 0.3 mM solution of catechol in phosphate buffer at 375 nm; readings were taken for 3 min and the mean rate per minute calculated. Eight to 16 single colonies from each construct were analysed. The activity was calculated as mU/mg total protein (20). The molar extinction coefficient for the reaction product (2-hydroxymuconic semialdehyde) is 4.4×10^4.

Expression of recombinant Rap protein

Recombinant Rap protein was expressed in the pBAD*Myc*His system (Invitrogen). The gene was amplified in a PCR reaction with the oligonucleotides RapS (5′-CGGCAGGCCTGGCGGCGTTCGCCACGATGTC-3′) and RapNc (5′-AGGAGGAACCCATGGGTCGTTTCG-3′). The PCR product was ligated as an *Nco*I/*Stu*I-cut fragment to pBAD*Myc*HisC digested with *Nco*I/*Sna*BI. The correct DNA sequence was verified before Rap was expressed and affinity purified on a Ni-NTA Superflow matrix (Qiagen).

RESULTS

Definition of the *rap* gene

There are five open reading frames on the pAL5000 DNA (1). Though only two, RepA and RepB, are necessary for replication, it has been observed that when additional sequences are present upstream of the minimal replicon, *M.smegmatis* cells recover faster upon electroporation (7). A small (115 amino acids) open reading frame, ORF5, is present in the upstream region and might encode an auxiliary factor. ORF5 starts immediately upstream of the L-site (nt 4452–4102) and is orientated in the opposite direction to the *repA* and *repB* genes. However, since the putative ORF5 product has no convincing similarity to any gene product in the GenBank database (P.S., unpublished), there remained the possibility that the DNA itself upstream of the L-site, through its many structural peculiarities (direct and inverted repeats) was responsible for the improved replication ability.

A closer study of different constructs carrying all or parts of the ORF5 (Table 1) showed that the improved replication is indeed due to a gene product and not to a *cis*-acting DNA structure. *Mycobacterium smegmatis* cells transformed with constructs carrying upstream sequences of varying length were monitored in parallel twice a day. The smallest functional replicon previously constructed was pUH52, but another 203 bp could be deleted, creating pDQ81, which carries no sequences upstream of the L-site and still replicates. Cells transformed with this construct needed ~5 days to form colonies of 0.5 mm in diameter. The recovery time for pUH52 which includes several repeated sequence motifs, but only half of ORF5, was similar to pDQ81. Comparisons between plates showed that when ORF5 was present (construct pUH83), transformants needed 3 days to form colonies of similar size. When all but the last 21 codons of *rap* were included (pUH96) the transformed cells formed 0.5 mm colonies in 4 days, indicating that a truncated *rap* still retains some functionality.

Mapping of *rap* transcription

The Rap ORF starts at base 4452 (Fig. 1). To define the start point of transcription, a preliminary mapping was undertaken using RT–PCR with the oligonucleotide ORF1A for the reverse

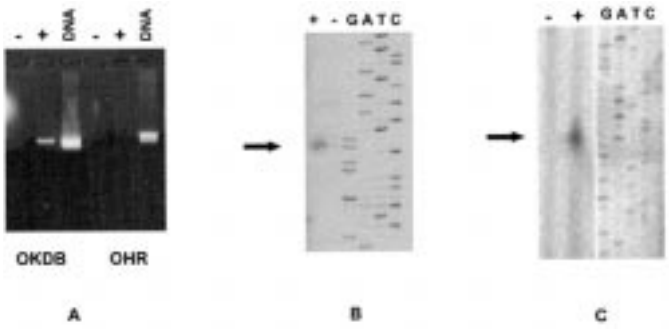

Figure 2. Definition of the 5′ end of *rap* mRNA. (**A**) RT–PCR analysis to narrow down the possible region of initiation. The oligonucleotide ORF1A (bp 4315–4340) was used for the reverse transcription reaction. The PCR reactions were done with ORF1A and different forward oligonucleotides as shown under the gel. A (–) sign indicates RNA from *M.smegmatis* cells lacking pAL5000 DNA; a (+) sign mRNA from *M.smegmatis* cells carrying pAL5000 as the construct pUH4 (7). The DNA used as a template for the positive control PCR was pUH4. (**B**) Primer extension analysis of the 5′ end of the *rap* mRNA using the oligonucleotide OGS1 (bp 4419–4438). The DNA sequence used as a size marker (a gift from L. Hamann, Research Centre Borstel) was from the gene for human LPS-binding protein sequenced with the oligonucleotide 5′-AATCGTCTGTGGACCCC-3′. A (–) sign indicates RNA from *M.smegmatis* cells lacking pAL5000 DNA; a (+) sign mRNA from *M.smegmatis* cells carrying pAL5000 as the construct pUH4. The arrow points to the 179 nt signal from the *rap* transcript. (**C**) Primer extension analysis as in (B) but with the oligonucleotide OKDL (bp 4531–4554). The DNA sequence is pUH52 sequenced with the same oligonucleotide as used in the primer extension reaction. The arrow points to the signal from the *rap* transcript.

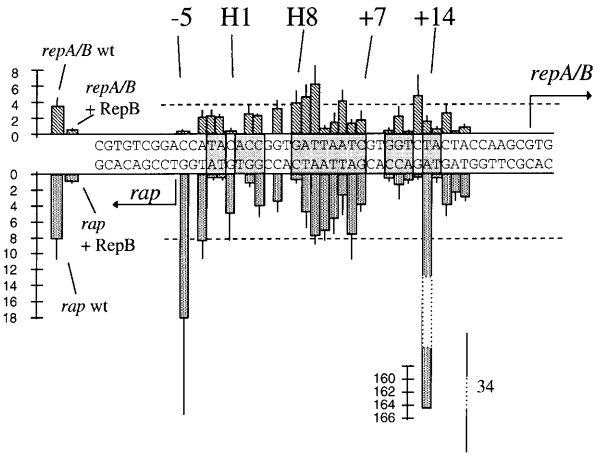

Figure 3. Mutational assay of the *rap-repA/B* promoter. The sequences are shown starting on the base corresponding to the 5′ end of the oligonucleotides used to generate mutated promoters in the *xylE* assay. Boxed and shaded sequences are symmetry motifs present in the inverse orientation on both DNA strands. The scale bars are graded in mU/mg protein; the scale for the *rap* assay is inverted with respect to the *repA/B* scale. Wild-type values are shown on the far left adjacent to values for when the repressor RepB is present in the cells; the error bars are standard deviations. Broken lines above the sequences indicate wild-type values for *repA/B* and *rap* respectively to aid comparison. Transcriptional start points are shown as arrows. The nomenclature of bases is shown above the sequence: H1 stands for the first base in the GC-box; H8 the last base in the GC-box; +1 is the first base downstream of this box, –1 the first base upstream of the box etc. Note that the bar for the +14 *rap* value is broken to fit on the page.

transcription reaction and a set of different oligonucleotides for the PCR reactions. Bands were obtained (Fig. 2A) with oligonucleotides mapping up to bp 4588 (OKDB) but not with an oligonucleotide mapping to bp 4608 (OHR).

The exact start point was then defined in primer extension experiments with the oligonucleotides OGS1 and OKDL (Fig. 2B and C). Primer extension assays are not always trustworthy in (G+C)-rich organisms owing to secondary structures in the mRNA; however, the RT–PCR assay had limited the area where transcription might start to a 20 nt window. The primer extension analysis located the *rap* transcript start to the T, base 4595. This is within the 20 nt window and is in fact within the H-site, only 37 bp from the start point of *repA/B* transcription.

The small distance implies that the promoters overlap to a great extent. It has already been noted that there are several palindromes and repeats in the H-site. In fact, it is possible to place a line of symmetry in the middle of the AT-box which shows many of the sequences on each side to be mirror images of each other. Thus the entire dual promoter can be treated as an imperfect palindrome. The shared sequence motifs fall into three areas: the AT-box (itself a palindrome), a –10 area and a –35 area. In the case of the *rap* promoter the latter two motifs fall closer to the transcription start point than for the *repA/B* promoter, but allowing for this, the structure would seem neither a back-to-back nor face-to-face promoter, but rather one bi-directional structure.

Mutational study of the dual promoter

Earlier promoter studies in mycobacteria have identified possible –10 and –35 regions as promoter elements (10–12) but no promoter has been thoroughly mapped. The structure of the dual *rap-repA/B* promoter, with shared possible proximal and distal elements, made it an excellent system for a thorough study of these potential elements. A mutation to the proximal element in one promoter would lead to a mutation in the distal motif in the other.

The *xylE* gene was selected as reporter gene for the study, as it has been shown to be a useful system for promoter fusions in mycobacteria (20). A complication in the assay was that RepB represses the *repA/B* promoter and hence a *xylE* fusion carried on a pAL5000-derived shuttle vector would be liable to interference from the repressor. To circumvent this, the fusion constructs were integrated into the *M.smegmatis* chromosome on the plasmid pDQ56, which carries the *xylE* reporter gene on the integrating vector pMV306 (4). There is only one attB site in the *M.smegmatis* genome (21), ensuring that the genetic environment would be the same for all integrants. For the promoter assay, *M.smegmatis* cells were transformed with the different constructs and transformants selected on Km plates. To enable comparisons between the two promoters, only one reporter gene was used and the promoters were inserted in one or the other direction front of *xylE*.

The construct pDQ56 on its own had negligible activity in a *xylE* assay (0.07 mU/mg). The wild-type *repA/B* promoter (pDQ156) scored 3.5 mU/mg, comparable to the 8 mU/mg scored by the wild-type *rap* promoter (pDQ157). When the *M.smegmatis* cells carrying pDQ156 and pDQ157 were transformed with pAL5000, the promoter activity dropped by an order of magnitude (Fig. 3), indicating that under normal replicating conditions, the *rap* promoter is repressed as well as the *repA/B* promoter. Repression was not complete, however; the residual activities 0.6 (*repA/B*) and 0.8 (*rap*) were significantly above that of the pDQ56 background.

Next, a series of oligonucleotides were synthesised, all with the same 5′ end but with different single base pair substitutions. Since promoter activity was thought to be dependent not only on binding affinity, but also on the energy needed to open the DNA helix, A was changed to T and G to C and vice versa, keeping the melting energy constant. Since the AT-box is the most extensive shared motif, it was considered potentially crucial to the dual activity and all bases in this box were mutated one at a time. In addition, most bases in the shared proximal and distal motifs were changed, as were some bases outside the shared areas.

A summary of this assay is shown in Figure 3. From the extensive symmetry of the promoters, one might have expected a down-mutation in one promoter to be accompanied by an up-regulation in the other. This was seldom observed, however, and strongly seen only in the case of the mutation –5, where the *rap* promoter was up-regulated concurrent with a down-regulation of *repA/B*.

The general conclusion which could be drawn from the mutational assay was that both the proximal and the distal regions are important for both promoters. Changes to the AT-box were less conclusive. Some changes to the proximal and distal boxes (e.g. –5 or +10 in *repA/B* and –2 or +10 in *rap*) were able practically to abolish promoter activity, reducing it to a greater degree than was seen in repression by RepB.

A fortuitous deletion occurred in one construct mutating the base –3 in the *rap* orientation. The mutation itself (pDQ224) showed wild-type activity. In the deletion (pDQ215), the 2 bp closest to the transcriptional start point are missing. This led to an inactivation of the promoter, indicating that the distance between the promoter elements and the start point is critical and in the case of the *rap* promoter, close to the shortest allowed. Changing the base immediately before the start point (–5) if anything led to an increase in promoter activity (Fig. 3).

To test if the lower melting energy for (A+T)-rich sequences was of importance for the AT-box function, the base A at position +5 was changed to a G (pDQ233 and pDQ234). This had only a moderate effect on *rap* (5.7 mU/ml) as well as on *repA/B* (2.8 mU/ml).

Replicative abilities of mutated plasmids

Plasmids carrying the *repA* and *repB* genes, the L-site and an H-site with some of the mutations described here have been tested in *M.smegmatis* for their ability to replicate (9). Most changes to the GC-box which introduce lower affinities for RepB binding have been shown to be detrimental to the replication ability while few changes to the AT-box and downstream sequences had any effect (9). The new mutations +4, +6, +11 and +14, inside and downstream of the AT-box were tested. Plasmids carrying the mutation +4, +6 and +11 replicated, but those with mutation +14 did not, in line with the results from the more limited earlier study.

Studies on Rap

If the *rap* gene product acts on the pAL5000 DNA it might be possible to enhance the replication of *rap*-less replicons by providing the gene product in *trans*. To test this, pDQ263 was constructed, an integrating construct expressing the *rap* gene from the +14 up-mutation of the *rap* promoter. pDQ265, a pUH52 equivalent construct carrying the hygromycin resistance gene was used as minimal replicon; the control DNA was pUH4, which carries all of pAL5000 and a Hy^R gene (7).

pDQ263 was integrated into *M.smegmatis* cells by electroporation; the transformants were made competent anew and transformed with pUH4 or pDQ265. As a control, pUH4 and pDQ265 were introduced into wild-type *M.smegmatis* cells.

Wild-type *M.smegmatis* cells when transformed with pDQ265, needed 5 days to form colonies, as in the case of pUH52. Cells with pDQ263 integrated into the chromosome needed 4 days to form colonies when subsequently transformed with pDQ265; this was a slight improvement over wild-type, but we never saw colonies after 3 days. In contrast, transformation with pUH4 readily yielded colonies in 3 days for both wild-type and pDQ263-transformed cells.

Rap was expressed as a Histidine-tagged recombinant protein using the pBAD*Myc*His system and purified Rap protein was tested for DNA-binding activity with different templates, either the *ori* region as used for RepB (8) or the *rap*-encoding DNA upstream of the L-site. However, no DNA-binding activity was observed for Rap to any of these templates. ATP and different ions (Mg^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+}, K^{+}) had no effect.

In gel-retardation assays with RepB, a higher affinity (10–50-fold) for RepB to the *ori* was observed when Rap was included in the mixture in equimolar amounts or higher (not shown). However, RepB showed the same increased affinity for the *ori* when incubated with BSA in similar concentrations to those used for Rap. Similar effects have been observed for a putative integration host factor of mycobacteria though in that case no BSA control was reported (22).

DISCUSSION

This work reports the definition of a replication-assisting gene (*rap*) on the *Mycobacterium* plasmid pAL5000, transcribed from a promoter in reverse to and overlapping the promoter driving transcription of the replication protein genes *repA* and *repB*. This double promoter was dissected in an extensive mutational study which showed the requirement for two motifs: a –10 element, and one element situated around position –35. These numbers are not quite valid for the *rap* promoter, but this seems to be owing to an unusually short distance between the elements and the start point of *rap* transcription.

Mycobacteria transformed with constructs carrying *rap* recover significantly faster than cells transformed with the pAL5000 minimal replicon. The differences in growth rate for different constructs cannot be attributed to differences in plasmid copy number, since no significant such differences have been observed between constructs (7). *Mycobacterium smegmatis* cells transformed with the truncated construct pUH96, encoding the first 95 amino acids of the Rap ORF showed an intermediary behaviour between those transformed with pUH83 (carrying all of *rap*) and those with pUH52 (which codes for the first 42 amino acids of *rap*). Adding more sequences upstream of *rap* had no additional effect and deleting the region with repeated sequence features between *rap* and the L-site was not deleterious to replication. The role of the structural features in this region remains undefined.

Since replicons lacking *rap* are still functional, either the product is non-essential or, more intriguingly, Rap can be substituted for by a host factor at lower efficiency. Searches in the GenBank database yielded no genes with significant similarity to *rap*, which indicates that any similarities must be on a higher level than the amino acid sequence.

Figure 4. Sequence alignment of the *rap* and *repA/B* promoters to some mapped mycobacterial promoters. The L5P1 and L5P2 are the promoters P1 and P2 from the *Mycobacterium* phage L5 (13); the other promoters are from Bashyam and Tyagi (10). Transcription starts on the last nucleotide in each sequence. Suggested –10 boxes (10) are underlined; for *rap* and *repA/B*, proximal and distal motifs suggested by the mutational assay are underlined.

The function of Rap could not be unravelled in this work. Providing Rap in *trans* had only a moderate effect on the replication of a pUH52-equivalent replicon. Hence, there seems to be little scope for using *rap* in *trans* to boost replication of mycobacterial plasmids.

The increased affinity of RepB for its target sites when incubated in the presence of purified Rap could be mimicked by using BSA instead of Rap. Thus, although this stabilising would be an appealing role for Rap, we would hesitate to see it as its main effect *in vivo*. We have not pursued the stabilisation studies further, but note that a similar enhancing activity has been reported for a mycobacterial IHF (22).

It is possible that Rap interacts with RepA, which has so far proved refractory to overexpression in *E.coli*. Rap and RepB both *trans*-activated the reporter constructs when used as baits in the HybridHunter (Invitrogen) yeast two-hybrid system (Q.Z. and P.S., unpublished) which makes such interaction studies difficult.

The 5′ end of the *rap* transcript was mapped in a primer extension analysis to a T, bp 4595. The Rap ORF is preceded by a 143 nt untranslated region, in contrast to the single nucleotide between the 5′ end and the translational start codon on the *repA/B* mRNA. The *rap* leader sequence is transcribed through the L-site. Constructs carrying the L-site alone are unable to replicate when the *repA* and *repB* gene products are supplied in *trans* (P.S., unpublished), an observation which might indicate that transcription from the *rap* promoter into the L-site is required as a means to open up the DNA helix to enable the initiation of replication. Since most mutations to the *rap* promoter would also affect RepB binding and/or the *repA/B* promoter, testing this is not a straightforward task. Mycobacterial *oriC* regions are located close to the transcription units *rpnA-rpmH-dnaA-dnaN* (23) and in *M.smegmatis* the *oriC* is situated between the *dnaA* and *dnaN* genes (24), implying that the degree of supercoiling of the DNA helix and thus function of *oriC* is affected by transcription from these two genes.

In the *xylE* assay, RepB on a replicon in *trans* repressed both *rap* and *repA/B* transcription, though in both cases there was a residual *xylE* activity even when RepB was present. The control DNA pDQ56 is almost inactive (0.07 mU/mg) in the assay while the repressed promoters showed values 10 times above this background level. This supports the model for pAL5000 replication where RepB represses its own synthesis through binding to the H-site but where a low-level read-through allows for a slow build-up of a RepB pool in the cells until the concentration is high enough for RepB to bind to the L-site and trigger replication. Of course this coarse model will have to be refined as new data become available.

Divergent promoters are a very common arrangement (reviewed e.g. in 25), but the extensive symmetry of the *rap* and *repA/B* promoters is unusual and indicated that the shared sequence motifs were important for promoter function in both directions. The mutational assay corroborates this, though it also showed that the promoters are less similar than the extensive symmetry would imply. Several mutations with dramatic effects on one promoter had little effect when introduced into the other sequence. It does not seem to be a case of mutual dependence; i.e. a down-mutation to one promoter did not necessarily lead to an up-regulation of the oppositely orientated one.

Comparing the *repA/B* and *rap* promoters with other described mycobacterial promoters, there are many similarities in the sequences (Fig. 4). The best fit is achieved if one does not try to align the AT-boxes of the two pAL5000 promoters, which indicates that these boxes may not be important as promoter elements. The most unusual feature is the start point of *rap* transcription, which is much closer to the suggested promoter elements than is the case for any other promoter. A demonstration that this distance is close to what is allowed is the fortuitous 2 bp deletion in pDQ215 which abolishes promoter activity.

Allowing for this oddity, the *rap* promoter can be compared to the P1 promoter from the *M.smegmatis* phage L5 (13). Without introducing any gaps into the sequences, not only can motifs from the proximal box be aligned to similar motifs in the P1 promoter, but the P1 promoter turns out to have similarities to the *rap* distal motif. Both promoters have an (A+T)-rich stretch between the motifs. The spacing between the proximal and distal boxes are no different for the *rap* promoter than for the others. Fitting the *repA/B* promoter into the same mould would need more sequence juggling. In the proximal region all promoters are more or less similar to the putative ‘extended –10’ promoters in mycobacteria (10). The distal motifs show far less similarities; this greater flexibility in the –35 than in the –10 motifs has been suggested for other mycobacterial promoters (11).

The up-mutation +14 for *rap* makes this promoter potentially useful for mycobacterial gene expression. The single-copy activity for the +14 promoter, which in addition can be repressed by RepB, is 164 mU/mg. In comparison, the activity of the *hsp60* promoter, measured when present on a plasmid and thus in several copies, is ~400 mU/mg (20).

In conclusion, the 120 bp *ori* region comprising the L-site, the H-site, the *rap* and *repA/B* promoters and the regions where the respective transcription starts, is one of the most biologically active regions described for mycobacteria. On these few base pairs, a large set of proteins must jostle for positions: two copies of the transcription machinery, at least three copies of RepB, the DNA replication machinery, in all probability RepA in however many copies, and possibly the Rap protein as well. Untangling the mechanistics behind the trafficking on this molecular bottleneck promises to be a fascinating task.

REFERENCES

1 Rauzier,J., Moniz-Pereira,J. and Gicquel-Sanzey,B. (1988) *Gene*, **71**, 315–321.
2 Snapper,S.B., Lugosi,L., Jekkel,A., Melton,R.E., Kieser,T., Bloom,B.R. and Jacobs,W.R.,Jr (1988) *Proc. Natl Acad. Sci. USA*, **85**, 6987–6991.
3 Ranes,M.G., Rauzier,J., Lagranderie,M., Gheorghiu,M. and Gicquel,B. (1990) *J. Bacteriol.*, **172**, 2793–2797.

4 Stover,C.K., de la Cruz,V.F., Fuerst,T.R., Burlein,J.E., Benson,L.A., Bennett,L.T., Bansal,G.P., Young,J.F., Lee,M.H., Hatfull,G.F., Snapper,S.B., Barletta,R.G., Jacobs,W.R.,Jr and Bloom,B.R. (1991) *Nature*, **351**, 456–460.
5 David,M., Lubinsky-Mink,S., BenZvi,A., Ulitzur,S., Kuhn,J. and Suissa,M. (1992) *Plasmid*, **28**, 267–271.
6 De Smet,K.A.L., Jamil,S. and Stoker,N.G. (1993) *Gene*, **136**, 215–219.
7 Stolt,P. and Stoker,N.G. (1996) *Microbiology*, **142**, 2795–2802.
8 Stolt,P. and Stoker,N.G. (1996) *J. Bacteriol.*, **178**, 6693–6700.
9 Stolt,P. and Stoker,N.G. (1997) *Nucleic Acids Res.*, **25**, 3840–3846.
10 Bashyam,M.D. and Tyagi,A.K. (1998) *J. Bacteriol.*, **180**, 2568–2573.
11 Bashyam,M.D., Kaushal,D., DasGupta,S.K. and Tyagi,A.K. (1996) *J. Bacteriol.*, **178**, 4847–4853.
12 Kenny,T.J. and Churchyard,G. (1996) *J. Bacteriol.*, **178**, 3564–3571.
13 Nesbit,C.E., Levin,M.E., Donnelly-Wu,M.K. and Hatfull,G.F. (1995) *Mol. Microbiol.*, **17**, 1045–1056.
14 Raleigh,E.A., Murray,N.E., Revel,H., Blumenthal,R.M., Westaway,D., Reith,A.D., Rigby,P.W.J., Elhai,J. and Hanahan,D. (1988) *Nucleic Acids Res.*, **15**, 1563–1575.
15 Snapper,S.B., Melton,R.E., Mustafa,S., Kieser,T. and Jacobs,W.R.,Jr (1990) *Mol. Microbiol.*, **4**, 1911–1919.
16 Baulard,A., Jourdan,C., Mercenier,A. and Locht,C. (1992) *Nucleic Acids Res.*, **15**, 4105.
17 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Second edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbor, NY.
18 Lydiate,D.J., Ashby,A.M., Henderson,D.J., Kieser,H.M. and Hopwood,D.A. (1989) *J. Gen. Microbiol.*, **135**, 941–955.
19 Mudd,E.A., Prentki,P., Belin,D. and Krisch,H.M. (1988) *EMBO J.*, **7**, 3601–3607.
20 Curcic,R., Dhanhayuthapani,S. and Deretic,V. (1994) *Mol. Microbiol.*, **13**, 1057–1064.
21 Peña,C.E.A., Stoner,J.E. and Hatfull,G.F. (1996) *J. Bacteriol.*, **178**, 5533–5536.
22 Pedulla,M.L., Lee,M.H., Lever,D.C. and Hatfull,G.F. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 15411–15416.
23 Salazar,L., Fsihi,H., de Rossi,E., Riccardi,G., Rios,C., Cole,S. and Takiff,H.E. (1996) *Mol. Microbiol.*, **20**, 283–293.
24 Quin,M., Madiraju,M.V.V.S., Zachariah,S. and Rajagopalan,M. (1997) *J. Bacteriol.*, **179**, 6311–6317.
25 Beck,C.F. and Warren,R.A.J. (1988) *Microbiol. Rev.*, **52**, 318–326.