JOURNAL OF BACTERIOLOGY, Nov. 1995, p. 6527–6535 Vol. 177, No. 22
0021-9193/95/$04.00+0

Mycobacterium smegmatis dnaA Region and Autonomous Replication Activity

MALINI RAJAGOPALAN,[1,2] MING-HUI QIN,[1] DONALD R. NASH,[1,2] AND MURTY V. V. S. MADIRAJU[1]*

Department of Microbiology[1] and Center for Pulmonary Infectious Disease Control,[2] The University of Texas Health Center at Tyler, Tyler, Texas 75710

Received 6 April 1995/Accepted 11 September 1995

Two key elements that are thought to be required for replication initiation in eubacteria are the DnaA protein, a *trans*-acting factor, and the replication origin, a *cis*-acting element. As a first step in studying the replication initiation process in mycobacteria, we have isolated a 4-kb chromosomal DNA fragment from *Mycobacterium smegmatis* that contains the *dnaA* gene. Nucleotide sequence analysis of this region revealed homologies with the *rpmH* gene, which codes for the ribosomal protein L34, the *dnaA* gene, which codes for the replication initiator protein DnaA, and the 5′ end of the *dnaN* gene, which codes for the beta subunit of DNA polymerase III. Further, we provide evidence that when cloned into pUC18, a plasmid that is nonreplicative in *M. smegmatis*, the DNA fragment containing the *dnaA* gene and its flanking regions rendered the former capable of autonomous replication in *M. smegmatis*. We suggest that the *M. smegmatis* chromosomal origin of replication is located within the 4-kb DNA fragment.

The genus *Mycobacterium* includes both rapid growers (e.g., *Mycobacterium smegmatis* and *M. fortuitum*) and slow growers (e.g., *M. tuberculosis*, *M. bovis* BCG, and *M. leprae*). The doubling time for rapid growers is approximately 3 to 4 h, compared with 20 to 24 h for slow growers. Some of the slow growers, such as *M. tuberculosis* and *M. leprae*, are major human pathogens. *M. tuberculosis* is a well-recognized opportunistic infectious agent for immunocompromised patients (24), and recent years have seen the emergence of *M. tuberculosis* strains that are resistant to one or more traditionally used antimycobacterial drugs. Our inability to control mycobacterial infections is due to a limited understanding of the basic metabolic processes, such as DNA replication and recombination, in these pathogens (33, 36). An improved understanding of the key steps involved in the replication process of both rapid growers and slow growers would enable one to identify metabolic targets against which new generation drugs could be directed (2).

Initiation of replication is believed to occur when DnaA, the initiator protein, interacts with the replication origin called the replicon or the replicator (15, 17). Detailed genetic studies carried out in *Escherichia coli* (15, 17) and limited studies carried out in *Bacillus subtilis* (9, 20–22) revealed that DnaA protein is required for replication initiation. DnaA protein initiates replication by selectively binding to repetitive units of nine-nucleotide DnaA protein recognition sequences, present in the origin of replication, called the DnaA boxes. This process triggers a cascade of events which result in replication initiation (4, 15, 17). In many bacteria, the *dnaA* gene is flanked by the *rpmH* and *rnpA* genes on the 5′ side and the *dnaN*, *recF*, and *gyrB* genes on the 3′ side (15, 22, 28). Exceptions to this gene order have been found in *Rhizobium meliloti*, *Caulobacter crescentus*, and *Synechocystis* sp. (16, 25, 38). Comparative analyses of the amino acid sequences of DnaA proteins from different bacteria have revealed a conserved ATP binding motif, GxxGxGKT (18, 27, 32), and another motif, GRDHTT, present in the carboxyl-terminal region. The region between these two motifs of the DnaA protein is also conserved (28).

With the exception of *E. coli*, the *dnaA* gene flanking regions of all bacteria thus far examined function as autonomously replicating sequences. For example, both the 5′ and 3′ flanking regions of the *dnaA* gene of *B. subtilis* (20), the 5′ flanking region of *Pseudomonas putida* (35), and the 3′ flanking region of *Streptomyces lividans* (37), a close relative of mycobacteria, function as autonomously replicating sequences. As a first step in identifying the key players in the replication initiation process in mycobacteria, we have cloned and sequenced a 4-kb DNA fragment containing the *dnaA* gene and its flanking regions from *M. smegmatis*. In addition, we provide evidence that a pUC18 plasmid containing this DNA fragment, but not pUC18 alone, replicated stably in *M. smegmatis*. These data suggest that the *dnaA* gene region of the *M. smegmatis* chromosome has features that support autonomous replication.

MATERIALS AND METHODS

Bacterial strains, plasmids, and chromosomal DNA preparations. *E. coli* DH5α (*recA1*, restriction deficient and host modification proficient) (26) was used for propagating all plasmids and for screening *M. smegmatis* genomic DNA libraries for the *dnaA* gene. The *M. smegmatis* genomic DNA cosmid library was kindly supplied by William R. Jacobs, Jr., Albert Einstein College of Medicine, Bronx, N.Y. *M. smegmatis* mc^{2}155, a high-transformation-proficient strain (14), was used as a host in transformation experiments involving cloning of the putative autonomously replicating sequences of *M. smegmatis*. All plasmids, unless otherwise stated, were derived from pUC18 and are described in Table 1. Chromosomal DNA preparations of *M. smegmatis* mc^{2}155 were prepared as described by Husson et al. (13). Oligonucleotide primers used in the study were synthesized in a Pharmacia Gene Assembler (Pharmacia Biotech, Piscataway, N.J.).

Media and growth and transformation conditions. *E. coli* cultures were grown in Luria-Bertani (LB) medium (19). For selection of recombinant clones, LB agar plates containing ampicillin (50 μg/ml), kanamycin (50 μg/ml), or both were used (19, 26). *E. coli* cultures treated with rubidium chloride were used in transformation experiments (19, 26).

M. smegmatis cultures were grown in 7H9 broth containing 0.05% Tween 80, 0.5% bovine serum albumin, and 0.2% dextrose. Electrocompetent *M. smegmatis* cells were prepared as described previously (14). Competent cells were resuspended in 10% glycerol, aliquoted, and stored at −70°C. Electroporation was carried out in a Bio-Rad Gene Pulser (Bio-Rad Laboratories, Richmond, Calif.) at 2.5 kV, 25 μF, and 1,000 ohms. Following electroporation, 7H9 broth was

* Corresponding author. Mailing address: Department of Microbiology, The University of Texas Health Center, Tyler, TX 75710. Phone: (903) 877-7731. Fax: (903) 877-7652. Electronic mail address: Murty@jason.uthct.edu.

TABLE 1. Bacterial strains and plasmids used

Strain, plasmid, or vector	Genotype or description	Reference
Bacterial strains		
E. coli DH5α	*recA1 hsdR*	10
M. smegmatis mc^2155	Efficient transformation strain obtained from *M. smegmatis* ATCC 607	14
Plasmids		
pUC18	Ampicillin resistant	34
pMR1	777-bp PCR product from *M. smegmatis* genomic DNA cloned into pUC18; ampicillin resistant	This study
pMR25	*M. smegmatis dnaA* cosmid; kanamycin resistant	This study
pMR40	A 4-kb *Pvu*II fragment containing the *dnaA* gene isolated from pMR25 cosmid and cloned into pUC18 in the *Sma*I site; ampicillin resistant	This study
pMR41	Same as pMR40 except that the 4-kb *Pvu*II insert was cloned into pUC18 in the opposite orientation; ampicillin resistant	This study
pMR42	A 1.3-kb *aph* gene responsible for conferring kanamycin resistance was cloned into pUC18; ampicillin and kanamycin resistant	This study
pMR43	A 456-bp *Bam*HI fragment of pMR40 was replaced with a 1.3-kb fragment containing the *aph* gene; ampicillin and kanamycin resistant	This study
pYUB12	*E. coli-Mycobacterium* shuttle plasmid vector; kanamycin resistant	14
pYUB18	*E. coli-Mycobacterium* shuttle cosmid vector; kanamycin resistant	14
Vectors[a]		
pMQ1	Generated upon deletion of a 0.5-kb *Bam*HI fragment ($B\text{-}B_1$) from pMR40	
pMQ2	Generated upon deletion of a 0.7-kb *Hin*dIII fragment ($H\text{-}H_1$) from pMR40	
pMQ3	Generated upon deletion of a 1.0-kb *Eco*RI fragment from pMR41	
pMQ4	1.2-kb *Xho*I fragment ($X_2\text{-}X_3$) cloned into pUC18	
pMQ5	0.85-kb *Xho*I fragment ($X_4\text{-}X_5$) cloned into pUC18	
pMQ6	0.5-kb *Bam*HI fragment ($B\text{-}B_1$) from pMR40 cloned into pUC18	
pMQ7	0.8-kb *Sma*I fragment ($S_1\text{-}S_2$) cloned into pUC18	
pMQ10	Generated upon deletion of a 3.0-kb *Xho*I-*Hin*dIII fragment ($H\text{-}X_5$ deletion) from pMR40	
pMQ11	0.2-kb *Ava*I fragment ($A_1\text{-}A_2$) cloned into pUC18	
pMQ12	0.7-kb *Hin*dIII fragment ($H\text{-}H_1$) from pMR40 cloned into pUC18	

[a] Plasmids used for determination of the sequence of the 4-kb DNA fragment. For definitions of the restriction enzyme cutting sites indicated in parentheses, see the legend to Fig. 1.

added, and cells were incubated for 3 h at 37°C, plated on 7H10 agar plates containing kanamycin (50 μg/ml), and incubated at 37°C. The plasmid DNA was recovered from *M. smegmatis* by using the electroduction protocol (see below).

Cloning of the *dnaA* gene. To facilitate cloning of the *dnaA* gene of *M. smegmatis*, degenerate oligonucleotide primers MVM8 (5′-GGAGGCCTGGG BAARACNCACYT-3′) and MVM9 (5′-TCCCCGCGGTSGTRTGRTCRCG DCC-3′) (B = C or G or T; D = A or G or T; N = A or C or G or T; R = A or G; S = C or G; Y = C or T) were used to amplify a 777-bp internal fragment of the *dnaA* gene by PCR (26). MVM8 and MVM9 primer sequences were based, respectively, on the putative ATP binding motif (GxxGxGKT) and the conserved motif (GRDHTT) in the DnaA protein (28). The underlined sequences in MVM8 and MVM9 represent the *Stu*I and *Sac*II restriction endonuclease recognition sites, respectively, that were incorporated for cloning purposes. Amplification was carried out in a Coy thermocycler (Coy Corporation, Grasslake, Mich.), using *Taq* DNA polymerase (Roche Molecular Systems, Inc., Branchburg, N.J.) and a PCR optimizer kit (Invitrogen Corporation, San Diego, Calif.) under the following cycling conditions: initial denaturation for 5 min at 95°C, followed by 6 cycles of amplification (1 min each of denaturation at 94°C, annealing at 55°C, and elongation at 72°C) and then 30 cycles of 1 min each of denaturation at 94°C, annealing at 60°C, and elongation at 72°C. A final elongation cycle of 10 min at 72°C was carried out to complete the synthesis of all amplified DNA strands.

The 777-bp PCR product was kinase treated and cloned into pUC18 by using standard protocols (26). Recombinant plasmid DNA was isolated, and the nucleotide sequence of the insert was determined in a Pharmacia Automated Laser Fluorescent (ALF) DNA sequencer, using autocycle and autoread sequencing kits. The DNA sequence thus obtained was compared with the sequence of the *E. coli dnaA* gene. A 33-mer oligonucleotide, MVM37 (A**AC**AA**GAC**CCC**GC**A TCG**ACC**GGTCG**CT**GGCC**GAG**), that showed 14 similarities (indicated in boldface) with the *E. coli dnaA* gene sequence was identified. The 33-mer oligonucleotide was synthesized in a Pharmacia DNA synthesizer, end labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase (26), and used to screen *M. smegmatis* genomic DNA libraries previously propagated in *E. coli* DH5α in a colony hybridization experiment (26). Hybridizations were carried out at 65°C, and membranes were washed at 65°C under highly stringent conditions (26). Cells containing the *dnaA* cosmids were identified and colony purified following a second screening. The identity of the *dnaA* gene in the positive cosmids was further established by two methods: (i) amplifying the 777-bp *dnaA* gene fragment by using MVM8 and MVM9 primers and (ii) probing the restriction enzyme-digested *dnaA* positive cosmid DNA with a radiolabeled 777-bp PCR product in a Southern hybridization experiment (26). In the latter experiment, the 777-bp PCR product was labeled by nick translation using [α-^{32}P]dCTP and nonradiolabeled dATP, dGTP, and dTTP (26). One recombinant cosmid, called pMR25, was selected and used for subcloning the *dnaA* gene-containing fragment.

Subcloning and nucleotide sequence determination. The pMR25 cosmid DNA was digested with *Pvu*II and electrophoresed on 0.8% agarose gels. Following electrophoresis, DNA fragments were transferred to Zetaprobe nylon membranes (Bio-Rad) by the capillary transfer method and probed with the ^{32}P-labeled 777-bp PCR fragment by Southern hybridization (26). A 4-kb DNA fragment was identified by autoradiography, and a corresponding DNA fragment was isolated from agarose gels, run under identical conditions, and gel purified by using a Bio-Rad PrepA gene purification kit. The purified fragment was ligated to pUC18 DNA previously digested with *Sma*I. This process eliminated the *Pvu*II site of the 4-kb DNA fragment. The ligation mixture was then used to transform *E. coli* DH5α. Ampicillin-resistant white colonies were identified, recombinant plasmid DNA was isolated, and the presence of the *dnaA* gene in the insert was established as described above. The pUC18 plasmids containing the insert in both orientations (pMR40 and pMR41) were selected and subsequently used to determine the nucleotide sequence of the 4-kb insert (Fig. 1).

The pMR40 and pMR41 DNAs were digested with restriction enzymes *Ava*I, *Bam*HI, *Hin*dIII, *Sma*I, *Eco*RI, and *Xho*I, individually and in various combinations. On the basis of restriction digest patterns, a physical map of the 4-kb insert region was prepared (Fig. 1). Digestion of pMR40 individually with *Bam*HI and *Hin*dIII followed by self-ligation of the vector produced pMQ1 and pMQ2, which lacked approximately 0.5- and 0.7-kb insert fragments, respectively. The 0.5- and 0.7-kb DNA fragments were gel purified and separately cloned into pUC18 to obtain pMQ6 and pMQ12, respectively. Digestion of pMR41 with *Eco*RI followed by self-ligation of the vector produced pMQ3, which lacked approximately a 1.0-kb insert DNA fragment. Similarly, digestion of pMR40 with *Xho*I and *Hin*dIII followed by self-ligation of the vector produced pMQ10, which lacked a 3.0-kb insert fragment. In addition, several DNA fragments were gel purified after digestion of pMR40 with various enzymes and subcloned into pUC18 in order to generate the respective plasmid derivatives (Table 1). DNA sequencing was performed in a Pharmacia ALF DNA sequencer, using autoread and autocycle sequencing kits. Fluorescence-labeled universal M13 forward and reverse sequencing primers were used in the sequencing reactions. In some cases, fluorescence-labeled primers based on the previously determined sequence were synthesized and used for determining the sequence (Fig. 1). Typically, a sequence of 300 to 400 bp per run with less than 2% ambiguities was obtained. Each clone was sequenced three to four times in order to confirm the nucleotide sequence and to resolve any ambiguities. Finally, in some cases, ambiguities were resolved

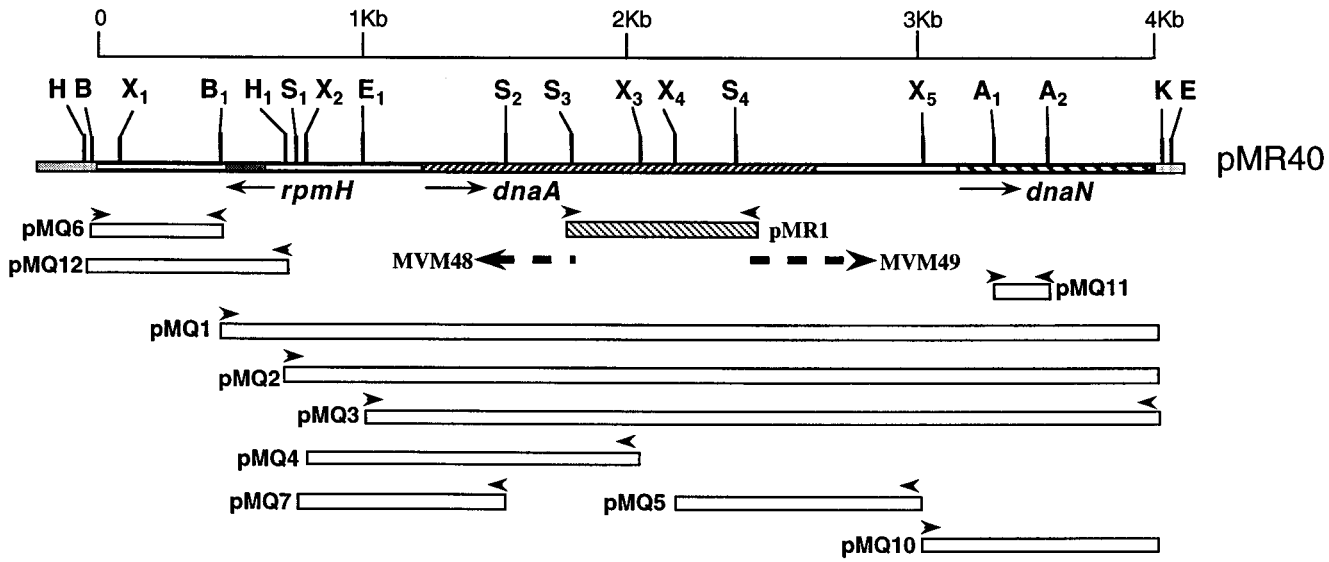

FIG. 1. Sequencing strategy. A 4-kb *Pvu*II DNA fragment containing the *dnaA* gene was isolated from cosmid pMR25 and cloned into the *Sma*I site of a pUC18 vector to produce pMR40. The positions of various restriction enzyme cutting sites and putative genes in the 4-kb insert are shown. The positions of the *rpmH*, *dnaA*, and *dnaN* genes in the 4-kb fragment are based on the nucleotide sequence (see Fig. 2). An arrow near the coding region of a gene indicates the direction of transcription. Different-size DNA fragments present in various pUC18-derived plasmids (pMQ1 to pMQ7 and pMQ10 to pMQ12) used for sequencing are shown as empty boxes. Arrowheads above the empty boxes indicate the direction of sequencing. Broken lines indicate the sequence obtained by using the fluorescence-labeled primers MVM48 and MVM49 synthesized in a Pharmacia Gene Assembler. Fluorescence-labeled M13 universal forward and reverse sequencing primers were used for determination of the DNA sequences of various pUC18-derived plasmids, using a Pharmacia ALF DNA sequencer. Abbreviations for restriction enzymes: A, *Ava*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sma*I; X, *Xho*I. For the purpose of clarity, restriction endonuclease sites of each category in the 4-kb fragment are numbered from left to right. ▨, multiple cloning site of the vector (not drawn to scale); ■, *rpmH* gene coding region; ▨, *dnaA* gene coding region; ▨, *dnaN* gene coding regions; ▨, the 777-bp PCR product cloned in pUC18 (pMR1).

by sequencing both DNA strands. The nucleotide sequences of the overlapping clones were aligned by using the Gene Jockey and DNA Strider software programs. The Clustal W 1.5 program was used for protein sequence alignments (11, 12).

Determination of autonomous replication activity. Nucleotide sequence analysis of the 4-kb *dnaA*-containing fragment revealed a single *Bam*HI recognition site at nucleotide position 453. Digestion of pMR40 with *Bam*HI produced a 456-bp fragment and a 6,240-bp fragment. The 6,240-bp fragment was gel purified, ligated with a 1.3-kb *Bam*HI fragment containing the aminoglycoside phosphotransferase (*aph*) gene (26), and used to transform *E. coli* DH5α to ampicillin and kanamycin resistance. Transformant colonies were purified, and their plasmid DNAs were analyzed by restriction digestion followed by agarose gel electrophoresis. One recombinant plasmid, pMR43, was used to transform *M. smegmatis* mc^2155 to kanamycin resistance. Again, transformant colonies were purified, and plasmid DNA was recovered by the electroduction protocol (3). In this method, recombinant plasmid-containing *M. smegmatis* cells were mixed with *E. coli* DH5α, washed, resuspended in 10% glycerol, and electroporated. Following electroporation, Luria broth was added, and the bacterial suspensions were incubated for 1 h at 37°C prior to plating on LB agar medium containing kanamycin (50 μg/ml) and incubated at 37°C. Since the generation time of *E. coli* is approximately 20 to 30 min compared with 3 to 4 h for *M. smegmatis*, and the colony morphology of *E. coli* is distinctly different from that of *M. smegmatis*, the colonies of *E. coli* appearing after overnight incubation were easily identified and selected for purification. Plasmid DNA was recovered from *E. coli* by using the alkali lysis protocol (26). The presence of insert in the plasmid was verified by digesting the plasmid DNA with *Pvu*II and probing with a radiolabeled 4-kb *Pvu*II fragment isolated from pMR25 (26). pYUB12 and pYUB18, the *E. coli-Mycobacterium* shuttle plasmid and cosmid, were used as controls in these experiments (14).

Nucleotide sequence accession number. The DNA sequence reported was deposited in GenBank and assigned accession number U17833.

RESULTS AND DISCUSSION

Cloning of the *dnaA* gene. Our initial attempts to complement *E. coli dnaA* mutants with *M. smegmatis* genomic plasmid expression libraries have not been successful (23). This is not surprising since only the *dnaA* genes from closely related enteric bacteria, not those from distantly related gram-positive bacteria, have been shown to complement *E. coli dnaA* mutants (29, 30, 39). Hence, to clone the *dnaA*-like gene from *M. smegmatis*, a reverse genetic approach as described in Materials and Methods was used. PCR amplification using degenerate oligonucleotide primers to conserved regions of eubacterial DnaA protein resulted in the production of a 777-bp fragment (data not shown). The nucleotide sequence of the PCR product, when searched by using BLAST (basic local alignment search tool) (1), showed significant homologies with *dnaA* gene sequences from other bacteria (data not shown). The PCR product thus provided the basis for obtaining cosmid DNA molecules containing the *dnaA* gene from the genomic DNA cosmid library as described above. From one such cosmid, a 4-kb fragment containing the *dnaA* gene was isolated and subcloned into pUC18.

DNA sequence analyses. A BLAST search with the 4-kb nucleotide sequence revealed homologies with *rpmH*, *dnaA*, and *dnaN* genes of several bacteria, including *Streptomyces coelicolor*, *Micrococcus luteus*, *B. subtilis*, and *E. coli* (data not shown; see below). The nucleotide sequence of the 4-kb region was computer analyzed by using the DNA Strider 1.2 program (Fig. 2). Three open reading frames (ORFs) beginning with GTG, referred to here as *orf1, orf2*, and *orf3*, were identified. Since many mycobacterial genes that initiate with GTG are known, this feature is not surprising (7). The locations of these ORFs correspond to nucleotide positions 621 to 478 on the bottom strand (*orf1*), nucleotide positions 1238 to 2725 (*orf2*) and 3225 to the end of the cloned fragment (*orf3*) on the top strand. *orf1*, *orf2*, and *orf3* showed homologies with *rpmH*, *dnaA*, and *dnaN* genes, respectively (data not shown; see below). No termination codon in *orf3* was detected. *orf1* is smaller than *orf2* and *orf3*. The direction of transcription of *orf1* appears to be opposite that of *orf2* and *orf3.*

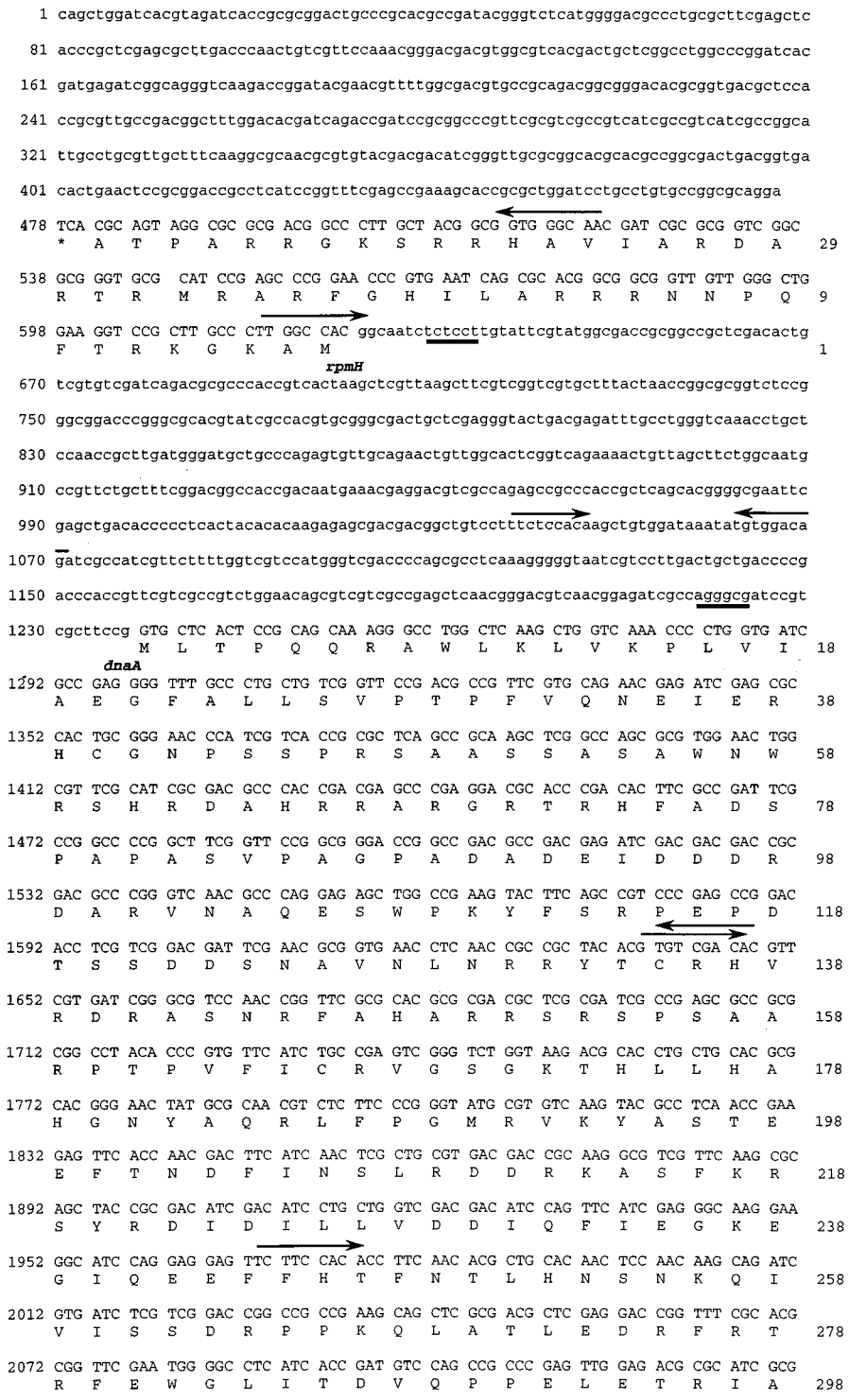

FIG. 2. Sequence of the 4-kb DNA fragment of *M. smegmatis* containing the *rpmH*, *dnaA*, and *dnaN* genes. The nucleotide sequence was determined by using autoread and autocycle kits and a Pharmacia ALF DNA sequencer. The nucleotide sequence of the top strand is shown, and the nucleotides are numbered on the left. The nucleotide sequences of the ORFs are presented in capital letters. Amino acids are indicated in single-letter codes below codons and are numbered on the right. The coding strand sequence of *orf1* (*rpmH*), as indicated in the text, corresponds to the bottom strand and hence is not shown. Asterisks indicate the stop codon. Positions and orientations of putative DnaA box sequences are marked by arrows above the sequence. Putative ribosome binding sequence are underlined.

(i) *orf1*. A small ORF initiating with GTG at position 621 of the bottom strand is predicted to encode the RpmH homolog. Seven bases upstream of the potential GTG start codon is the sequence AGGAGA, which has some resemblance to the *E. coli* consensus ribosome binding sequence. The presumed protein is 47 residues in length and has a molecular mass of 5.5 kDa, which is in agreement with the molecular masses of the RpmH proteins in *B. subtilis* (5.2 kDa), *M. luteus* (5.4 kDa), *E.*

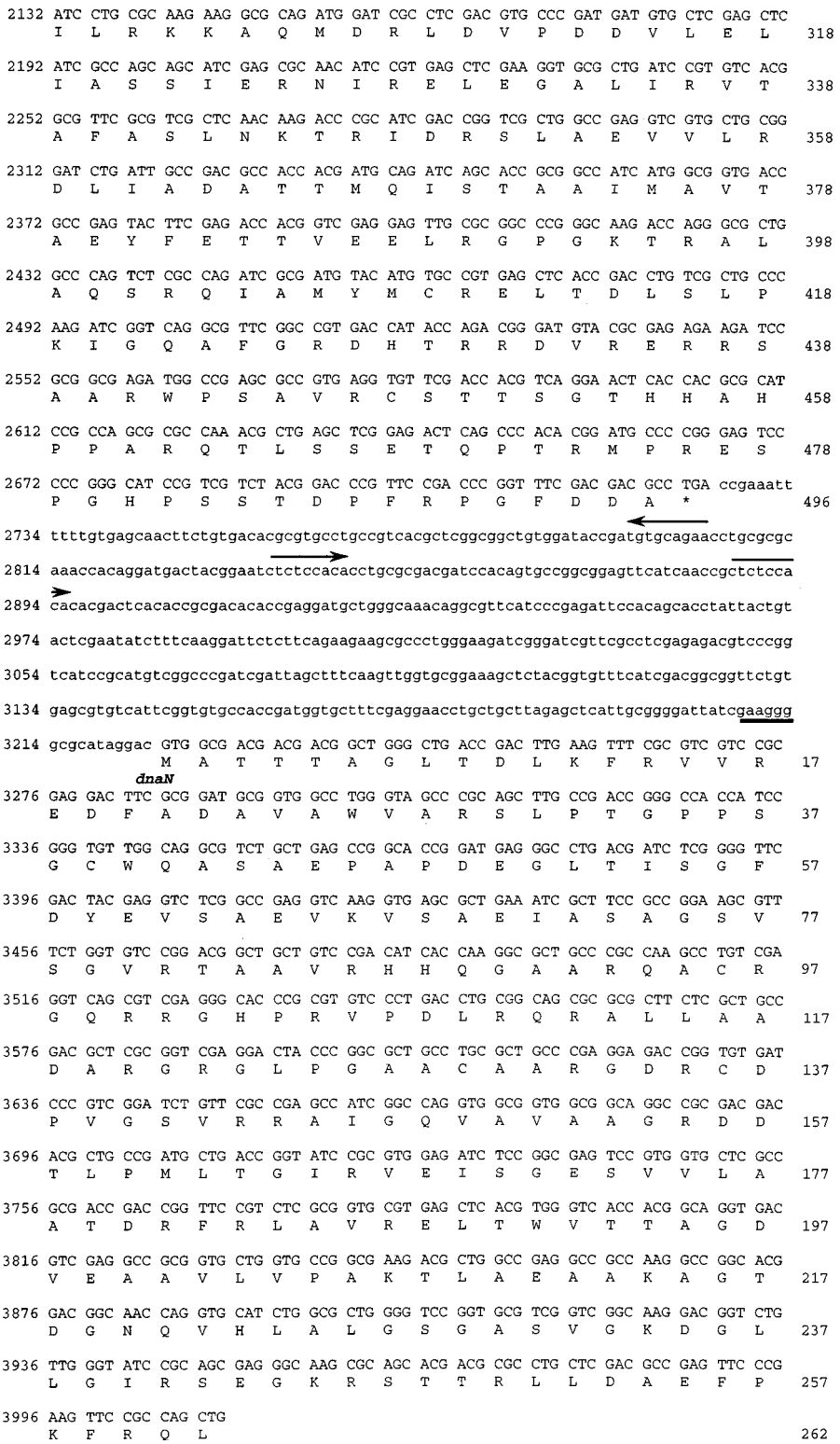

FIG. 2—*Continued.*

coli (5.4 kDa), and *S. coelicolor* (5.5 kDa). The predicted protein fits very well into the typical homology pattern of the known RpmH proteins (Fig. 3A) and is 72, 69, 55, and 53% similar to the proteins of *S. coelicolor*, *M. luteus*, *B. subtilis*, and *E. coli*, respectively (data not shown). As in other bacteria, the putative RpmH protein of *M. smegmatis* contains the conserved KRTFQP motif in the amino-terminal region (Fig. 3A).

(ii) *orf2*. A large ORF initiating with one of the three po-

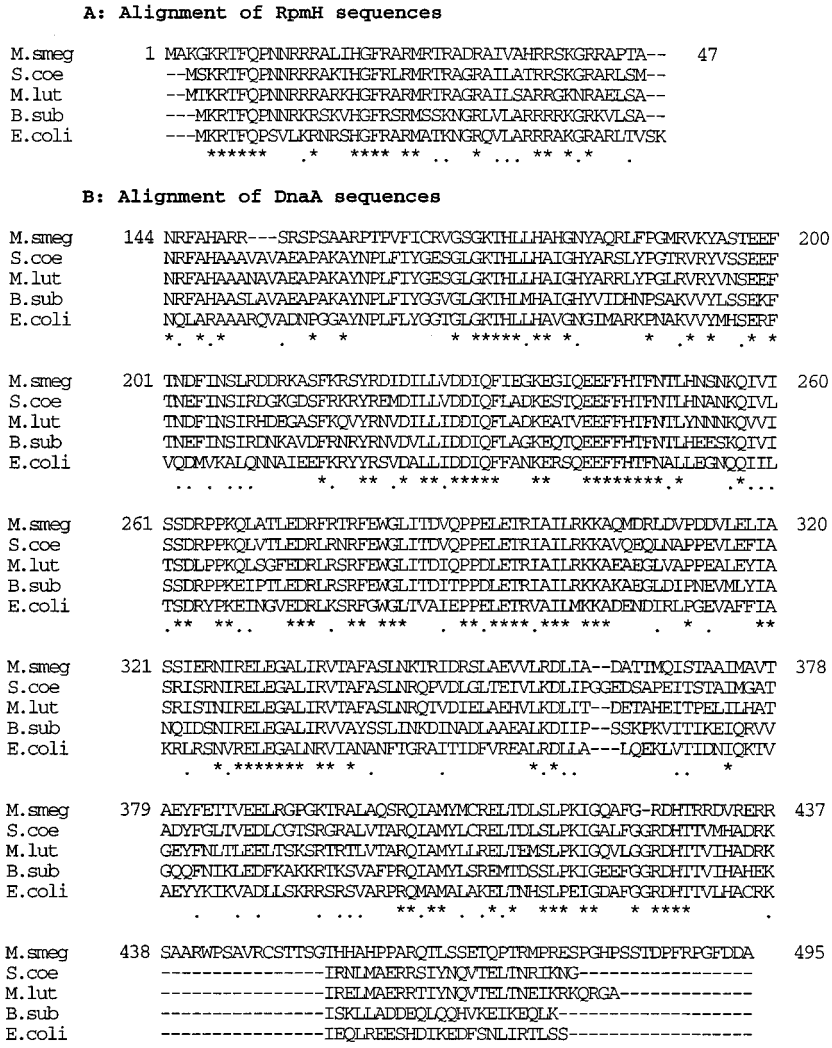

FIG. 3. Alignment of the amino acid sequences of *M. smegmatis* (M.smeg) RpmH (A) and DnaA (B) proteins with sequences of the respective proteins from *S. coelicolor (S.coe), M. luteus* (M.lut), *B. subtilis* (B.sub), and *E. coli*. Alignments were carried out by using the Clustal W 1.5 program. Gaps have been created for determining the homologies. Identical amino acids are marked with asterisks, and conserved substitutions are marked with dots. *M. smegmatis* RpmH and DnaA amino acid residues are numbered. Alignments started with amino acid 144 of the *M. smegmatis* DnaA protein (see Fig. 2).

tential GTG start codons at positions 1238, 1286, and 1331 and terminating at position 2725 was identified. Another potential but smaller ORF initiating with ATG at position 1061 was excluded because a stop codon, TGA, was located 81 nucleotides downstream from it. Of the three GTG start codons, the GTG at 1238 was tentatively chosen because the ORF extending from this position was the largest. Also, 14 bases upstream of the GTG at nucleotide position 1238 is a potential ribosome binding sequence, AGGGCG. Initiation from nucleotide positions 1286 and 1331 has not been excluded, however. The predicted protein from the largest ORF would have 495 amino acids and a molecular mass of 55.9 kDa, which is similar to the size of *M. luteus* DnaA protein (56.9 kDa) but is larger than those of the *E. coli* (52.6 kDa) and *B. subtilis* (50.8 kDa) DnaA proteins (28). The putative DnaA protein of *S. coelicolor* is 656 amino acid residues in length and is larger (73.1 kDa) than all known DnaA proteins (5). The deduced amino acid sequence of the *dnaA* gene of *M. smegmatis*, particularly the 352 amino acids in the carboxyl-terminal region, show substantial similarity to the corresponding regions of other bacteria (Fig. 3B). As is the case in other bacteria, this region contains several characteristic motifs such as GxxGxGKT, RELEGA, and GRDHT (28). A phylogenetic tree based on the sequence analyses carried out with the Clustal W 1.5 program (11, 12) revealed that the aligned region (Fig. 3B) of *M. smegmatis* is 62, 57, 50, and 41% similar to those of *S. coelicolor*, *M. luteus*, *B. subtilis* and *E. coli*, respectively (data not shown). The 143-residue amino-terminal region of *M. smegmatis* DnaA protein compared with that of *S. coelicolor* exhibited 44 amino acids that were identical and 22 that were similar (data not shown).

(iii) *orf3*. The GTG at nucleotide position 3225 was chosen as a potential start codon because the ORF extending from it is the largest. Eleven bases upstream of the putative start codon is GAAGGG, a sequence resembling that of the *E. coli* consensus ribosome binding sequence. The deduced amino acid sequence of *orf3* showed significant similarities with the amino-terminal parts of the DnaN sequences of other bacteria (data not shown).

The *dnaA* gene flanking regions. The intergenic region of *rpmH-dnaA* (i.e., the 5′ flanking region of *dnaA*) and/or the

TABLE 2. Putative DnaA box sequences in the *dnaA* gene region of *M. smegmatis*

Putative DnaA box	Sequence[a]	Location[b]
Coding strand	TTCTCCACA (1)	1038–1046
	TTGGCCACG (2)	614–622
	GTGTCGACA (2)	1639–1647
	TCTTCCACA (2)	1968–1976
	CTCTCCACA (2)	2837–2845
	CTCTCCACA (2)	2887–2895
Complementary strand[c]	CTGTCCACA (1)	1070–1062
	TTCTGCACA (2)	2804–2796
	GTGTCGACA (2)	1648–1640
	TTGCCCACC (2)	521–513
Consensus	TTG/CTCCACA	

[a] Using Gene Jockey sequence analysis software, DnaA box-like sequences with one to two mismatches present within the 4-kb region (Fig. 2) were identified. The numbers in parentheses indicate the number of mismatches with the consensus sequence (TTGTCCACA) for the DnaA boxes of G+C-rich bacteria.
[b] Nucleotide position of each DnaA box in the 4-kb insert of pMR40.
[c] Defined on the basis of the *dnaA* gene sequence.

intergenic region of *dnaA-dnaN* (i.e., the 3′ flanking region of *dnaA*) in many bacteria contain the DnaA protein recognition sequences called DnaA boxes. The A+T content of this region is usually higher than that of the rest of the chromosome. We found that the A+T contents of the 5′ and 3′ flanking regions of the *dnaA* gene of *M. smegmatis* were 39 and 44%, respectively. The A+T and G+C contents of the 4-kb *dnaA* gene region were 35 and 65%, respectively, as is typical of the mycobacterial genome (6). The consensus sequence of DnaA boxes in bacteria with high G+C contents has been reported to be TTGTCCACA (5, 8, 28). No such sequences were found within the 4-kb *dnaA* gene region of *M. smegmatis*. However, several putative DnaA box-like sequences with one to two mismatches with the consensus sequence were found (Table 2; Fig. 2). Two of these putative DnaA boxes were present in the intergenic region of *rpmH-dnaA*, and three were present in the intergenic region of *dnaA-dnaN*. Three putative DnaA boxes were found within the coding region of the *dnaA* gene, and two were found within the coding region of the *rpmH* gene. Of the 10 putative DnaA boxes, 4 were oriented opposite the other 6 (Fig. 2). On the basis of these data, we suggest TT(G/C)TCC ACA as the consensus sequence for the DnaA boxes of *M. smegmatis*. Precise biochemical experiments with DNA fragments containing putative DnaA boxes and purified DnaA protein will be necessary to define the exact sequences of the DnaA boxes of *M. smegmatis.*

Identification of autonomously replicating sequences. With the exception of *E. coli*, the chromosomal origin of replication in all eubacteria that have been examined is localized near the *dnaA* gene (15, 22, 28, 31). As a first step in identifying the putative chromosomal replication origin of *M. smegmatis*, we examined the ability of the *M. smegmatis* DNA fragment, containing the *dnaA* gene and its flanking regions, to support autonomous replication when present in plasmids normally nonreplicative in mycobacteria. In these experiments, *M. smegmatis* was transformed separately with a pUC18 plasmid containing the 1.3-kb *aph* gene (pMR42) and a pUC18 plasmid containing the 1.3-kb *aph* gene and the 3.5-kb *dnaA* gene region of *M. smegmatis* (pMR43). As a control, *M. smegmatis* was transformed with an *E. coli-Mycobacterium* shuttle vector (pYUB12). Kanamycin-resistant colonies were obtained with pMR43 (Fig. 4C) and pYUB12 (Fig. 4D) but not with pMR42 (Fig. 4B). Untransformed *M. smegmatis* did not produce any kanamycin-resistant colonies (Fig. 4A). Approximately 10^6 transformants per μg of electroporated DNA were obtained with pMR43, and the transformation frequency was comparable with that obtained with the pYUB12 control vector (0.6×10^6 transformants per μg of DNA). The kanamycin-resistant colonies of *M. smegmatis* obtained with the pMR43 vector could result either from stable extrachromosomal replication of plasmids or stable integration of the introduced plasmid onto the chromosome. To distinguish between these two possibilities, plasmid DNA from kanamycin-resistant *M. smegmatis* (Fig. 4C) was recovered by electroduction into *E. coli* (see

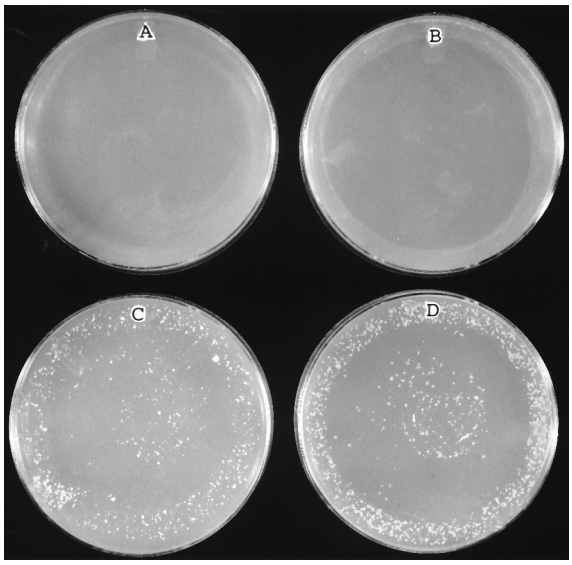

FIG. 4. Autonomous replication activities of plasmids containing *M. smegmatis* sequences. *M. smegmatis* mc^{2}155 cells were electrotransformed with the plasmids indicated below, plated on 7H10 agar plates containing kanamycin, incubated at 37°C for 4 days, and photographed. (A) No plasmids; (B) pMR42, a pUC18 plasmid containing the *aph* gene; (C) pMR43, a pUC18 plasmid containing the 3.5-kb genomic DNA fragment of *M. smegmatis* and the 1.3-kb *aph* gene; (D) pYUB12, an *E. coli-Mycobacterium* shuttle vector.

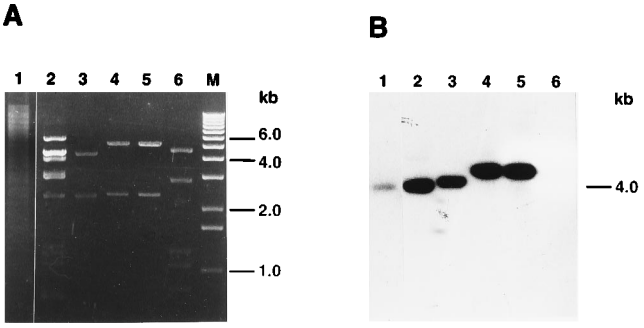

FIG. 5. Analysis of plasmids containing autonomously replicating sequences. (A) Agarose gel electrophoresis of genomic DNA and plasmids containing cloned sequences of *M. smegmatis*. Genomic and plasmid DNAs were digested with restriction enzymes as indicated. Following digestion with restriction enzymes, samples were electrophoresed on 0.8% agarose gels, stained with ethidium bromide, and photographed. Lanes: 1, *M. smegmatis* genomic DNA digested with *Pvu*II; 2, pMR25 cosmid DNA digested with *Pvu*II; 3, pMR40 DNA digested with *Pvu*II; 4, pMR43 DNA digested with *Pvu*II; 5, pMR43 plasmid DNA recovered from kanamycin-resistant *M. smegmatis* and digested with *Pvu*II; 6, pYUB12 DNA digested with *Hin*cII: M, markers. (B) Autoradiogram of agarose gel shown in panel A. The DNA in the agarose gel from panel A was transferred to nylon membranes, probed with a radiolabeled 4-kb *Pvu*II fragment from pMR25 in a Southern hybridization experiment, and visualized by autoradiography. Lanes 1 to 6 are the same as in panel A.

Materials and Methods). Our ability to recover the introduced pMR43 plasmid DNA from kanamycin-resistant *M. smegmatis* suggests that the introduced plasmid had not integrated into the chromosome but rather replicated extrachromosomally. Since pUC18 plasmids do not replicate on their own in mycobacteria, these data suggest that the *M. smegmatis dnaA* gene region rendered pUC18 plasmids capable of autonomous replication in *M. smegmatis.*

To further establish that the recovered plasmid is similar to that of the input plasmid, restriction digestion followed by agarose gel electrophoresis was carried out (Fig. 5A). Following electrophoresis, DNA was transferred to nylon membranes and probed with the ^{32}P-labeled 4-kb *Pvu*II fragment containing the *dnaA* gene (26). DNA bands were visualized by autoradiography (Fig. 5B). The *E. coli-Mycobacterium* shuttle vectors pYUB12 and pYUB18 served as controls (14). A single hybridization signal was observed with both the introduced plasmid, pMR43 (Fig. 5B, lane 4), and the recovered plasmid, pMR43 (Fig. 5B, lane 5). Since the size of the insert from the recovered plasmid DNA (Fig. 5, lanes 5) was similar to the size of that from the introduced plasmid DNA (Fig. 5, lanes 4), it was concluded that no change or modification of pMR43 occurred during transformation and growth in *M. smegmatis*. Whether any changes occurred in regions not covered by the restriction enzymes tested remains to be determined by sequencing. No hybridization signal was detected with pYUB12 (Fig. 5B, lane 6) or pYUB18 (data not shown). These data indicated that the DNA sequences present in pMR43 did not have any homology with the replication origin sequences present in the pYUB12 and pYUB18 vectors (14). The DNA fragment containing the autonomously replicating sequence was derived from the *M. smegmatis* chromosome, since digestion of the genomic DNAs of *M. smegmatis* (Fig. 5A, lane 1) and a *dnaA*-positive cosmid from the *M. smegmatis* genomic DNA library (Fig. 5A, lane 2) with *Pvu*II showed a 4-kb band upon hybridization (Fig. 5B, lanes 1 and 2). Similarly, digestion of pMR40, the pUC18 plasmid containing the 4-kb insert from the pMR25 cosmid with *Pvu*II, also showed one band upon hybridization (Fig. 5B, lane 3). The position of the hybridization signal in lane 3 corresponding to the *Pvu*II fragment from pMR40 was higher than the hybridization signal produced by the *Pvu*II digestion of the genomic DNA and pMR25 cosmid DNA (Fig. 5B, lanes 1 and 2). This is because cloning of the 4-kb *Pvu*II fragment into the *Sma*I site of the pUC18 vector eliminated the *Pvu*II site of the insert. Thus, the *Pvu*II fragment released from the pMR40 vector included a 323-bp region in addition to the 4-kb insert.

Although the observation that the DNA fragment containing the *M. smegmatis dnaA* gene region functions as an autonomously replicating sequence is indicative of the presence of replication origins, it remains to be proven whether it is the actual site for replication initiation in *M. smegmatis*. Since putative DnaA box-like sequences were found in the intergenic regions of *rpmH-dnaA* and *dnaA-dnaN*, it will be important to determine whether the 5′ or 3′ flanking region or both flanking regions of the *dnaA* gene are required for autonomous replication activity. Detailed experiments to delimit the size of the replication origin should allow the construction of minichromosomes of *M. smegmatis* in order to begin to understand the molecular mechanisms involved in replication initiation. Identification of such sequences from slow growers such as *M. tuberculosis* should allow one to systematically carry out genetic and biochemical experiments for understanding the growth rate differences between slowly and rapidly growing mycobacterial species.

ACKNOWLEDGMENTS

We acknowledge with gratitude receiving the *E. coli-Mycobacterium* shuttle vectors and *M. smegmatis* genomic DNA libraries from William R. Jacobs, Jr., and the valuable help of A. John Clark with Clustal sequence alignments. We also thank S. Black and C. Wu for help with computer analysis of the sequence data, and we thank Vincent Steingrube and Nora Linderoth for critical comments on the manuscript.

This research was supported by New Faculty Startup funds provided by The University of Texas Health Center at Tyler to M.V.V.S.M.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–415.
2. **Bates, J. H.** 1995. Tuberculosis chemotherapy: the need for new antituberculosis drugs is urgent. Am. J. Respir. Crit. Care Med. **151:**942–943.
3. **Baulard, A., C. Jourdan, A. Mercenier, and C. Locht.** 1992. Rapid mycobacterial plasmid analysis by electroduction between *Mycobacterium* sp. and *Escherichia coli*. Nucleic Acids Res. **20:**4105.
4. **Bramhill, D., and A. Kornberg.** 1988. A model for initiation at origins of DNA replication. Cell **54:**915–918.
5. **Calcutt, M. J., and F. J. Schmidt.** 1992. Conserved gene arrangement in the origin region of the *Streptomyces coelicolor* chromosome. J. Bacteriol. **174:** 3220–3226.
6. **Clark-Curtiss, J. E.** 1990. Genome structure of mycobacteria, p. 77–96. *In* J. McFadden (ed.), Molecular biology of the mycobacteria. Academic Press, London.
7. **Dale, J. W., and A. Patki.** 1990. Mycobacterial gene expression and regulation, p. 173–198. *In* J. McFadden (ed.), Molecular biology of the mycobacteria. Academic Press, London.
8. **Fujita, M. Q., Yoshikawa, H., and Ogasawara, N.** 1990. Structure of the *dnaA* region of *Micrococcus luteus*: conservation and variations among eubacteria. Gene **93:**73–78.
9. **Fukuoka, T., S. Moriya, H. Yoshikawa, and N. Ogasawara.** 1990. Purification and characterization of an initiation protein for chromosomal replication, DnaA, in *Bacillus subtilis*. J. Biochem. **107:**732–739.
10. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–561.
11. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. Clustal V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:**189–191.
12. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:**151–153.
13. **Husson, R. H., B. J. James, and R. A. Young.** 1990. Gene replacement and expression of foreign DNA in mycobacteria. J. Bacteriol. **172:**519–524.
14. **Jacobs, W. R., Jr., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. Udani, W. Jones, R. G. Barletta, and B. R. Bloom.** 1991. Genetic systems for mycobacteria. Methods Enzymol. **204:**537–555.
15. **Kornberg, A., and T. Baker.** 1991. DNA replication. W. H. Freeman and Company, New York.
16. **Margolin, W., D. Bramhill, and S. R. Long.** 1995. The *dnaA* gene of *Rhizobium meliloti* lies within an unusual gene arrangement. J. Bacteriol. **177:** 2892–2900.
17. **Marians, K. J.** 1992. Prokaryotic DNA replication. Annu. Rev. Biochem. **61:**673–719.
18. **Mildvan, A. S., and D. C. Fry.** 1987. NMR studies of the mechanism of enzyme action. Adv. Enzymol. **59:**241–243.
19. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
20. **Moriya, S., T. Atlung, F. G. Hansen, H. Yoshikawa, and N. Ogasawara.** 1992. Cloning of autonomously replicating sequences [ars] from the *Bacillus subtilis* chromosome. Mol. Microbiol. **6:**309–315.
21. **Moriya, S., W. Firshein, H. Yoshikawa, and N. Ogasawara.** 1994. Replication of a *Bacillus subtilis* oriC plasmid in vitro. Mol. Microbiol. **12:**469–478.
22. **Ogasawara, N., and H. Yoshikawa.** 1992. Genes and their organization in the replication origin region of the bacterial chromosome. Mol. Microbiol. **6:** 629–634.
23. **Rajagopalan, M., M. H. Qin, V. A. Steingrube, D. R. Nash, R. J. Wallace, Jr., and M. V. V. S. Madiraju.** 1995. Amplification and cloning of *Mycobacterium tuberculosis dnaA* gene. Gene **163:**75–79.
24. **Reichman, L. B., and E. S. Hershfield.** 1994. Tuberculosis: a comprehensive international approach. Marcel Dekker, Inc., New York.
25. **Richter, S., and W. Messer.** 1995. Genetic structure of the *dnaA* region of the cyanobacterium *Synechocystis* sp. strain PCC6803. J. Bacteriol. **177:**4245–4251.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Serrano, R.** 1988. Structure and function of proton translocating ATPase in plasma membranes of plants and fungi. Biochim. Biophys. Acta **947:**1–28.
28. **Skarstad, K., and E. Boye.** 1994. The initiator protein DnaA: evolution, properties and function. Biochim. Biophys. Acta **1217:**111–130.

29. **Skovgaard, O.** 1990. Nucleotide sequence of a *Proteus mirabilis* DNA fragment homologous to the *60K-rnpA-rpmH-dnaA-dnaN-recF-gyrB* region of *Escherichia coli*. Gene **93:**27–34.
30. **Skovgaard, O., and F. G. Hansen.** 1987. Comparison of *dnaA* nucleotide sequences of *Escherichia coli*, *Salmonella typhimurium*, and *Serratia marcescens*. J. Bacteriol. **169:**3976–3981.
31. **Smith, D. W., T. W. Yee, C. Baird, and V. Krishnapillai.** 1991. Pseudomonad replication origins: a paradigm for bacterial origins? Mol. Microbiol. **5:**2581–2587.
32. **Walker, J. E., M. Saraste, M. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α and β subunits of ATP synthase, myokinases and other ATP requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
33. **Wheeler, P. R., and C. Ratledge.** 1994. Metabolism of *Mycobacterium tuberculosis*, p. 353–385. *In* B. R. Bloom (ed.), Tuberculosis: pathogenesis, protection, and control. American Society for Microbiology, Washington, D.C.
34. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp8 and pUC19 vectors. Gene **33:**103–119.
35. **Yee, T. W., and D. W. Smith.** 1990. Pseudomonas chromosomal replication origins: a bacterial class distinct from *Escherichia coli*-type origins. Proc. Natl. Acad. Sci. USA **87:**1278–1282.
36. **Young, D. B.** 1994. Strategies for new drug development, p. 559–568. *In* B. R. Bloom (ed.), Tuberculosis: pathogenesis, protection, and control. American Society for Microbiology, Washington, D.C.
37. **Zakrzewska-Czerwinksa, J., and H. Schrempf.** 1992. Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. J. Bacteriol. **174:**2688–2693.
38. **Zweiger, G., and L. Shapiro.** 1994. Expression of *Caulobacter dnaA* as a function of the cell cycle. J. Bacteriol. **176:**401–408.
39. **Zyskind, J. W., J. M. Cleary, W. S. A. Brusilow, N. Harding, and D. W. Smith.** 1984. Chromosomal replication origin from the marine bacterium *Vibrio harveyi* functions in *Escherichia coli: oriC* consensus sequence. Proc. Natl. Acad. Sci. USA **80:**1164–1168.