Journal of General Microbiology (1992), **138**, 1717–1727. *Printed in Great Britain*

The nucleotide sequence of the promoter, 16S rRNA and spacer region of the ribosomal RNA operon of *Mycobacterium tuberculosis* and comparison with *Mycobacterium leprae* precursor rRNA

KAREN E. KEMPSELL,[1]† YUAN-EN JI,[1] IRIS C. E. ESTRADA-G,[2]‡ M. JOSEPH COLSTON[2] and ROBERT A. COX[1]*

[1] *Laboratory of Developmental Biochemistry and* [2] *Leprosy and Mycobacterial Research, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK*

(*Received 22 August 1991; revised 21 October 1991; accepted 29 October 1991*)

***Mycobacterium tuberculosis* H37Rv has a single *rrn* (ribosomal RNA) operon. The operon was cloned and a region of 1536 nucleotides was sequenced, starting 621 bp upstream from the 5′-end of the 16S rRNA coding region and continuing to the start of the 23S rRNA coding region. The 16S rRNA sequence inferred from the gene sequence was found to differ in one position from *Mycobacterium bovis* (nucleotide 1443) and from *Mycobacterium microti* (nucleotide 427). A single putative promoter was identified on the basis of similarities with the sequence of *rrn* operons of *Bacillus subtilis* and *Escherichia coli*. The regions of similarity include a −35 box, a −10 box, a stringent response element, antitermination signals, potential RNAase III processing sites and features of precursor rRNA secondary structure. Sequences upstream from the 5′-end of *Mycobacterium leprae* 16S rRNA were also investigated. Homologous schemes of secondary structure were deduced for precursor rRNA of both *M. tuberculosis* and *M. leprae*; although the principal features are common to both species there are notable differences.**

Introduction

Mycobacteria are of interest because they include important pathogens such as *Mycobacterium leprae* and *Mycobacterium tuberculosis*, which have doubling times of 12 d and 1 d respectively (Shepard, 1960; Winder & Rooney, 1970). Several other mycobacteria (e.g. *Mycobacterium africanum*, *Mycobacterium bovis* and *Mycobacterium microti*) are very closely related to *M. tuberculosis* and are classified as members of the *M. tuberculosis* complex (Wayne, 1982). The members of this complex have a single rRNA operon and RFLP analysis has revealed that this operon and its 5′- and 3′-flanking sequences are highly conserved among members of the complex (K. E. Kempsell, I. C. E. Estrada-G, M. J. Colston & R. A. Cox, unpublished work).

This report describes further analysis of the rRNA gene family of *M. tuberculosis* H37Rv. The nucleotide sequence of the promoter region, the 16S rRNA coding region, the intercistronic region and the 5′-end of the 23S rRNA coding region was established. The promoter region of *M. tuberculosis*, which serves as a model for all members of the *M. tuberculosis* complex, was investigated because of our interest in the control of rRNA synthesis in slow-growing mycobacteria and its role in cell proliferation. The complete 16S rRNA sequence allows clarification of the relation of *M. tuberculosis* with other members of the *M. tuberculosis* complex and with other slow-growing mycobacteria.

In addition, we have investigated the nucleotide sequence of the putative leader sequence of the precursor rRNA (pre-rRNA) transcript of *M. leprae* (cf. Sela & Clark-Curtiss, 1991). The nucleotide sequence of the *M. leprae* rRNA operon was recently published (Liesack *et al.*, 1990, 1991; Sela & Clark-Curtiss, 1991). Comparisons of the data for *M. tuberculosis* and *M. leprae* reveal interesting similarities and differences. Our results suggest that the single *rrn* operon of both species

* Author for correspondence. Tel. (081) 959 3666; fax (081) 906 4477.

† Present address: Department of Surgery, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK.

‡ Present address: Departamento de Immunología, Escuela Nacional de Ciencias Biologicas, Mexico City, Mexico.

The nucleotide sequence data reported in this paper have been submitted to EMBL and have been assigned the accession numbers X58889, X58890 and X58891.

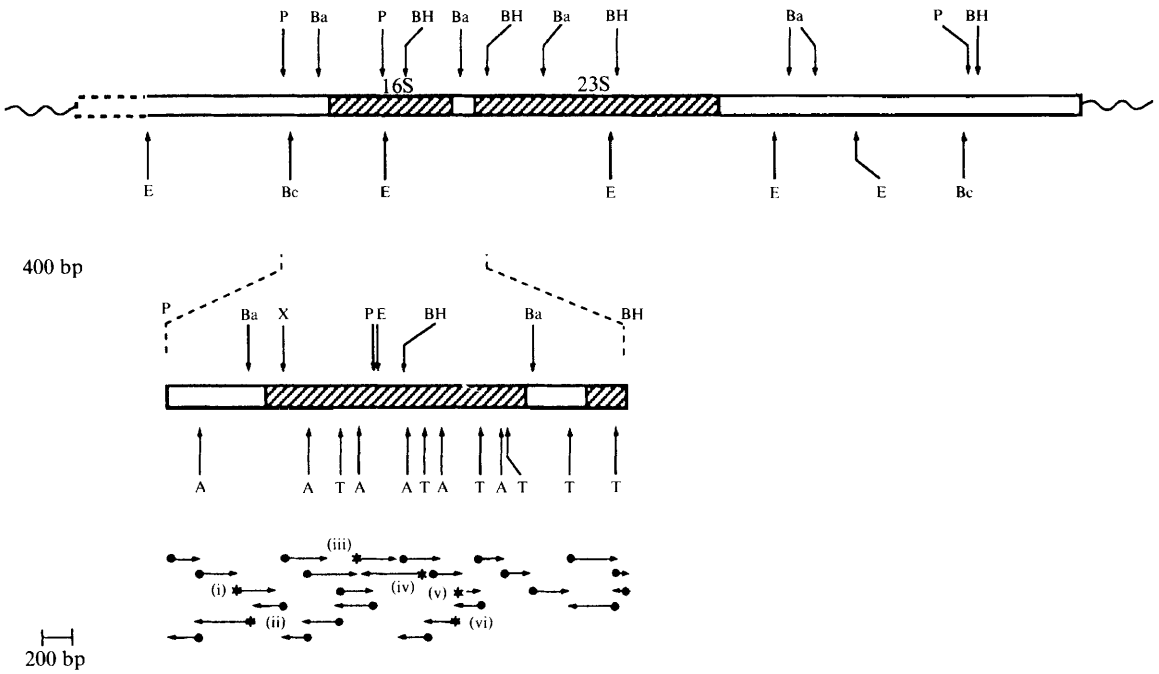


Fig. 1. Restriction site map, sub-cloning and sequencing strategy for the *M. tuberculosis* rRNA cistron. Hatched portions represent 16S and 23S rRNA coding regions and unhatched portions denote flanking sequences. Restriction endonucleases used in cloning were: A, *Alu*I; Ba, *Bal*I; Bc, *Bcl*I; BH, *Bam*HI; E, *Eco*RI; P, *Pst*I; T, *Taq*I; X, *Xho*I. No sites were detected for *Hin*dIII or *Bgl*II. The regions specified by the arrows with asterisks were sequenced using the oligonucleotide primer indicated using roman numerals (see Table 1). ~, bacteriophage λ EMBL3 sequences.

conforms to the general pattern established for a wide range of bacteria (for review see King *et al.*, 1986), in which a single transcript (pre-rRNA) is processed to yield mature 16S rRNA and 23S rRNA.

Methods

Materials. All chemicals and enzymes were obtained from suppliers described previously (Cox *et al.*, 1991). T7 sequencing kit was supplied by Pharmacia. DH5αF′ competent cells were supplied by Gibco/BRL. M13/pUC #1212-sequencing and #1233-reverse-sequencing primers were obtained from New England Biolabs.

Strain and culture. Strains of *Escherichia coli* K12 used in library propagation and DNA cloning were LE392 and DH5αF′. Bacterial cells were grown on LB medium (Maniatis *et al.*, 1982) at 37 °C. Bacteriophage λ EMBL3 and derivatives were propagated using strain LE392 grown in LB medium plus 10 mM-$MgSO_4$ at 37 °C for all manipulations. Competent cells of strain DH5αF′ were used for transformations with the recombinant plasmids pUC8 and pUC18.

Cloning of the 16S rRNA of M. tuberculosis H37Rv. A *Sau*3A partial library of *M. tuberculosis* H37Rv genomic DNA cloned into the *Bam*HI site of bacteriophage λ EMBL3 was provided by Dr E. Davies, Laboratory for Leprosy and Mycobacterial Research, NIMR, London. Bacteriophage plaques were transferred in duplicate onto Hybond-N nylon membranes according to the manufacturer's instructions. Duplicate filters were then individually hybridized with either probe A [^{32}P-labelled oligonucleotide g1 (see Table 1)] or with probe B [^{32}P-labelled oligonucleotide cg6 (see Table 1)], which recognize respectively sequences near to the 5′- and 3′-ends of a bacterial 16S rRNA (Cox *et al.*, 1991). Six recombinants hybridized with both probes. DNA was isolated from each of them using the 'maxi-preparative' method as described in Maniatis *et al.* (1982). All six clones were found to contain an insert which spanned an identical 11·6 kbp fragment of the rRNA operon characterized by the restriction map shown in Fig. 1. A representative recombinant, bacteriophage λ EMBL3-TB1, was selected for sequence analysis.

Cloned M. leprae cDNA. Clones of the *M. leprae* rRNA operon were isolated from a cosmid Lawrist 4 library (Cox *et al.*, 1991). The putative leader sequence of precursor rRNA and part of the coding sequence was found in a 901 bp *Sau*3A/*Pst*I fragment, which was sequenced by standard methods.

Sequencing methods. A 1·2 kbp *Pst*I fragment containing 5′-16S rRNA gene sequences and upstream sequences was cloned into the *Pst*I site of plasmid pUC8. A 1·1 kbp *Bam*HI fragment containing 3′-16S rRNA sequences, the spacer region and the 5′-end of 23S rRNA was cloned into plasmid pUC18. Appropriate *Alu*I and *Taq*I fragments of each of these recombinants were subcloned into pUC18 for sequence determination by the subcloning routine shown in Fig. 1. The 184 bp region located between the 3′-end of the *Pst*I and the 5′-end of the *Bam*HI fragment was synthesized by the PCR reaction using primers g3 and cg5 (see Table 1) and the product was isolated using Geneclean. The same product was obtained when either genomic DNA or phage λ EMBL3-TB1 DNA was used as substrate. All nucleotide sequences were determined by the dideoxy chain-termination method (Sanger *et al.*, 1977). PCR amplified fragments were sequenced in triplicate on both strands using a modified Sequenase protocol (Winship, 1989).

Double-strand plasmid template was made single-stranded for sequencing by the NaOH-precipitation method (Murphy & Ward, 1989). Single strand template was then sequenced in triplicate using

Table 1. *Deoxyribonucleotide primers*

(i)–(vi), These primers, identified by roman numerals, were used in sequencing.
A, B, These primers were radiolabelled at the 5′-end with ^{32}P and used to screen the *M. tuberculosis* H37Rv/EMBL3 library.

Primer	Sequence	Nucleotide position in *M. tuberculosis* 16S rRNA gene sequence*
tbp1 (i)	5′(GGTGAGTCTCGGTGCCGAGATCG)3′	312–334
tbp2 (ii)	5′(GCCAGTCTAATACAAATCCGGCT)3′	405–427
g3 (iii)	5′(GTGCCAGCAGCCGCGGTAATACG)3′	1126–1148
cg5† (iv)	5′(CGCTTGTGCGGGCCCCCGTCAATT)3′	1531–1554
v9 (v)	5′(ACTCGTGAGAGACTGCCGGGGTCA)3′	1758–1781
v11 (vi)	5′(GAGTTGACGTCGTCCCCGCCTTCCTC)3′	1787–1813
g1 (A)	5′(TTGGAGAGTTTGATCCTGGCTC)3′	627–648
cg6 (B)	5′(GGTACGGCTACCTTGTTACGACTT)3′	2105–2128

* See Fig. 2.
† c indicates that the nucleotide sequence is complementary to the RNA-like strand.

M13/pUC specific #1212-sequencing and #1233-sequencing primers and either Sequenase or T7 (Pharmacia) sequencing kits according to the manufacturer's instructions. When necessary, sequence-specific primers (tbp1–v11) were used (see Table 1 and Fig. 1).

Results

The recombinant phage λ EMBL3-TBI has the same map of restriction endonuclease sites (see Fig. 1) as that established previously for the rRNA operon by analysis of genomic DNA (K. E. Kempsell, I. C. E. Estrada-G, M. J. Colston & R. A. Cox, unpublished work). The cloning strategy shown in Fig. 1 was used to establish the nucleotide sequence 621 bp upstream from the 5′-end of the 16S rRNA cistron (1536 bp), through the spacer region (276 bp) to the 5′-end of the 23S rRNA cistron (the first 105 bp), as shown in Fig. 2.

M. tuberculosis 16S rRNA coding region

The 16S rRNA sequence, inferred from the gene sequence, comprises 1536 bp. Previously Rogall *et al.* (1990) used PCR amplification of a 5′-portion of the *M. tuberculosis* H37Rv 16S rRNA gene to obtain sequence data from the coding region, namely nucleotides 117–261 (nucleotides 738–882, Fig. 2), and nucleotides 429–498 (nucleotides 1040–1119, Fig. 2). Identical sequences were obtained for *M. bovis*, *M. bovis* BCG and *M. africanum* (Rogall *et al.*, 1990). The partial sequence data of Rogall *et al.* (1990) and the sequence presented in Fig. 2 are in agreement. Also, sequences near to the 3′-end (nucleotides 1295–1480) were obtained previously directly from 16S rRNA by the use of reverse transcriptase (Estrada-G *et al.*, 1989). Apart from several unidentified nucleotides the sequence determined using reverse transcriptase is very similar to the data shown in Fig. 2. The primary sequence can be folded into a secondary structure which is typical of eubacteria (see Fig. 3).

Comparison of the complete 16S rRNA gene sequence of *M. tuberculosis* with that of *M. bovis* (Suzuki *et al.*, 1988*a*) revealed one difference, the insertion/deletion of a C residue located in the hairpin loop region of helix 47 [nucleotides 2022–2104, Fig. 2; nucleotides 1401–1483 in the inferred 16S rRNA sequence (Fig. 3)]. The sequence for the hairpin loop is 5′CUCG3′ in *M. tuberculosis* (see also Estrada-G *et al.*, 1989) compared with 5′CUG3′ for *M. bovis*, as reported previously (Suzuki *et al.*, 1988*a*).

The sequence of nucleotides 46–470 of the coding region of 16S rRNA gene of *M. microti* was obtained earlier (EMBL Data Bank accession number X58889; K. E. Kempsell, I. C. E. Estrada-G, M. J. Colston & R. A. Cox, unpublished work). Compared with *M. tuberculosis*, the *M. microti* sequence is identical except for the deletion of a T residue located at position 427 which leads to the loss of an A-T basepair in helix 17 (Fig. 3).

The anti-Shine–Dalgarno sequence (Shine & Dalgarno, 1974) comprises a tract of ten pyrimidines located at the 3′-end of 16S rRNA (see Figs 2 and 3). At least nine mRNA species of members of the *M. tuberculosis* complex have been sequenced (for review see Dale & Patki, 1990). The putative Shine–Dalgarno sequence in each of the mRNAs is located six to twelve nucleotides upstream from the AUG or GUG start codon, and has the potential of forming five to eight basepairs with the 3′-end of 16S rRNA (see Table 2). *E. coli* has a shorter anti-Shine–Dalgarno sequence (seven nucleotides), and the Shine–Dalgarno sequence is closer (4–7 nucleotides) to the AUG start codon. The functional significance of these differences is not clear.

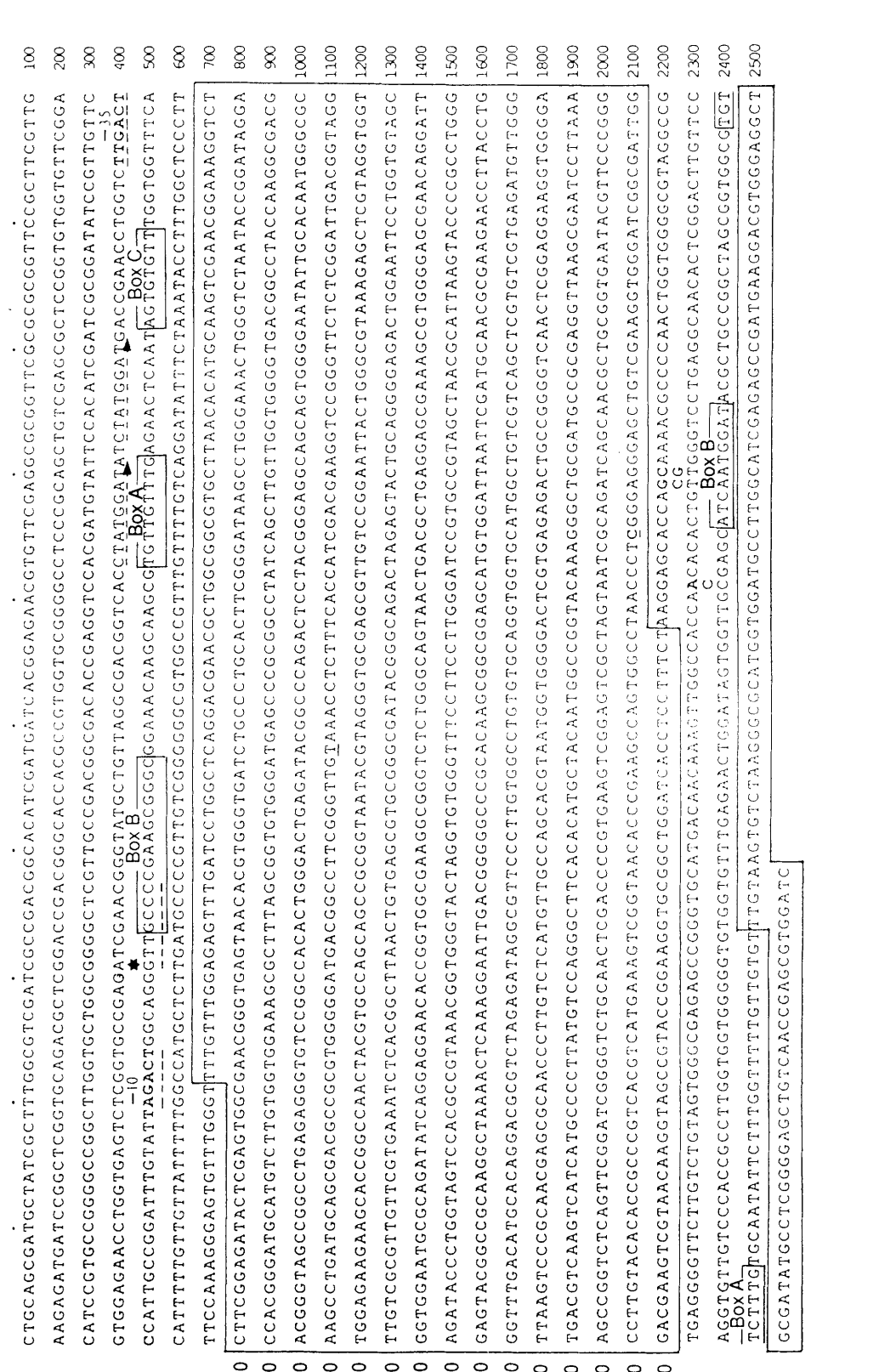


Fig. 2. Nucleotide sequence of the entire 16S rRNA gene of *M. tuberculosis* H37Rv, upstream and downstream flanking sequences, and the 5′-terminus of the 23S rRNA gene (EMBL Data Library accession number X58890). The coding sequence of the 16S rRNA gene and the 5′-terminus of the 23S rRNA gene are boxed. Nucleotides which differ from the published sequence for *M. bovis* (Suzuki *et al.*, 1988*a*) are highlighted in bold with the alternative nucleotides given below. The underlined T residue (position 427) is an insertion compared with sequence for *M. microti*. The underlined C residue at nucleotide position 2064 is an insertion compared with the sequence for *M. bovis* (Estrada-G *et al.*, 1989). −10 and −35 show the location of consensus-like −10 and −35 promoter elements. Possible antitermination elements (Box A, Box B and Box C) are marked appropriately. ★ marks the putative site for the start of transcription. The 'stringent' control element is underlined. An A-T rich 8 bp repeat sequence (underlined, broken arrow) may also be a promoter element.

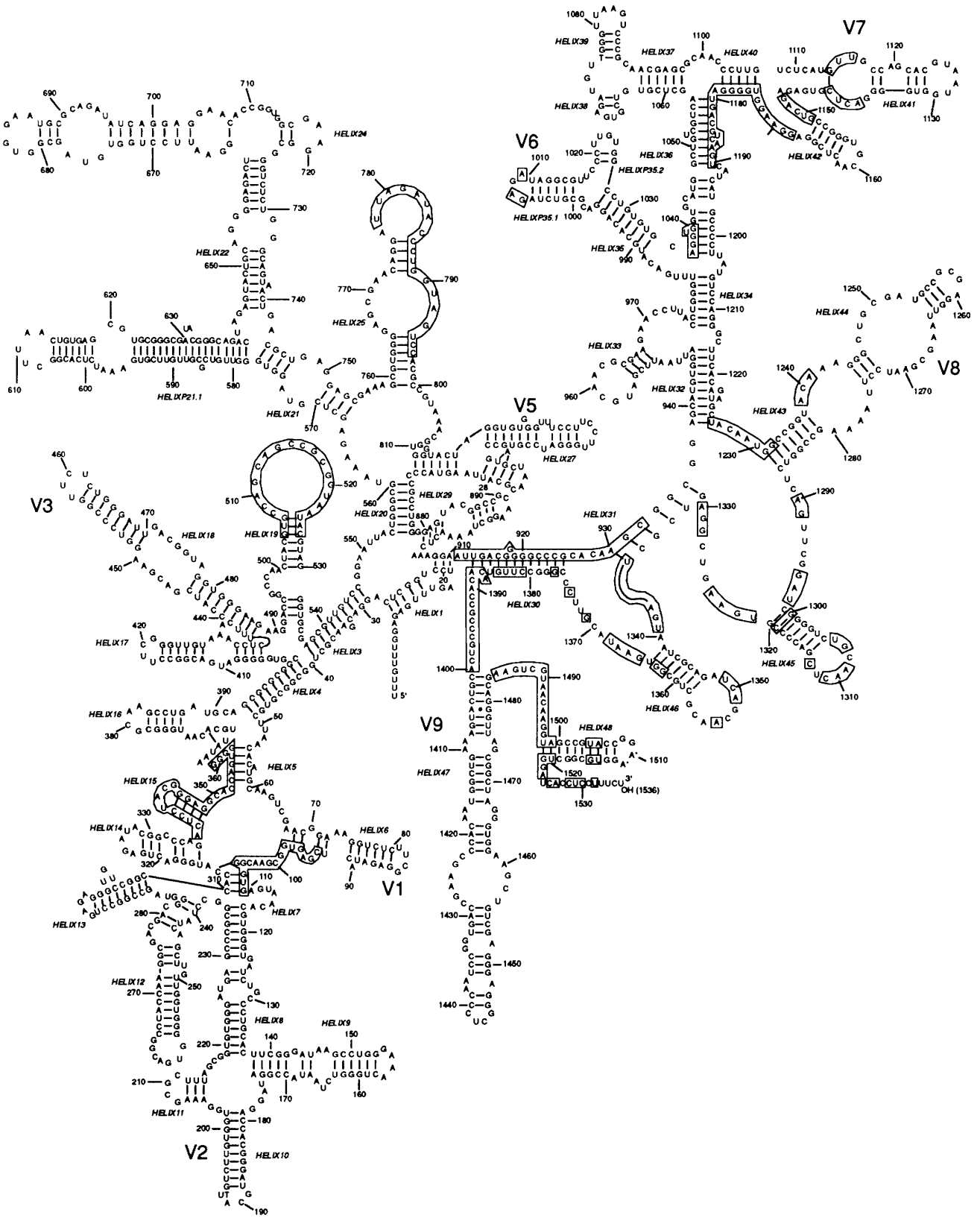


Fig. 3. Schematic secondary structure of *M. tuberculosis* 16S rRNA. The sequence was inferred from the gene (see Fig. 2). The scheme is based on that of Noller (1984) for *E. coli* and is annotated as described by Dams *et al.* (1988). The boxed nucleotides are highly conserved among eubacteria. V1, V2 etc. are variable regions. The *M. bovis* sequence (Suzuki *et al.*, 1988*a*) is identical except for the deletion of a C residue (position 1443) in the hairpin loop region of helix 47.

Table 2. *Shine–Dalgarno and anti-Shine–Dalgarno sequences of the M. tuberculosis complex*

Sequences are aligned to maximize G–C basepair formation between Shine–Dalgarno and anti-Shine–Dalgarno sequences.
N_x ranges from four to seven nucleotides in *E. coli* mRNA.
Guanine residues participating in 'wobble' (G–U) basepairing are denoted by use of a lower-case letter 'g'.

Gene	Sequence
E. coli	
16S rRNA (anti-Shine–Dalgarno)*	3′ UUCCUCC 5′
mRNA (hypothetical Shine–Dalgarno)	5′ AAGGAGG - - - N_x - - A/G UG
M. tuberculosis complex	
16S rRNA (anti-Shine–Dalgarno)†	3′ UCUUUCCUCC 5′
mRNA (hypothetical Shine–Dalgarno)	5′ AGAAAGGAGG - - - N_x - - A/G UG
M. tuberculosis mRNA (Shine–Dalgarno)‡	
dnaJ analogue	5′ - A g AGG g GG - - - N_9 - - GUG
10–12 kDa antigen	5′ - - - - - GGAGGG - N_8 - - GUG
19–22 kDa antigen	5′ - - AAAGGAG - - - N_6 - - GUG
32 kDa antigen	5′ - - g A g GGA - - - - N_7 - - AUG
38 kDa antigen	5′ AGAAAGG - - - - - N_6 - - GUG
65 kDa antigen	5′ - - - - - GGAGGAA - N_{10} - - AUG
M. bovis BCG mRNA (Shine–Dalgarno)‡	
α antigen	5′ - - AAAGG g G - - - N_7 - - AUG
18 kDa antigen	5′ - - g AAGGAG - - - N_8 - - AUG
23 kDa antigen	5′ - - - - - GGAGGAA - N_7 - - GUG

* Shine & Dalgarno (1974).
† See Fig. 3 and Suzuki *et al.* (1988*a*).
‡ For a compilation of mRNA sequences see Dale & Patki (1990).

Transcription of the M. tuberculosis rrn operon and processing of the transcript

Sequence elements implicated in the control of transcription of the *rrn* operon and in processing the pre-rRNA transcript to form mature 16S and 23S rRNA (see Fig. 2) were identified by comparison with elements of known function first established for *Bacillus subtilis* (Loughney *et al.*, 1983; Ogasawara *et al.*, 1983; Stewart & Bott, 1983) and *E. coli* (Young & Steitz, 1978; Berg *et al.*, 1989; Gourse *et al.*, 1989; Li *et al.*, 1984). These elements have since been recognized in a wide range of bacteria (for review see King *et al.*, 1986).

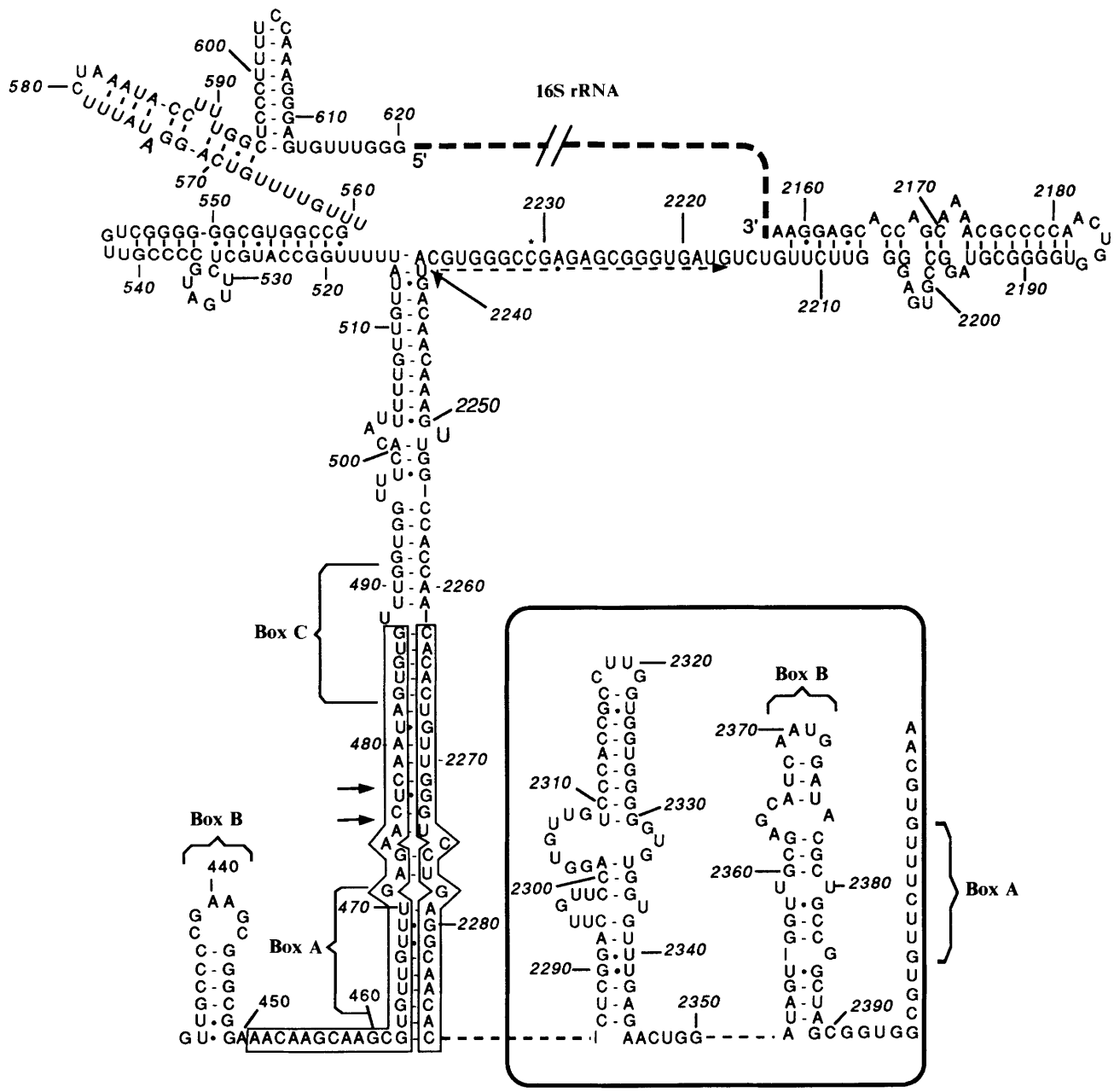
In total, 11 putative elements were identified, namely −35 box, −10 box, the 5′-end of precursor rRNA (or start of transcription), a stem structure formed between the leader region of the transcript and the intergenic (spacer) sequences separating mature 16S rRNA and 23S rRNA, RNAase III processing sites within this stem, and also a second potential processing site in the spacer region, the antitermination signals Box A, Box B, Box C in the leader region and Box A and Box B but no Box C in the spacer region, as discussed below.

Consensus sequences established for *rrn* operons of *B. subtilis* include the −35 box [5′TTGAC(T/A)3′], −10 box [5′TA(T/C)(T/A)(T/A)T3′] and the site for the start of transcription (see below). The corresponding sequences for the *rrn* operon of *M. tuberculosis* (see Fig. 2) are −35 box (5′-TTGACT3′) and −10 box (5′TAGACT3′).

The start site for transcription of the *rrn* operon is implicated in the growth-rate regulation and stringent control of transcription. Transcription of tRNA genes is regulated in a similar way (Duester *et al.*, 1982). The 5′-end of transcripts from tRNA and rRNA genes have features which distinguish them from transcription of genes coding for proteins (Ogasawara *et al.*, 1983). The established sequences include 5′GTGGTAG3′ for *E. coli* $tRNA^{leu}$ (Duester *et al.*, 1982); 5′GCCGGCG3′ for the *E. coli rrnG* operon (Li *et al.*, 1984); 5′GTTGTTA3′ for *B. subtilis* tRNA (Ogasawara *et al.*, 1983); and 5′G(T/C)(T/C)G(C/T)TT3′ for *B. subtilis rrn* operons (Ogasawara *et al.*, 1983). The corresponding sequence for *M. tuberculosis* (see Fig. 2) is 5′GTTGCCC3′, which conforms to the general pattern.

Antitermination signals which influence the response of RNA polymerase to pause sites and terminator sites have been found in both the leader and spacer regions of the *rrn* operons of *E. coli* (Berg *et al.*, 1989; Gourse *et al.*, 1989; Li *et al.*, 1984). The *E. coli rrn* antitermination

Fig. 4(*a*)



motif appears to be widespread (Berg *et al.*, 1989; King *et al.*, 1986). The elements Box A, Box B and Box C, which were first identified in bacteriophage λ, were subsequently found in *rrn* operons of *E. coli*, but in the order Box B, Box A and Box C in the leader region and Box B and Box A (but no Box C) in the spacer region (see, for example, Berg *et al.*, 1989). The three motifs (Box B, Box A and Box C) are found within approximately 70 bp immediately following the *E. coli rrnG* P2 promoter (Li *et al.*, 1984). These motifs have their homologues within 65 bp downstream from the putative start of transcription of the *rrn* operon of *M. tuberculosis* (see Fig. 2).

Box B is a region of hyphenated dyad symmetry with the potential for the RNA-like strand to form a hairpin loop. The sequence 432–446 (5′GCCCGAAGCGGGC3′; Fig. 2) is very similar to the Box B element (5′GCCCTGAAGAAGGGC3′) of the *nutL* gene of bacteriophage λ (Friedman & Gottesman, 1983). The Box B element tentatively identified in the spacer region (2366–2375) has a different sequence.

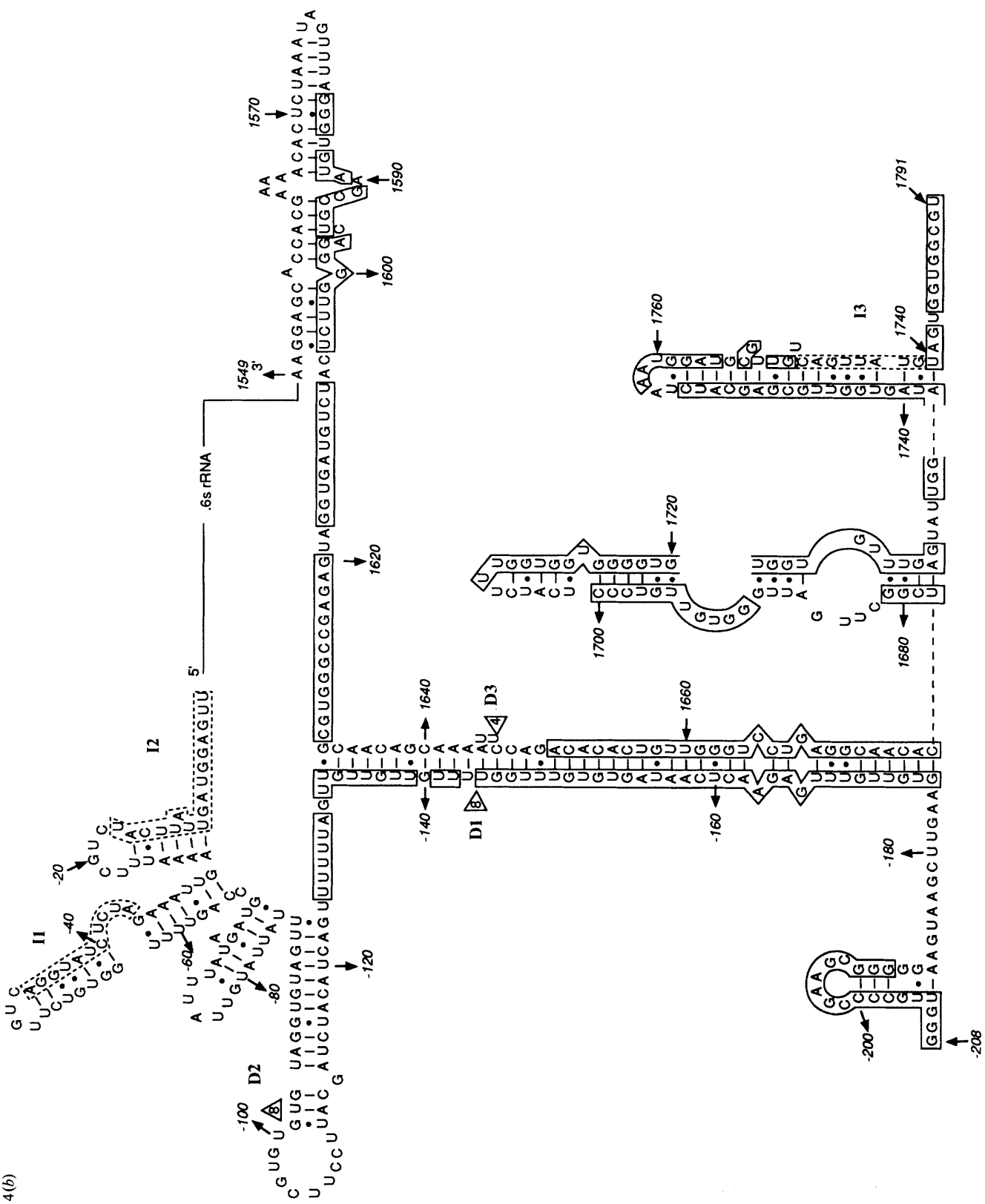

I1
I2
I3
D1
D2
D3
5'
3'
16s rRNA
1549
1570
1590
1600
1620
1640
1660
1680
1700
1720
1740
1760
1791
-20
-40
-60
-80
-100
-120
-140
-160
-180
-200
-208

Fig. 4(*b*)

Fig. 4(*c*)

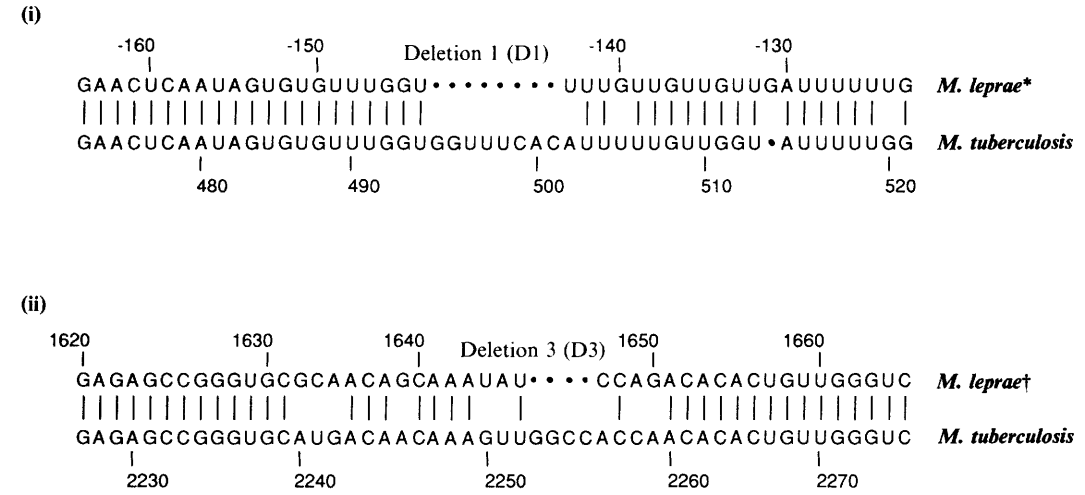


Fig. 4. Scheme for the secondary structure of precursor rRNA of *M. tuberculosis* H37Rv and *M. leprae*. (*a*) *M. tuberculosis* precursor rRNA. The region of the stem which is similar to *B. subtilis* pre-rRNA (Ogasawara *et al.*, 1983) is boxed. The arrows indicate possible RNAase III cleavage sites. Box A, Box B and Box C indicate putative antitermination signals (Li *et al.*, 1984). Residues 2287–2390 implicated in the processing of 23S rRNA are framed. ◄------► denotes an imperfect 12 bp palindrome; mismatched cytosine residues are marked by an asterisk. Nucleotides are numbered as in Fig. 2. (*b*) Comparison of the calculated secondary structures of *M. tuberculosis* and *M. leprae* precursor rRNAs. The boxes indicate regions of identical sequence: regions I1, I2 and I3, enclosed by a broken line, are sequences not present in *M. tuberculosis*; D1 ⧐, D2 ⧐ and D3 ⧐ indicate where a number of extra nucleotides (number indicated in the triangle) are present in the *M. tuberculosis* sequence. These sequences specific for *M. tuberculosis* are shown in (*c*). Nucleotides are numbered from the 5′-end of the 16S rRNA coding region. The minus sign denotes upstream sequences. The sequence of the spacer region was reported by Liesack *et al.* (1991). (*c*) Comparison of the nucleotide sequences of *M. tuberculosis* and *M. leprae* in (i) the leader, and (ii) the spacer regions of pre-rRNA which are believed to interact through basepairing to form a stem structure, as shown in (*a*) and (*b*). * EMBL Data Library accession number X58891; † Liesack *et al.* (1991).

The Box A motif has a more conserved sequence which is 5′(C/T)GCTCTT(T)A3′ in bacteriophage λ (Friedman & Gottesman, 1983), 5′TGCTCTTTA3′ in the *rrnG* operon of *E. coli* (Li *et al.*, 1984) and 5′AGTTCTTTG3′ in the *rrnO* operon of *B. subtilis* (Ogasawara *et al.*, 1983). Homologous sequences are found in positions 463–471 (5′TGTTGTTTG3′) and 2398–2406 (5′TGTTCTTTG3′) of Fig. 2.

The Box C element has the consensus sequence 5′GG(T/C)GT(G/A)(T/C)G3′ in bacteriophage λ (Friedman & Gottesman, 1983) and 5′TGTGTGGG3′ in the *rrnG* operon of *E. coli* (Li *et al.*, 1984); the *M. tuberculosis* homologue appears to be 5′AGTGTGTTT3′ (positions 483–498, Fig. 2).

The putative elements described above are in accord with the notion that the leader sequence of pre-rRNA of *M. tuberculosis* comprises 191 nucleotides, compared with 110–175 nucleotides noted for the leader region of *rrn* operons of *B. subtilis* (Ogasawara *et al.*, 1983). A potential secondary structure was derived for the transcript by the method of Staden (1984), including the putative stem structure formed by basepairing between leader and spacer sequences (see Fig. 4*a*), described by Young & Steitz (1978) for *E. coli* and by Ogasawara *et al.* (1983) for *B. subtilis*. The stem region (Fig. 4*a*) has 82% homology with its *B. subtilis* counterpart. The *B. subtilis* sequence (Ogasawara *et al.*, 1983) is 5′GUUCUUUGA-AAA↓CU↓AAA3′ (the arrows indicate RNAase III cleavage sites) compared with the *M. tuberculosis* sequence 5′GUUGUUUGAGAA↓CU↓CAA3′ (the broken arrows indicate possible RNAase III cleavage sites). Part of the leader region (positions 464–478) is largely repeated in the spacer region (positions 2396–2416), namely 5′GCGUGUUGUUUGUGCAA-UAC3′ (see Figs 2 and 4*a*). This repeat sequence is a candidate for one strand of a second stem structure (including an RNAase III processing site) formed by interaction with sequences downstream from the 3′-end of 23S rRNA (cf. *B. subtilis*; Ogasawara *et al.*, 1983). In the *rrn* operons of *B. subtilis* (Stewart & Bott, 1983), Mycoplasma (Taschke & Herrman, 1986) and Streptomyces (Pernodet *et al.*, 1989), the Box A motif is located within the stem structure close to the RNAase III processing site, as we have found for *M. tuberculosis* (see Fig. 4*a*).

The spacer region (Fig. 2) comprises 276 bp, which is the same length as the *M. bovis* BCG spacer (Suzuki *et al.*, 1988*a*) but which is 6 bp shorter than the *M. leprae* spacer (Liesack *et al.*, 1991). We did not detect tRNA genes in the spacer region (see Fig. 2). In this respect,

M. tuberculosis resembles *M. bovis* BCG (Suzuki *et al.*, 1988*a*), *M. leprae* (Liesack *et al.*, 1991), both operons of *Mycoplasma capricolum* (Sawada *et al.*, 1984), the single operon of *Mycoplasma hyopneumoniae* (Taschke & Herrmann, 1986), *Streptomyces ambofaciens* (Pernodet *et al.*, 1989), *Streptomyces coelicolor* (Bayliss & Bibb, 1988), *Streptomyces lividans* TK21 (Suzuki *et al.*, 1988*b*), and eight out of ten of the *B. subtilis rrn* operons (Loughney *et al.*, 1982). However, tRNA genes have been located downstream of the *rrn* operon of *M. tuberculosis* (Bhargava *et al.*, 1990).

M. leprae precursor rRNA

The nucleotide sequence of the *Sau*3A/*Pst*I fragment of the *M. leprae rrn* operon was established and was combined with published data for the spacer region separating the 16S rRNA and 23S rRNA genes (Liesack *et al.*, 1991) to generate a possible secondary structure for the putative precursor rRNA (see Fig. 4*b*). Comparison of Figs 4(*a*) and 4(*b*) reveals that the same overall scheme of secondary structure may apply to both *M. tuberculosis* and *M. leprae*, although there are appreciable differences in detail. For example, the putative stringent response and antitermination signals, and RNAase III processing sites are very similar; however, the leader sequence is longer in *M. leprae* (208 nucleotides) than in *M. tuberculosis* (191 nucleotides).

The leader sequence presented in Fig. 4(*b*) differs from the published sequence (Sela & Clark-Curtiss, 1991) by the presence of two additional C residues in the putative Box B antitermination signal forming a run of four consecutive C residues.

The differences in the secondary structure of precursor rRNA of *M. leprae* and *M. tuberculosis* are attributable to insertions and deletions in the *M. leprae* leader and spacer regions compared with *M. tuberculosis* (see Fig. 4*c*). The bihelical stem region formed between part of the leader sequence and part of the spacer region provides evidence for compensating changes; for example, deletion D1 [Fig. 4*c*(i)] in the leader sequence is matched by deletion D3 in the spacer region [Fig. 4*c*(ii)]. The high degree of homology (see Fig. 4*b*) between the secondary structures proposed for *M. leprae* and *M. tuberculosis*, together with the evidence for compensating changes, lends support to the schemes proposed in Fig. 4.

Discussion

Members of the *M. tuberculosis* family have a single *rrn* operon (K. E. Kempsell, I. C. E. Estrada-G, M. J. Colson & R. A. Cox, unpublished work). They also have fewer ribosomes than many other bacteria. However, their complement of ribosomes varies according to the growth rate (Winder & Rooney, 1970). More than 80% of the RNA fraction of a bacterium is rRNA, so that the ribosome complement is reflected in the RNA:DNA ratio. In the mid-exponential phase this ratio is 2:1 for *M. tuberculosis*, 4:1 for *Mycobacterium smegmatis* and 20:1 for *E. coli* (Winder & Rooney, 1970), that is the complement of ribosomes are in the proportions 1:2:10. *M. tuberculosis* has a single *rrn* operon, *M. smegmatis* has two (Bercovier *et al.*, 1986) and *E. coli* has seven (Kiss *et al.*, 1977). The single operon of *M. tuberculosis* provides about 70% of the number of ribosomes synthesized by one operon of *E. coli*, on the basis of the assumption that the seven operons are equivalent. The evidence that *M. tuberculosis* maintains growth-rate-dependent control of ribosome biosynthesis rests on the observation that the RNA:DNA ratio varies according to growth conditions whereas the ratio RNA:protein remains unchanged (Winder & Rooney, 1970). The sequence data presented above for *M. tuberculosis* has similarities with elements of the promoter and leader regions of the *rrn* operons of *B. subtilis* and *E. coli*, suggesting that they have similar mechanisms for the control of rRNA synthesis.

On the basis of its similarity with *M. tuberculosis* precursor rRNA, the single *rrn* operon of *M. leprae* appears functional (see also Sela & Clark-Curtiss, 1991). Indeed, the RNA:DNA ratio of *M. leprae* indicates that each cell contains at least 2000 ribosomes (Estrada-G *et al.*, 1988). Thus it is unlikely that the exceptionally slow growth rate of *M. leprae* arises from a defect in its capacity for ribosome biosynthesis. A more likely explanation is that, for other reasons, *M. leprae* is incapable of independent growth and survives as an obligate parasite (for review see Wheeler, 1990).

The sequence of the *rrn* operon of *M. bovis* extending from 100 bp upstream from the start of the 16S rRNA coding region through the spacer region to the start of the 23S rRNA coding region (2007 bp) was reported previously (Suzuki *et al.*, 1988*a*). There are four differences compared with the equivalent *M. tuberculosis* sequence. The single difference in the 16S rRNA coding region was discussed above; the three other differences are found in the spacer region [positions 2169(G/C), 2170(C/G) and 2261(A/C)] as shown in Fig. 2. These four differences in 2007 bp result in 99·8% similarity in sequence. Members of the *M. tuberculosis* complex are also very closely related (98% similarity in 16S rRNA sequences) to other slow-growing mycobacteria (Rogall *et al.*, 1990; Stahl & Urbance, 1990; Cox *et al.*, 1991). This high degree of similarity in sequence suggests that the principal features of the leader region of the *M. tuberculosis rrn* operon are present not only in other members of the *M. tuberculosis* complex but also in other slow-growing mycobacteria. We have presented supporting evidence to show that this inference is true in the case of *M. leprae*.

This investigation received support from the UNDP/WHO Special Programme for Research Training in Tropical Diseases, and from the British Leprosy Relief Association (LEPRA). Dr Yuan-en Ji is supported by a grant from LEPRA.

References

BAYLISS, H. A. & BIBB, M. J. (1988). Transcriptional analysis of the 16S rRNA gene of the *rrnD* gene set of *Streptomyces coelicolor* A3(2). *Molecular Microbiology* **2**, 569–579.

BERG, K. L., SQUIRES, C. & SQUIRES, C. L. (1989). Ribosomal RNA operon anti-termination: function of leader and spacer region Box B–Box A sequences and their conservation in diverse micro-organisms. *Journal of Molecular Biology* **209**, 345–358.

BERCOVIER, H., KAFRI, O. & SELA, S. (1986). Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochemical and Biophysical Research Communications* **136**, 1136–1141.

BHARGAVA, S., TYAGI, A. K. & TYAGI, J. S. (1990). tRNA genes in mycobacteria: organization and molecular cloning. *Journal of Bacteriology* **172**, 2930–2934.

COX, R. A., KEMPSELL, K., FAIRCLOUGH, L. & COLSTON, M. J. (1991). The 16S ribosomal RNA of *Mycobacterium leprae* contains a unique sequence which can be used for identification by the polymerase chain reaction. *Journal of Medical Microbiology* **35**, 284–290.

DALE, J. W. & PATKI, A. (1990). Mycobacterial gene expression and regulation. In *Molecular Biology of the Mycobacteria*, pp. 173–198. Edited by J. McFadden. London: Surrey University Press.

DAMS, E., HENDRIKS, L., VAN DE PEER, Y., NEEFS, J. M., SMITS, G., VENDENBEMPT, L. & DE WACHTER, R. (1988). Compilation of small subunit RNA sequences. *Nucleic Acids Research* **16**, r87–r173.

DUESTER, G., ELFORD, R. M. & HOLMES, W. M. (1982). Fusion of the *Escherichia coli* $tRNA^{Leu}$ promoter to the *galK* gene: analysis of sequences necessary for growth-rate-dependent regulation. *Cell* **30**, 855–864.

ESTRADA-G, I. C. E., COLSTON, M. J. & COX, R. A. (1989). Determination and evolutionary significance of nucleotide sequences near to the 3′-end of 16S ribosomal RNA of mycobacteria. *FEMS Microbiology Letters* **61**, 285–290.

FRIEDMAN, D. I. & GOTTESMAN, M. (1983). The lytic mode of lambda development. In *Lambda II*, pp. 21–52. Edited by R. Hendrix, J. Roberts, F. Stahl & R. Weisberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

GOURSE, R. L., DE BOER, H. A. & NOMURA, M. (1989). DNA determinants of rRNA synthesis in *E. coli* growth rate dependent regulation, feedback inhibition, upstream activation antitermination. *Cell* **44**, 197–205.

KING, T. C., SIRDESKMUKH, R. & SCHLESSINGER, D. (1986). Nucleolytic processing of ribonucleic acid transcripts in procaryotes. *Microbiological Reviews* **50**, 428–451.

KISS, A., SAIN, B. & VENETIANER, P. (1977). The number of rRNA genes in *Escherichia coli*. *FEBS Letters* **79**, 77–79.

LI, S. C., SQUIRES, C. L. & SQUIRES, C. (1984). Antitermination of *E. coli* rRNA transcription is caused by a control region segment containing *nut*-like sequences. *Cell* **38**, 851–860.

LIESACK, W., PITULLE, C., SELA, S. & STACKEBRANDT, E. (1990). Nucleotide sequence of the 16S rRNA from *Mycobacterium leprae*. *Nucleic Acids Research* **18**, 5558.

LIESACK, W., SELA, S., BERCOVIER, H., PITULLE, C. & STACKEBRANDT, E. (1991). Complete nucleotide sequence of the *Mycobacterium leprae* 23S and 5S rRNA genes plus flanking regions and their potential in designing diagnostic oligonucleotide probes. *FEBS Letters* **281**, 114–118.

LOUGHNEY, K., LUND, E. & DAHLBERG, J. E. (1982). tRNA genes are found between the 16S and 23S rRNA genes in *Bacillus subtilis*. *Nucleic Acids Research* **10**, 1607–1624.

LOUGHNEY, K., LUND, E. & DAHLBERG, J. E. (1983). Ribosomal RNA precursors of *Bacillus subtilis*. *Nucleic Acids Research* **11**, 6709–6721.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

MURPHY, G. & WARD, E. S. (1989). Sequencing of double-stranded DNA. In *Nucleic Acids Sequencing, a Practical Approach*, pp. 99–115. Edited by C. J. Howe & E. S. West. Oxford: IRL Press.

NOLLER, H. (1984). Structure of ribosomal RNA. *Annual Review of Biochemistry* **53**, 119–231.

OGASAWARA, N., MORIYA, S. & YOSHIKAWA, H. (1983). Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. *Nucleic Acids Research* **11**, 6301–6318.

PERNODET, J.-L., BOOCARD, F., ALEGRE, M.-T., GAGNAT, J. & GUERINEAU, M. (1989). Organization and nucleotide sequence analysis of a ribosomal RNA cluster from *Streptomyces ambofaciens*. *Gene* **59**, 33–46.

ROGALL, T., FLOHR, T. & BOTTGER, E. C. (1990). Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. *Journal of General Microbiology* **136**, 1915–1920.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.

SAWADA, M., MUTO, A., IWAMI, M., YAMAO, F. & OSAWA, S. (1984). Organization of ribosomal RNA genes in *Mycoplasma capricolum*. *Molecular and General Genetics* **196**, 311–316.

SELA, S. & CLARK-CURTISS, J. E. (1991). Cloning and characterization of the *Mycobacterium leprae* putative ribosomal RNA promoter in *Escherichia coli*. *Gene* **98**, 123–127.

SHEPARD, C. C. (1960). The experimental disease that follows the injection of human leprosy bacilli into the footpads of mice. *Journal of Experimental Medicine* **112**, 445–454.

SHINE, J. & DALGARNO, L. (1974). The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 1342–1346.

STADEN, R. (1984). Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research* **12**, 505–519.

STAHL, D. A. & URBANCE, J. W. (1990). The division between fast- and slow-growing species correspond to natural relationships among the mycobacteria. *Journal of Bacteriology* **172**, 116–124.

STEWART, G. C. & BOTT, K. F. (1983). DNA sequence of the tandem ribosomal RNA promoter for *B. subtilis* operon *rrnB*. *Nucleic Acids Research* **11**, 6289–6300.

SUZUKI, Y., NAGATA, A., ONO, Y. & YAMADA, I. (1988*a*). Complete nucleotide sequence of the 16S rRNA gene of *Mycobacterium bovis* BCG. *Journal of Bacteriology* **170**, 2886–2889.

SUZUKI, Y., ONO, Y., NAGATA, A. & YAMADA, T. (1988*b*). Molecular cloning and characterization of an rRNA operon in *Streptomyces lividans* TK21. *Journal of Bacteriology* **170**, 1631–1636.

TASCHKE, C. & HERRMANN, R. (1986). Analysis of transcription and processing signals of the 16S–23S rRNA operon of *Mycoplasma hyopneumoniae*. *Molecular and General Genetics* **205**, 434–441.

WAYNE, L. G. (1982). Microbiology of tubercle bacilli. *American Review of Respiratory Diseases* **125**, 31–41.

WHEELER, P. R. (1990). Recent research into the physiology of *Mycobacterium leprae*. *Advances in Microbial Physiology* **31**, 70–124.

WINDER, F. G. & ROONEY, S. A. (1970). Effects of nitrogenous components of the medium on the carbohydrate and nucleic acid content of *Mycobacterium tuberculosis* BCG. *Journal of General Microbiology* **63**, 29–39.

WINSHIP, P. R. (1989). An improved method for directly sequencing PCR amplified material by using dimethylsulphoxide. *Nucleic Acids Research* **17**, 1266.

YOUNG, R. A. & STEITZ, J. A. (1978). Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 3503–3597.