3840–3846 *Nucleic Acids Research, 1997, Vol. 25, No. 19*

Mutational analysis of the regulatory region of the *Mycobacterium* plasmid pAL5000

Pelle Stolt* and Neil G. Stoker

Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Received July 1, 1997; Revised and Accepted August 15, 1997

ABSTRACT

The regulatory region of the *Mycobacterium fortuitum* plasmid pAL5000 was studied *in vivo* and *in vitro* by mutational analysis. This region comprises the origin of replication for the plasmid and the start point of transcription for the *repA/B* genes, which encode the two replication proteins RepA and RepB. In this region there are two binding sites for RepB: a low-affinity site which is probably the origin of replication and a high-affinity-site which overlaps the promoter and implies an autoregulated expression of RepB. The high-affinity site contains two 8 bp palindromes, as well as an inverted repeat structure. By introducing point mutations into each of these motifs and monitoring changes to RepB binding in a gel-retardation assay, it was shown that the central, GC-rich palindrome (the GC-box) is the most important motif for protein binding. Mutations in the second, AT-rich palindrome (the AT-box) had no effect on protein binding and the inverted repeat structure *per se* was not needed, though some single-base changes affected binding to one or other of the DNA strands. These mutations were subsequently tested *in vivo* for their effects on plasmid replication in *Mycobacterium smegmatis*. Any change to the GC-box abolished replication, but changes to the other motifs were dependent on the position of the changed base, again indicating that the inverted repeats are not essential and that the AT-box is part of the promoter and not primarily recognised by RepB. The mutated plasmids did not show any changes in copy number to that of the wild-type. The expression of RepB was boosted by introducing a stronger promoter upstream of the *repA/B* genes. The resulting plasmid was capable of increasing to a degree in *trans* the copy number of other plasmids carrying the *ori* region, but was unstable when present on its own in *M.smegmatis*.

INTRODUCTION

The plasmid pAL5000 (1; GenBank Accession Number M23557) was first isolated from *Mycobacterium fortuitum* (2). It is the most studied mycobacterial replicon and constructs based on the pAL5000 origin are able to replicate in a wide range of mycobacteria. Several *Escherichia coli–Mycobacterium* shuttle vectors have been constructed based on this replicon (3–7).

We have previously shown (8–9) that the minimal functional replicon of pAL5000 comprises a *cis*-acting site, presumably the origin itself, and two genes, *repA* and *repB*, coding for replication proteins (Fig. 1A). The *repA* and *repB* genes overlap by 1 bp and are transcribed as a single RNA species (9). Plasmids carrying the origin but lacking the *repA/B* genes are able to replicate if those two genes are present in *trans*.

Both the RepA (277 aa) and RepB (119 aa) proteins show similarities to replication proteins from eubacterial plasmids; RepA is similar to the Rep proteins from ColEII-type plasmids (10) and RepB is similar to the product of an open reading frame (ORF2) from the plasmid pMB1 from *Bifidobacterium longum* (11).

RepA has not been studied in detail and its role in pAL5000 replication remains unclear. RepB has DNA-binding properties, binding specifically to two sites in the *ori* region (Fig. 1B). One site (the L-site) is the probable origin of replication; here RepB binds to one strand of the DNA helix only (9), possibly triggering replication in the process. The second binding site (the H-site) is immediately upstream of the *repA/B* promoter, where RepB binds in two copies, to autoregulate its expression. The affinity of RepB for the H-site is some 10-fold higher than that for the L-site. This dual role for a replication protein as autorepressor and initiator of replication is common in plasmids, e.g. P1, pSC101 and F (12–16).

There is no structural similarity between the H-site and the L-site. The H-site has two 8 bp palindromes as well as an inverted repeat of 5 bp outside these motifs, while there are no such structural motifs in the L-site (Fig. 1B). RepB binds to the H-site on both sides of the DNA helix in a staggered fashion, whereas the binding to the L-site is to one strand only, as indicated by DNaseI-footprinting (9).

The region comprising the RepB-binding sites and the promoter region of the *repA/B* genes is clearly an important regulatory region for the plasmid. This paper presents a more detailed investigation of the structure of the H-site. We have introduced mutations into the different motifs and have used gel-retardation assays to monitor changes to the RepB binding pattern. In addition, we have made plasmids carrying mutated H-sites and tested their ability to replicate in *Mycobacterium smegmatis*. The

*To whom correspondence should be addressed at present address: Division of Molecular Infection Biology, Research Centre Borstel, Parkallee 22, D-23845 Borstel, Germany. Tel: +49 4537 188 486; Fax: +49 4537 188 686; Email: pstolt@fz-borstel.de

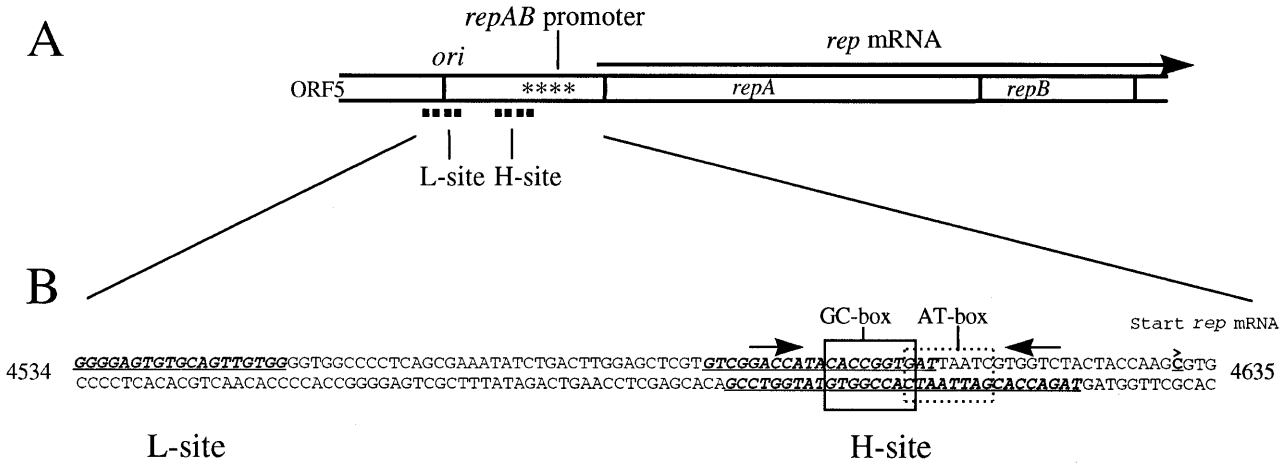


Figure 1. (**A**) Schematic map of the minimal replicon of pAL5000. The binding sites for RepB are marked with dotted lines. The *rep* promoter region is indicated with asterisks. The exact boundaries of the *ori* region have not been determined. The putative ORF5 gene product is not necessary for replication. (**B**) Sequence of the RepB-binding region. The numbering refers to the sequence of pAL5000 in GenBank (Accession No. M23557). The bases protected from DNaseI cleavage by RepB binding are boldfaced and underlined. The start point of transcription for the *rep* mRNA is shown as an arrowhead above an underlined bold base. The AT-box is boxed with dotted lines; the GC-box is boxed with plain lines. The inverted repeat structure in the H-site is marked with arrows.

effect on plasmid copy number of increasing the *repA/B* expression was also investigated.

MATERIALS AND METHODS

Materials

Escherichia coli strain DH5α (17) was used throughout to manipulate plasmid DNA. Constructs created in this study are shown in Table 4. *Escherichia coli* cells were grown in TY medium (16 g tryptone, 10 g yeast extract/l) with or without the addition of kanamycin (Km; final concentration 50 µg/ml), ampicillin (Ap; 50 µg/ml), or chloramphenicol (Cm; 40 µg/ml). *Mycobacterium smegmatis* strain mc^{2}155 (18) was grown in Lemco medium (Difco) or on Lemco agar plates.

DNA extraction

Plasmid DNA was isolated from *E.coli* cells by standard procedures (19). For large-scale plasmid preparations, Wizard midipreps (Promega) were used. Mycobacterial plasmid DNA was extracted through electroduction (20) into *E.coli* cells followed by standard DNA preparations.

Electroporation

Competent *M.smegmatis* cells were prepared as described by Snapper *et al.* (19). Transformation with 0.1–1 µg DNA was performed using 300 µl aliquots of cells.

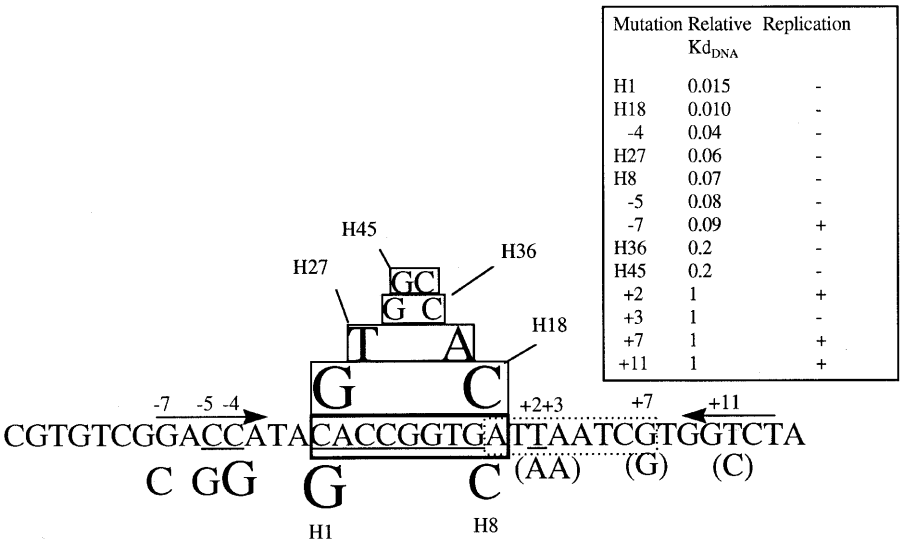
Copy number determination

Relative plasmid copy numbers were determined as single cell resistance (SCR) to Km (21) as described earlier (8).

Table 4. DNA constructs used in the investigations described in this work

Plasmid	Characteristics
pUH11	pUH4 with *Asc*I-deletion (152–1687) Δ*repA/B*
pUH12	pYUB12 with *Asc*I-deletion (152–1687) Δ*repA/B*
pUH36	pAL5000 region 3875–752 (*repA*) in pUC18
pUH52	pAL5000 region 3875–1093 in pUC18
pUH61	pAL5000 region 3875–1093 in pUC18; Km^R (Tn*903*)
pUH77	pAL5000 region 3861–4837 in pUC18; Km^R (Tn*903*)
pDQ31	pAL5000 bp387to 4589 (L-site) in pUC18 *Sma*I-site
pDQ51	Wild-type H-site + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ52	H-site with mutation (–5) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ55	pDQ11 with *M.bovis* BCG Hsp60 promoter from pMV261as 200 bp fragment in *Eco*RV site upstream of H-site
pDQ57	H-site with mutation (–7) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ58	H-site with mutation (+2) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ59	H-site with mutation (+11) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ60	H-site with mutation (+3) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ61	H-site with mutation (H1) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ62	H-site with mutation (+7) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ63	H-site with mutation (H45) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ64	H-site with mutation (–4) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ65	H-site with mutation (H8) + *repA/B* in pDQ31+ Km^R (Tn*903*)
pDQ66	Hsp60 promoter + H-site and *repA/B* from pDQ55 as *Xba*I cassette in pDQ11
pDQ71	pDQ66 + Km^R (Tn*903*)

The numbering refers to the pAL5000 sequence GenBank M23557 (1). The plasmid pYUB12 was described previously (18). The constructs in the pUH series have been described previously (8). All pDQ constructs were made for the present study.



Mutation	Relative Kd_{DNA}	Replication
H1	0.015	-
H18	0.010	-
-4	0.04	-
H27	0.06	-
H8	0.07	-
-5	0.08	-
-7	0.09	+
H36	0.2	-
H45	0.2	-
+2	1	+
+3	1	-
+7	1	+
+11	1	+

Figure 2. Mutational analysis of the H-site. The 5′ end of the sequence shown corresponds to the 5′ end of the oligonucleotides used in the PCR reactions to make templates for RepB binding and replicons with mutated H-sites. The GC- and AT-boxes are boxed by a plain and dotted line respectively. The arrows mark the inverted repeat structure. The different mutations introduced are shown above and below the sequence. The boxed bases are mutations where pairs of bases were changed in order to preserve the palindromic structure of the GC-box. The numbering of the mutations are H1–H8 for bases inside the GC-box, positive numbers for bases downstream of the GC-box and negative numbers for bases upstream of the GC-box. The sizes of the mutated bases shown are roughly proportional to the impact of the base-change on DNA binding. Bases essential for DNA replication are underlined. The numeric value given in the table is the Kd_{DNA} for the binding of RepB to the mutated target relative to wild-type which is set to unity.

DNA-binding assays

Templates for DNA-binding assays with mutated H-sites were prepared in PCR reactions using forward oligonucleotides based on the sequence 5′ CGTGTCGGACCATACACCGGTGATTAA 3′ (oligonucleotide OKDH; 9) but with mutations corresponding to those shown in Figure 2. The two exceptions were the oligonucleotides used to introduce the mutations +7 (5′ CGTGTCGGACCATACACCGGTGATTAATGGTGG 3′) and +11 (5′ CGTGTCGGACCATACACCGGTGATTAATCGTGCTCT 3′). The 5′ ends of the oligonucleotides were radioactively labelled using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (Promega). The oligonucleotide OFP1 (5′ GAGCAGATCGTCGCTTGCCA 3′; 9) was used as reverse primer; all templates are 118 bp in size. *Pfu* DNA polymerase (Stratagene) was used in all PCR reactions. The expression of recombinant RepB protein as an MBP-fusion protein and the demonstration of specific binding by the purified RepB protein to the H-site have been reported previously (9). The binding assays were carried out as described earlier (a 15 min incubation at 37°C in a buffer consisting of 100 mM Tris–HCl pH 8.0, 10 mM dithiothreitol, 20 mM $MgCl_2$; 1 µg salmon-sperm DNA per reaction; complexes separated by electrophoresis on a 6% non-denaturing polyacrylamide gel) and the calculation of Kd_{DNA} was done as described. All binding experiments were carried out in triplicate with a wild-type titration in parallel to each assay, in order to minimise dilution errors. Templates for the L-site were prepared as described previously (9).

In the experiments with RepB binding to the H-site, the RepB–MBP fusion protein was used. The binding pattern for this protein has been shown to be identical to the pattern for RepB with the MBP moiety cleaved off (9).

Construction of plasmids with mutated regulatory region

The construct pDQ31, carrying the L-site but not the H-site, was created by amplifying the region between bp 3871and 4589 of pAL5000 using the primers ORF1E (5′ GCGCGATATCGAGCCGAGAAC 3′) and OKDB (5′ ACGAGCTCCAAGTCAGATAT 3′; 9), which were treated with 5 U T4 polynucleotide kinase (PNK; Promega) and ATP (1 mM) for 30 min prior to the PCR reaction. The PCR product was ligated into *Sma*I-digested pUC18.

The wild-type control region of pAL5000 coding for the *repA/B* genes and the H-site, but lacking the L-site, was amplified in a PCR reaction using the oligonucleotides OΔL (5′ CAGCGAGATATCTGACTTGGAGCT 3′) and ORF2B (CGACACCGGATCCCCAATTGCGTTA; 9). These oligonucleotides introduce an *Eco*RV and a *Bam*HI site, respectively and the PCR product was cloned in pBSKS– (Stratagene) in the *Eco*RV–*Bam*HI sites, creating pDQ11.

To introduce mutated binding-sites into the regulatory region, the oligonucleotides described for the gel-retardation assay were used as forward primers in PCR reactions with ORF2B as reverse primer to amplify the mutated H-sites together with the *repA/B* genes. The forward primers were kinased with ATP and PNK and the products cloned in the *Eco*RV–*Bam*HI sites of pBSKS–.

DNA sequencing of the mutated constructs was done with the modified dideoxy-chain termination method (22) using the Auto Read sequencing kit and the Automatic Laser Fluorescent DNA sequencer (Pharmacia).

To add the L-site to these constructs, the cloned regions were ligated to the vector pDQ31 as blunt/*Bam*HI fragments (after cutting with *Hin*dIII, filling in with the Klenow polymerase [Promega] and cutting with *Bam*HI) in the *Bam*HI site and the filled-in *Xba*I site of pDQ31.

The Km^R gene from Tn*903* (23,24) was added to all the final constructs on a *Bam*HI–*Bgl*II fragment ligated to the *Bam*HI site in the replicons (downstream of the *repA/B* genes) in such an orientation that transcription from the Km^R gene was away from *repA/B* as not to interfere with the replicon. The resulting series of mutated replicons is listed in Table 4.

The *repA/B*-overexpressing construct pDQ66 was made by introducing the Hsp60 promoter from *M.bovis* BCG (25) which was cut out from the plasmid pMV261 (5) on a *Pvu*II–*Eco*RV fragment and ligated to the *Eco*RV site of pDQ11. The cassette was then cut out from the vector with *Xba*I and placed in the *Xba*I site of pDQ31 in the same orientation as in the wild-type plasmid. Adding the Km^R gene on a blunt-ended *Bam*HI–*Bgl*II fragment in the blunted *Hin*dIII site downstream of *repA/B* created pDQ71. Again, transcription from the Km^R gene was away from the *repA/B* genes.

RESULTS

Mutational analysis of the H-site

The H-site overlaps the promoter region of the *repA/B* promoter (9; Fig. 1B) and RepB binding here presumably serves to autoregulate its expression. The regions protected from DNaseI cleavage are staggered and DNA-binding assays indicate that two molecules of RepB bind to this site in a cooperative fashion (9). There are three notable structural features in the H-site: two 8 bp palindromes and one 5 bp inverted repeat. The palindrome GATTAATC (the AT-box) was suspected on grounds of its high (A+T)-content, to be a feature of the promoter, while the GC-rich and overlapping palindrome CACCGGTG (the GC-box) might be a recognition sequence for RepB.

To test these assumptions, a set of oligonucleotides was constructed, with specific mutations either in one or the other palindrome or in the repeats. In the GC-box, pairs of bases were exchanged, to keep the palindromic structure intact. Some single-base changes were also made to this box. Base changes were always C→G and A→T or the reverse. These oligonucleotides were used in PCR reactions with the oligonucleotide OFP1 as reverse primer and the products were used in gel-retardation assays using purified RepB protein.

To score for the relative importance to binding affinity, we titrated RepB onto the target and calculated the Kd_{DNA} for the respective templates as described previously (9). These values were compared with the binding-constant for RepB to the wild-type H-site in parallel reactions and the relative binding constants determined.

The results of these assays are shown in Figure 2. The mutation with the most dramatic effect was the changing of the first C of the GC-box to a G (mutation H1). This led to a virtual abolition of binding, with the affinity reduced almost a 100-fold. Changing both the C and the final G of the box, preserving the palindrome, practically abolished binding (mutation H18; Fig. 3A). The final G of the GC-box, when changed to a C (mutation H8), had a less strong, but still pronounced effect.

Changing the other bases in the GC-box had far less impact on the binding (Fig. 2). These bases were only changed in pairs, in order to preserve the palindrome. Mutations to the four central bases of the GC-box had little effect (5-fold reduction) on RepB binding ability.

The AT-box was much less important for RepB binding than the GC-box. None of the three single-base changes introduced (mutations +2, +3 and +7), had any significant effect. The changing of bases even closer to the transcription start, though destroying the inverted repeat structure (mutation +11), produced a wild-type binding pattern. The upstream sequence of the inverted repeat was more important to binding than the downstream sequence with base-changes reducing the binding 10–25-fold (Fig. 2).

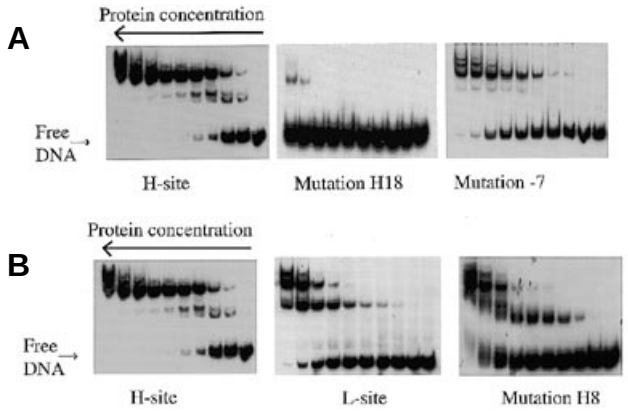


Figure 3. (**A**) RepB-binding patterns to some mutated H-sites compared with wild-type. Each lane represents a two-fold dilution from the previous one, with the highest concentrations (2600 nM protein) in the leftmost lanes. Free template is marked by an arrow. (**B**) Similarities between RepB binding to the L-site and to the mutation H8. The template for the L-site was synthesised as described (9).

The binding patterns for RepB to the mutated targets –7, –5, –4 and H8 indicated that the protein bound in one copy only in contrast to the wild-type (Fig. 3A and B). At high protein concentrations, there is a slowly migrating band in the wild-type H-site binding pattern while at lower protein concentrations a band appears which migrates more rapidly, concomitant with a weakening of the slowly migrating band (Fig. 3A). The mutations –4, –5, –7 and H8 produced DNA-binding patterns with only one band or where the shift from slower to faster-migrating bands occurred at far higher protein concentrations. Indeed some of these patterns, most pronouncedly that for the mutation H8, were virtually indistinguishable from that of RepB to the L-site (Fig. 3B) where it is thought to bind in only one copy.

In vitro testing of H-site mutations

To test for the influence of the different motifs on the replication abilities of pAL5000, replicons were constructed with mutated regulatory sites by amplifying the *repA/B* region including the H-site in PCR reactions. The PCR products were cloned in pBluescript vectors. None of these constructs, which all lacked the L-site and thus the putative *ori*, was able to replicate in *M.smegmatis*. The defective replicons were then excised and introduced into a pBluescript vector carrying the L-site but not the H-site (pDQ31) and the resulting constructs were electroporated into *M.smegmatis.* As a control, we made similar constructs where the amplification was performed using the oligonucleotide OΔL which produces a wild-type regulatory region, still lacking the L-site. This wild-type construct did not replicate in *M.smegmatis*, but when fused to the L-site in pDQ31, the resulting construct (pDQ51) was viable.

The results of the *in vivo* assay are summarised in Figure 2. All the changes to the GC-box which we tested abolished the replicative ability of the plasmid. In contrast, the mutations +2

and +7 to the AT-box produced viable plasmids. The –4 and –5 mutations, however, abolished replication. The mutation +11, which destroys the inverted-repeat motif downstream of the AT-box, yielded a viable plasmid.

The mutation –7 yielded a viable plasmid, although the transformation efficiency of the construct (pDQ57) was consistently ten times lower than that of the other constructs. The other viable mutants transformed with an efficiency comparable with that of pDQ51. There was no difference in growth rate between cells carrying pDQ57 and those carrying pDQ51. Copy-number determinations using the method of Nordström (21) showed no significant differences between pDQ51 and the mutant plasmids, including pDQ57 (not shown).

It took 4–5 days for cells transformed with any of these plasmids, including pDQ51, to form colonies on Km plates, in contrast to cells transformed with the shuttle vector pYUB12, which carries all of pAL5000 (5). Such cells typically form colonies after 3 days. This indicated that the initial expression of *repA/B*, immediately upon transformation, was less efficient in the manipulated plasmids. This conclusion was further supported by co-transformation experiments where the construct pUH11 (8) was electroporated into *M.smegmatis* together with pYUB12 or the pDQ series of constructs. The plasmid pUH11 carries the hygromycin resistance gene from *Streptomyces hygroscopicus* (26) and lacks much of *repA* and all of *repB* and thus is unable to replicate on its own but can be activated in *trans*. Cells were readily co-transformed with the pair pUH11/pYUB12, but none of the pDQ series would support replication, not even pDQ51, which carries the wild-type H-site. However, when cells were first transformed with pDQ51 or one of the viable mutant constructs, and then transformed with pUH11 in a second step, this construct could be introduced with high efficiency. This supported the notion that the initial *repA/B* expression is important for plasmid viability.

Effect of *repA/B* expression on plasmid copy number

Since the experiments above showed the importance of a sufficient level of RepA and/or RepB in the cells for establishing a plasmid population, it was thought that one possible way of raising the copy number of the plasmid would be to increase the expression of the *repA/B* genes.

To boost *repA/B* expression we introduced the Hsp60 promoter from *M.bovis* BCG (25) into pDQ51 between the L-site and the H-site in such an orientation that the *repA/B* genes would be transcribed. We expected this construct (pDQ66) to have higher *repA/B* expression, since the Hsp60 promoter is very strong and not regulated by RepB.

This construct was tested for its ability to support replication in *trans*, by co-transforming *M.smegmatis* cells with pDQ66 and pUH77. The construct pUH77, which has been described earlier (8), carries a 1 kb region comprising the *ori*, as well as the Km^R gene from Tn*903* but lacks *repA/B*. There is only the Ap resistance marker on pDQ66 and so selecting on Km for pUH77 would co-select for pDQ66, since this plasmid is necessary for pUH77 to replicate. As a control, pUH77 was transformed together with pUH56 (8), which carries a wild-type *ori* and *repA/B* and has wild-type pAL5000 replication characteristics, but lacks a Km resistance marker.

The co-transformation efficiency for the pair pDQ66/pUH77 was 1–5% that of the wild-type pair pUH56/pUH77. The spread in SCR values to Km for pDQ66/pUH77 was greater than that for pUH56/pUH77 (Fig. 4), but the values were always higher than for the wild-type. Thus, the increased amount of RepA and/or RepB in the cells carrying pDQ66 seems to have a positive effect on the copy number of the activated pUH77. The increase in copy-number estimated from SCR measurements was not more than 1.4–2-fold.

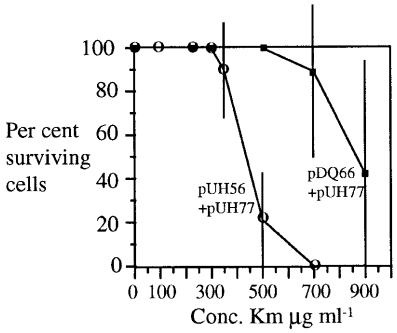


Figure 4. Relative copy numbers measured by single-cell resistance (SCR) to Km. The values are means from six (pUH56+pUH77) and seven (pDQ66+pUH77) single colonies respectively. ○, pUH56+pUH77; ■, pDQ66+pUH77. The concentration of Km where the curves deviate from the 100% line are taken as the SCR value.

This low transformation frequency and the greater spread in SCR values indicated that the high level of replication protein(s) from pDQ66 was deleterious to the cells. To be able to select for the pDQ66 replicon in *M.smegmatis* cells, we introduced the Km resistance gene into this construct, creating pDQ71. The transformation effiency for pDQ71 alone was very low; lower than the co-transformation efficiency for the pair pDQ66/pUH77, and the transformants did not show higher resistance to Km. When the plasmid was recovered from these cells, restriction enzyme digestions revealed that deletions and rearrangements had occurred (not shown), indicating that the replicon with a too high expression of *repA/B* is not stable in *M.smegmatis*.

It was not clear from the above experiments which gene product, RepA or RepB, was deleterious to the plasmid in too high a concentration. To investigate this, we did double transformations of *M.smegmatis* with two pairs of plasmids, pDQ66+pYUB285 and pDQ71+pUH36. The plasmid pYUB285 carries the pAL5000 minimal replicon but has a deletion in *repA* which makes it non-replicating (27); however, it carries *repB*. This construct has a Km^R gene and so pDQ66 was used as helper plasmid. The construct pUH36 (8) carries the *ori* and *repA* but lacks *repB*. It lacks a Km^R gene [there is a typing error in the paper by Stolt and Stoker (8)] and thus pDQ71 was used as helper in this case. Neither pYUB285 nor pUH36 can replicate on its own and we had shown above that pDQ71 (and thus also pDQ66) do not replicate in *M.smegmatis*. Thus, transformants with any of these pairs able to grow on kanamycin would indicate that the pDQ plasmid had been stabilised by the second replicon.

The pair pDQ66+pYUB285 did not transform *M.smegmatis*, whereas pDQ71/pUH36 did. This indicates that the cells can tolerate high levels of RepA but not of RepB. The frequency of transformation was ~10% of that for the pair pUH56+pUH77 and about twice as high as for pDQ66+pUH77.

DISCUSSION

This paper presents a dissection of the regulatory region of the *M.fortuitum* plasmid pAL5000. Central to this region is the so-called H-site, where the replication protein RepB binds with high affinity. The structure of the H-site was probed by specific mutations to bases in the three different structural motifs present. These motifs are: a GC-rich palindrome (GC-box), an AT-rich palindrome (AT-box) and a 5 bp inverted repeat. The mutations were tested for changes to RepB binding *in vitro* as well as for plasmid viability *in vivo*.

The integrity of the GC-box was shown to be crucial to replication. All changes to this box produced non-replicating plasmids. The initial C of the GC-box was the most important single base in the H-site for RepB binding. If this base was changed to a G, the binding constant dropped by two orders of magnitude. Changing the first and last nucleotides, keeping the palindromic structure of the box intact, virtually abolished binding. Other bases were less important for binding; pairwise changes to the other bases in the box, keeping the palindrome intact, reduced RepB binding between 5- and 25-fold.

Mutations to the AT-box sometimes abolished replication but in no case did they affect RepB binding *in vitro*, which indicates that this box is indeed part of the promoter structure, rather than of a recognition motif for RepB. The exception was the first G of the AT-box, but this base is also part of the GC-box and thus might have a dual role. Support for this role for the AT-box is given by the observation (9) that 11 bp of the *repA/B* promoter region, including the AT-box, can be found in the promoter region for the *Lactococcus lactis dnaE* gene (28).

If the AT-box is a part of the promoter, single-base pair changes might preserve a functional promoter whilst leaving the RepB binding unaffected. The binding sites for RepB, as defined by DNaseI-footprinting experiments, are staggered (Fig. 1B; 9) and the DNA region protected from DNaseI cleavage is probably larger than the area actually in contact with RepB. Thus, even though the protein would occlude the AT-box, it would not have to be in actual contact with the DNA in this region.

Apart from the GC-box, where all mutations resulted in loss of replication ability, observations of changes to the DNA-binding properties of a mutated motif *in vitro* could not be used to predict whether the change would lead to a replicating or a non-functional plasmid *in vivo*. Thus only two out of three mutations to the upstream sequence of the inverted repeat abolished replication, though the effects on RepB binding were of a similar magnitude. No changes downstream of the GC-box had any effect on RepB binding. The mutation +11 produced a wild-type binding pattern *in vitro*, while the corresponding mutation in the upstream motif, mutation –5, showed reduced RepB binding and could not support replication. This suggests that the inverted repeat structure is not important as such, but the upstream bases act as part of a binding site. Indeed, the mutation –7, while reducing Kd_{DNA} >10-fold and producing a binding pattern indicating that RepB only binds to one strand of this construct, nevertheless produced a replicating plasmid, albeit with reduced transformation efficiency. It is not clear why the other two mutations in this motif, while showing similar binding patterns *in vitro*, led to non-replicating plasmids.

The mutations H8, –7, –5 and –4 had an effect on the binding pattern of RepB which supports our argument that two copies of RepB bind in a cooperative fashion to the H-site. In particular, the

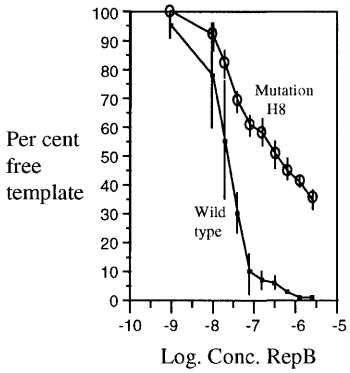


Figure 5. Binding curves for RepB to the wild-type H-site and the mutation H8. ■, wild-type; ○, mutation H8. The concentration of RepB at which 50% of the template is bound is an approximation of the Kd_{DNA} for the interaction.

binding pattern to the mutated target H8 was virtually indistinguishable from that to the L-site (Fig. 3B). From DNaseI-footprinting experiments we have shown that binding to the L-site is to one strand of the DNA helix only (9). If the mutations destroyed one binding motif of the H-site, such a pattern of only one protein molecule would be expected. The slope of the binding curve for the H8 mutation is less steep than that for the wild-type H-site (Fig. 5), which is also seen for binding to the L-site (9). This is another indication that there is no cooperative binding to the mutated site. Different binding patterns for replication proteins in dual roles as autorepressors and replication initiators have been shown for other plasmids; e.g. binding as autorepressor in dimeric form and as initiator as monomer (29–31) but whether this is the case for RepB remains to be determined.

This work shows that the initial expression of the *repA/B* genes is important to the biology of pAL5000. In co-transformation experiments, none of the plasmids created was able to support the replication of a second *ori* in *trans* (construct pUH11), even though the copy number of these mutated plasmids did not seem to differ from wild-type. When cells already carrying pDQ51 (wild-type H-site) or one of the viable mutant constructs were made competent, they could be transformed with pUH11 with high efficiency, in fact higher than for the pDQ vectors themselves. Thus, the limiting factor for the transformation efficiency of these constructs is the initial expression of *repA/B*. Once there is a pool of the Rep proteins in the cells, the amounts are sufficient to support replication of a second plasmid lacking these genes. In fact, the efficiency of transformation with pUH11 of cells already carrying pDQ51, was if anything greater than that for wild-type cells with pDQ51, indicating that the initial expression of *repA/B* is a contributing factor to the transformation efficiency of any pAL5000-based construct.

The construct pDQ51 is not strictly wild-type, since the cloning procedure has changed the number of bases between the H-site and the L-site, but its copy number is similar to that of the wild-type replicon pUH61 (8). Thus, the exact number of nucleotides between the H-site and the L-site seems not to be a deciding factor for plasmid viability. Indeed, in pDQ66, introducing the Hsp60 promoter means there are ~200 bp between the two sites and this construct still replicates in *M.smegmatis* if there is another *ori* present in *trans* to reduce the level of RepB (see below). The fine-tuning of *repA/B* expression in pAL5000 has not

been elucidated and it is probable that the region between the H-site and the L-site together with the *ori* is involved in this regulation.

Plasmids lacking the L-site altogether are unable to replicate. The smallest replicon constructed to date is pUH52 (8) which carries the H-site, L-site, *repA/B* and a further 210 bp upstream of the L-site. This limits the origin area to that between the H-site and the additional 210 bp present on pUH52, which strongly supports the notion that the L-site is indeed the origin of replication.

Raising the number of Rep proteins in the cells might be one way of increasing the copy number of the plasmid. We had initially naively hoped that mutations abolishing or severely decreasing RepB binding could force the protein to bind instead to the L-site, prematurely triggering replication of the plasmid and raising copy number. This turned out not to be the case; the strongly negative mutations were non-viable and the viable constructs all had copy numbers very similar to the wild-type pDQ51.

The construct pDQ71, with the Hsp60 promoter driving *repA/B* expression, was not stable in *M.smegmatis*; the transformation efficiency was negligible and plasmids recovered from any transformants we obtained had suffered deletions and rearrangements. Co-transformation experiments showed that plasmids carrying the replication region but lacking *repB* could stabilise the *rep*-overexpressing constructs, whereas *repB*-carrying plasmids could not. Thus, excessive levels of RepB block the replication of pAL5000. Similar deleterious effects of overexpressing replication proteins are seen in P1 (32). In the viable double transformants, the presence of extra binding sites on the second plasmid might serve to titrate RepB and keep the protein concentration at an acceptable level.

An *ori* activated by pDQ66 has a slightly higher copy-number than one activated by a wild-type plasmid (Fig. 4), but the small difference and the instability of pDQ66 argue against any widespread use of this plasmid to boost copy numbers. The small increase in copy numbers observed indicates that there are factors other than the RepA or RepB level that determine pAL5000 copy numbers.

ACKNOWLEDGEMENTS

This work was carried out as part of the Glaxo Wellcome Action TB initiative. We thank Ken Duncan for critically reading the manuscript.

REFERENCES

1 Rauzier,J., Moniz-Pereira,J. and Gicquel-Sanzey,B. (1988) *Gene* **71**, 315–321.
2 Labidi,A., David,H.L. and Roulland-Dussoix,D. (1985) *Microbiol. Lett.* **30**, 221–225.
3 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.
4 Ranes,MG., Rauzier,J., Lagranderie,M., Gheorghiu,M. and Gicquel,B. (1990) *J. Bacteriol.* **172**, 2793–2797.
5 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.
6 David,M., Lubinsky-Mink,S., BenZvi,A., Ulitzur,S., Kuhn,J. and Suissa,M. (1992) *Plasmid* **28**, 267–271.
7 De Smet,K.A.L., Jamil,S. and Stoker,N.G. (1993) *Gene* **136**, 215–219.
8 Stolt,P. and Stoker,N.G. (1996) *Microbiology* **142**, 2795–2802.
9 Stolt,P. and Stoker,N.G. (1996) *J. Bacteriol.* **178,** 6693–6700.
10 Hiraga,S.I., Sugiyama,T. and Itoh,T. (1994) *J. Bacteriol.* **176**, 7233–7243.
11 Rossi,M., Brigidi,P., Gonzalez Vara y Rodriguez,A. and Matteuzzi,D. (1996) *Res. Microbiol.* **147**, 133–143.
12 Linder,P., Churchward,G., Xia,G.X., Yu,Y.Y. and Caro,L. (1985) *J. Mol. Biol.* **181**, 383–393.
13 Rokeach,L.A., Sogaard-Andersen,L. and Molin,S. (1985) *J. Bacteriol.* **164**, 1262–1270.
14 Vocke,C. and Bastia,D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2252–2256.
15 Abeles, AL. (1986) *J. Biol. Chem.* **261**, 3548–3555.
16 Wada,C., Imai,M., and Yura,T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8849–8853.
17 Hanahan,D., Jessee,J. and Bloom,F.R. (1991) *Methods Enzymol.* **204**, 63–114.
18 Snapper,S.B., Melton,R.E., Mustafa,S., Kieser,T. and Jacobs,W.R.Jr. (1990) *Mol. Microbiol.* **4**, 1911–1919.
19 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*. 2nd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbor, NY.
20 Baulard,A., Jourdan,C., Mercenier,A. and Locht,C. (1992) *Nucleic Acids Res.* **15**, 4105.
21 Nordström,K. (1993) In K. G. Hardy, ed., *Plasmids: A Practical Approach*. 2nd edn. Oxford University Press, pp 2–38.
22 Zimmermann,J., Voss, H., Schwager, C., Stegemann, J., Erfle,H., Stucky,K., Kristensen,T. and Ansorge,W. (1990) *Nucleic Acids Res.* **18**, 1067.
23 Berg,D.E., Jorgensen,R, and Davies,J. (1978) In D. Schlessinger, ed., *Microbiology*-1978. Washington DC: American Society for Microbiology, pp. 13–15.
24 Nomura,N., Yamagishi, H. and Oka,A. (1978) *Gene* **3**, 39–51.
25 Thole,J.E.R., Keulen,W.J., Kolk,A.H.J., Groothuis,D.G., Berwald,L.G., Tiesjema,R.H. and VanEmbden,J.D.A. (1987) *Infect. Immunol.* **55,** 1466–1475.
26 Lydiate,D.J., Ashby,A.M., Henderson,D.J., Kieser,H.M. and Hopwood,D.A. (1989) *J. Gen. Microbiol.* **135**, 941–955.
27 McAdam,R.A., Weisbrod,T.R., Martin,J., Scuderi,J.D., Brown,A.M., Cirillo,J.D., Bloom,B.R. and Jacobs,W.R.Jr (1995) *Infect. Immunol.* **63**, 1004–1012.
28 Araya,Y., Ishibashi,N., Shimamura,S., Tanaka,K. and Takahashi,H. (1993) *Biosci. Biotech. Biochem.* **57,** 88–92.
29 Sugiura,S., Tanaka,M., Masamune,Y. and Yamaguchi,K. (1990) *J. Biochem.* **107**, 369–376.
30 Manen,D., Upegui-Gonzalez,L.C. and Caro,L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8923–8927.
31 Ishiai,M., Wada,C., Kawasaki,Y. and Yura,T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3839–3843.
32 Chattoraj,D.K., Snyder,K.M. and Abeles,A.L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2588–2592.