Nucleotide sequences of the spacer-1, spacer-2 and trailer regions of the *rrn* operons and secondary structures of precursor 23S rRNAs and precursor 5S rRNAs of slow-growing mycobacteria

Yuan-en Ji,¹ Karen E. Kempell,^{1†} M. Joseph Colston² and Robert A. Cox¹

Author for correspondence: Robert A. Cox. Tel: +44 81 959 3666. Fax: +44 81 906 4477.

Laboratory of Developmental Biochemistry¹ and Leprosy and Mycobacterial Research², National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The single ribosomal RNA (*rrn*) operons of slow-growing mycobacteria comprise the genes for 16S, 23S and 5S rRNA, in that order. PCR methodology was used to amplify parts of the *rrn* operons, namely the spacer-1 region separating the 16S rRNA and 23S rRNA genes and the spacer-2 region separating the 23S rRNA and 5S rRNA genes of *Mycobacterium avium*, *Mycobacterium intracellulare*, '*Mycobacterium lufu*' and *Mycobacterium simiae*. The amplified DNA was sequenced. The spacer-2 region, the 5S rRNA gene, the trailer region and the downstream region of the *rrn* operon of *Mycobacterium tuberculosis* were cloned and sequenced. These data, together with those obtained previously for *Mycobacterium leprae*, were used to identify putative antitermination signals and RNase III processing sites within the spacer-1 region. Notable features include two adjacent potential Box B elements and a Box A element. The latter is located within a sequence of 46 nucleotides which is very highly conserved among the slow-growers which were examined. The conserved sequence has the capacity to interact through base-pairing with part of the spacer-2 region. Secondary structures for mycobacterial precursor 23S rRNA and for precursor 5S rRNA were devised, based on sequence homologies and homologous nucleotide substitutions. All the slow-growers, including *M. leprae*, conform to the same scheme of secondary structure. A putative motif for the intrinsic termination of transcription was identified approximately 33 bp downstream from the 3'-end of the 5S rRNA gene. The spacer-1 and spacer-2 sequences may prove a useful supplement to 16S rRNA sequences in establishing phylogenetic relationships between very closely related species.

Keywords: slow-growing mycobacteria, *rrn* operon, antitermination elements, precursor 23S rRNA, precursor 5S rRNA

INTRODUCTION

Mycobacteria are acid-fast Gram-positive bacteria. Many members of this genus are characterized by slow growth. These slow-growers include the human pathogens *Mycobacterium leprae* and *Mycobacterium tuberculosis*, which have doubling times of 12 d and 1 d respectively

(Shepard, 1960; Winder & Rooney, 1970). Although ribosome biosynthesis is essential to cell proliferation, the strategy (or strategies) used by slow-growing mycobacteria to control the production of new ribosomes is not known. One facet of this strategy is the regulation of the biosynthesis of rRNA. The slow-growing mycobacteria have a single rRNA (*rrn*) operon (Bercovier *et al.*, 1986) comprising genes for the rRNA family in the order 16S rRNA, 23S rRNA and 5S rRNA (see Fig. 1). The operon is transcribed as a single unit (precursor rRNA or pre-rRNA) which is cleaved by RNases to precursors of 16S rRNA (pre-16S rRNA), 23S rRNA (pre-23S rRNA) and

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

The EMBL accession numbers for the nucleotide sequence data reported in this paper are X74054–X74063 and X75599–X75602.

Table 1. The mycobacterial strains used in this study

Species*	Strain	Approx. time for colonies to develop on agar medium (d)
<i>Mycobacterium avium</i> (<i>M. av.</i>)	NCTC 8559	14
<i>Mycobacterium intracellulare</i> (<i>M. in.</i>)	NCTC 10682	14
' <i>Mycobacterium lufu</i> ' (<i>M. lu.</i>)	†	14
<i>Mycobacterium simiae</i> (<i>M. si.</i>)	TMC 5135 (<i>M. habana</i>)	10
<i>Mycobacterium leprae</i> (<i>M. le.</i>)	Armadillo tissue	Non-culturable
<i>Mycobacterium tuberculosis</i> (<i>M. tu.</i>)	H ₃₇ R _v (NCTC 7416)	21

* The shortened name, shown in parentheses, is used in Fig. 2. *M. habana* was first described following its isolation from cases of pulmonary tuberculosis (Valdivia *et al.*, 1971), and has been found to be closely related to *M. simiae* serovar 1 (Meissner & Schroder, 1975), and currently isolates are formally referred to as *M. simiae* (Bruckner & Colonna, 1993). 'M. lufu' is an environmental mycobacterium isolate from Zaire.

† The strain was kindly provided by Dr S. R. Pattyn, Institute of Tropical Medicine, Antwerp, Belgium.

5S-rRNA (pre-5S rRNA). Mature rRNA species are derived from the precursors by further RNase treatment. Thus, correct processing reactions are essential for the synthesis of mature rRNA.

Previously we described the promoter region of the *rrn* operon of *M. leprae* (see also Sela & Clark-Curtiss, 1991) and *M. tuberculosis* (Kempsell *et al.*, 1992; Ji *et al.*, 1994), and identified the site for the start of transcription (Ji *et al.*, 1994). Also, the nucleotide sequences of the leader and spacer-1a regions of the *rrn* operon (see Fig. 1) were obtained for eight pathogenic mycobacteria. These data were used to identify a common scheme of secondary structure for pre-16S rRNA (Ji *et al.*, 1994).

This report describes further analysis of the *rrn* operons of *Mycobacterium avium*, *Mycobacterium intracellulare*, '*Mycobacterium lufu*', which is as sensitive to diamino diphenyl sulphone as *M. leprae* (Portaels, 1980), a strain of *Mycobacterium simiae* which is as effective as *M. leprae* in protecting mice against infection with *M. leprae* (Singh *et al.*, 1989), *M. tuberculosis* and *M. leprae*. The results provide insight into the primary and secondary structures of pre-23S rRNA, pre-5S rRNA and the trailer regions of pre-rRNA.

METHODS

Materials. All general chemicals were of analytical grade. Polynucleotide kinase and T₄ ligase were supplied by Pharmacia. Sequenase (USB) sequencing kit was purchased from Cambridge Biosciences. [³⁵S]dATP-αS and [γ-³²P]ATP were from Amersham. GeneClean kit was purchased from Bio101. AmpliTaq DNA polymerase was supplied by Perkin Elmer Cetus. Oligonucleotide primers were prepared with an automated DNA synthesizer (model 370A, Applied Biosystems).

Bacterial strains and cultures. Mycobacterial strains are listed in Table 1, together with an indication of the growth rate. All mycobacterial cultures were maintained on Löwenstein-Jensen agar slopes and grown for use in liquid culture at 37 °C in modified Dubos medium (Dubos & Davis, 1946). To modify mycobacterial cell walls for subsequent genomic DNA isolation,

glycine was added to 200 mM after 2 weeks growth and incubation continued for a further 2 weeks, after which time the cells were harvested. Other bacterial cells were grown on LB medium (Maniatis *et al.*, 1982) at 37 °C. Competent cells of *Escherichia coli* strain DH5αF' were used for transformation with the recombinant plasmids pUC8, pUC18 and pBluescriptII KS (Ji *et al.*, 1994).

Isolation of DNA. Late-exponential-phase cells were harvested and washed twice in 50 mM Tris/HCl, 10 mM EDTA, 0.3 M sucrose and resuspended in the same buffer at 1 g wet wt cells per 10 ml. Lipase (type VII, Sigma) and lysozyme (egg white, Pharmacia) were added to give a final concentration for each of 2 mg ml⁻¹ and the reaction mixture was kept at 37 °C for 2 h. The cells were centrifuged and resuspended in 6 M guanidinium chloride, 0.1% (v/v) Tween 80, 10 mM EDTA, 1 mM 2-mercaptoethanol and left at -20 °C for 16 h. An equal volume of chloroform was added to the suspension and mixed vigorously. After centrifugation the supernatant was removed to a fresh tube, overlaid with 2.5 vols ice-cold ethanol and mixed by inversion. The DNA precipitate was recovered by centrifugation and washed with 70% (v/v) ethanol. This was then resuspended in 10 mM Tris/HCl (pH 7.6), 10 mM EDTA, 100 mM NaCl, 1.0% (w/v) SDS, 0.5 mg proteinase K ml⁻¹ at 37 °C for 4–6 h. The SDS was removed by incubation at 4 °C for 2 h then centrifugation. The supernatant was subjected to two rounds of chloroform extraction; the DNA was then precipitated and resuspended in TE buffer by conventional means.

DNA polymerase chain reaction (PCR). Bacterial DNA (1–100 ng) was subjected to PCR in a total volume of 50 µl, with 1 unit of *Taq* polymerase (Advanced Biotechnologies), 50 mM KCl, 10 mM Tris/HCl pH 8.3, 1.5 mM MgCl₂, 0.01% gelatin, 100 pmol of each of two appropriate primers and 200 µM of each dNTP (dATP, dCTP, dGTP, dTTP). The 50 µl mixture was covered by 50 µl light mineral oil. The relevant gene fragment coding for spacer-1 was synthesized using primer combination 1 (5'-GCC AAG GCA TCC ACC ATG C-3'; and 5'-ATT GAC GGG GGC CCG CAC AAG CG-3') as described by Ji *et al.* (1994). The gene fragment coding for spacer-2 was synthesized using primer combination 2 (primer S2-1, 5'-CCA GTT CTC TCA CCA GGG G-3'; and S2-2, 5'-CAC ACA CAT CGC AAC CAC-3'). The target for primer S2-1 is a sequence located approximately 210 bp upstream from the 3'-end of the

Fig. 1. Strategies used in this investigation. (a) *rrn* operon of *M. tuberculosis* and other slow-growing mycobacteria. The *M. tuberculosis* 5.4 kbp *Bam*HI fragment was subcloned into plasmid pUC8 as described previously (Kempsell *et al.*, 1992). Appropriate regions were sequenced (see Methods) as indicated in the enlarged section. The filled circles show the locations of the primers used. Regions of genomic DNA were also amplified and sequenced (see Methods). The locations of the target sites for primers used for amplification are shown by arrowheads. (b) The transcript, pre-rRNA, of the *rrn* operon (approx. 5.3 kbp). The interaction through base-pairing between the 5'-flanking (leader) region and part of the spacer-1 (spacer-1a), which brings the 5'- and 3'-ends of 16S rRNA together, is shown schematically. The interaction between the remaining part of spacer-1 (spacer-1b) region and spacer-2, which brings the 5'- and 3'-ends of 23S rRNA together, is also indicated. (c) Processing of pre-rRNA yielding intermediate products. The interactions between leader and spacer-1a sequences in pre-16S rRNA and between spacer-1b and spacer-2 in pre-23S rRNA are represented by dots. The structure of pre-16S rRNA was described previously (Ji *et al.*, 1994). Pre-23S rRNA, pre-5S rRNA and the end of transcription are the focal points of this investigation.

23S rRNA gene of *M. leprae* (position 3224–3242, EMBL Data Library accession number X56657; Liesack *et al.*, 1991). The target for primer S2-2 is positions 42–59 of the 5S rRNA gene (see Fig. 2). Amplification was achieved using 36 cycles. The reaction mixture was heated to 94 °C for 1 min, kept at 58 °C for 1 min, then heated to 72 °C for 2.5 min, and this cycle was repeated 35 more times. Finally, the solution was heated to 94 °C for 15 s, then kept at 58 °C for 1 min and at 72 °C for 5 min.

PCR cloning. The products of PCR were separated by agarose gel electrophoresis. DNA was recovered from the gel by using glass milk (Vogelstein & Gillespie, 1979). DNA was re-suspended in 20 µl water. Bluescript plasmid was digested with *Eco*RV and incubated with *Taq* polymerase (1 unit per µg plasmid; 20 µl vol.) using standard buffer conditions (50 mM KCl, 10 mM Tris/HCl pH 8.3, 1.5 mM MgCl₂ and 200 µg BSA ml⁻¹) in the presence of 2 mM dTTP. The modified plasmid was purified and ligated with amplified DNA as described by Ji *et al.* (1994).

DNA sequencing. At least three colonies from the same ligation reaction were selected for study. In each case the sequence of the inserted DNA was determined by standard procedures using the dideoxy chain-termination method, as described by Ji *et al.*

(1994). The newly synthesized DNA was labelled with [³⁵S]dATP-αS and the products were separated by polyacrylamide gel electrophoresis.

Alignment of sequences. Computer-aided analysis of the alignment of nucleotide sequences of the leader and spacer regions was performed using PILEUP, which is part of the GCG sequencing analysis software package (Devereux *et al.*, 1984). PILEUP can perform automatic multiple sequence alignments according to the relationship between the sequences.

RESULTS

The scope of this investigation and the sequencing strategies are summarized in Fig. 1. Sequence data were obtained for the 3'-end of 16S rRNA, the spacer-1 region, the 5'- and 3'-ends of the 23S rRNA gene, the spacer-2 region and the 5'-end of the 5S rRNA gene. The species studied were *M. avium*, *M. intracellulare*, '*M. lufu*' and *M. simiae*. The nucleotide sequence of the spacer-1 region and the 5'-end of 23S rRNA of the *rrn* operon of *M. tuberculosis* H₃₇R_V was reported previously (Kempsell *et al.*, 1992). Further sequence data were obtained for this myco-

Table 2. Length of spacer-1 spacer-2 and trailer regions of the *rrn* operon of mycobacteria and other species

For details of the mycobacterial spacers see Fig. 2 and Ji *et al.* (1994). (a) Sequence data of Liesack *et al.* 1991. (b) *M. bovis* (Suzuki *et al.*, 1988) has a sequence which is identical with that of *M. tuberculosis*. (c) *Streptomyces ambofaciens*: Pernodet *et al.* (1989). (d) The end of transcription was identified by a string of T-residues on the RNA-like strand downstream from the 5S rRNA gene. However, the absence of a nearby sequence of interrupted dyad symmetry is inconsistent with spontaneous termination. (e) *Bacillus subtilis*: Ogasawara *et al.* (1983). (f) Two tRNA genes separate spacer-1a and spacer-1b. –, Sequence not known.

Species	Length (bp)				
	Spacer-1			Spacer-2	Trailer
	Spacer-1a	Spacer-1b	Total		
<i>M. intracellulare</i>	130	150	280	83	–
<i>M. avium</i>	130	149	279	–	–
' <i>M. lufu</i> '	130	151	281	79	–
<i>M. simiae</i>	130	154	284	92	–
<i>M. leprae</i> ^(a)	129	157	286	85	31
<i>M. tuberculosis</i>	129	149	278	77	33
<i>M. bovis</i> ^(b)	129	149	278	–	–
<i>S. ambofaciens</i> ^(c)	201	102	303	81	41 ^(d)
<i>B. subtilis</i> ^(e)	80	81	339 ^(f)	68	38

bacterium downstream from the 3'-end of 23S rRNA, namely the spacer-2 region, the 5S rRNA gene, trailer and 3' flanking sequences.

The sizes of the spacer-1, spacer-2 and trailer regions are summarized in Table 2. The sequence data, which are summarized in Fig. 2, together with data for the leader and spacer-1a sequences obtained previously (Ji *et al.*, 1994), provide an overall view of the *rrn* operons. Nucleotide positions are related to the start of transcription. The relevant sequences of the *rrn* operon of *M. leprae* (Liesack *et al.*, 1991) are included for comparison.

Spacer-1

The nucleotide sequence of spacer-1 was obtained for *M. avium*, *M. intracellulare*, '*M. lufu*' and *M. simiae* by amplification of genomic DNA using an appropriate primer combination (see Fig. 1 and Methods). These data are presented in Fig. 2 together with the previously published sequences of *M. tuberculosis* (Kempell *et al.*, 1992), *M. bovis* (Suzuki *et al.*, 1989) and *M. leprae* (Liesack *et al.*, 1991). All the sequences conform to a general pattern which is described below.

Elements important to transcription

The efficiency with which the *rrn* operon is transcribed is influenced by antitermination signals which govern the response of RNA polymerase to potential pause sites and rho-dependent terminator sites. Three antitermination elements, Box A, Box B and Box C, were first identified in bacteriophage λ and were later found in the leader and

spacer-1 regions of the *rrnG* operon of *E. coli*. For convenience, the two sets are distinguished by the subscripts L (for leader) and S (for spacer). In the leader region of the *rrnG* operon the elements were found in the order Box B_L, Box A_L and Box C_L; and two elements Box B_S and Box A_S were found in the spacer region (Berg *et al.*, 1989).

Three putative elements, Box B_L, Box A_L and Box C_L, have a similar organization in the leader region of the *rrn* operon of slow-growing mycobacteria as they have in the *rrnG* operon of *E. coli* (Kempell *et al.*, 1992; Ji *et al.*, 1994). In the spacer-1b region (see Fig. 2) two potential Box B elements (Box B_S' and Box B_S'') are located in tandem, followed by potential Box A_S and Box C_S sequences.

In each of the six species Box B_S' is followed by a very highly conserved sequence of 46 nucleotide residues [positions 1921 to 1967 (S1-169 to S1-215)], and 44 positions are identical in all six species. All 46 positions were identical in five species and the exception, *M. tuberculosis*, differed from the others by the insertion of a G residue [position 1929 (S1-177)] and two substitutions [positions 1947 (S1-195) and 1949 (S1-197) Fig. 2]. Both Box A [positions 1936 (S1-184) to 1948 (S1-196)] and Box C [(positions 1955 (S1-203) to 1962 (S1-210)] lie within the highly conserved region. The motifs Box B_S', Box B_S'', Box A_S, Box C_S are present within a span of 76 bp, in accord with the disposition found for Box B, Box A and Box C in the leader region of the *rrnG* operon of *E. coli* (see for example Berg *et al.*, 1989).

The highly conserved 46 bp region of spacer-1b has features in common with the 29 bp invariant sequence

Fig. 2. Sequences of the spacer-1, 5'- and 3'-ends of the 23S rRNA coding region, spacer-2, 5S rRNA, trailer region and 3'-flanking sequences of the *rrn* operon of *M. tuberculosis* and several slow-growing mycobacteria. See Ji *et al.* (1994) for details of spacer-1a sequences. The sequences were aligned using PILEUP (see text). Except for *M. leprae* and *M. tuberculosis*, gene sequences (see Fig. 1) were amplified and sequenced as described in Methods. α , see Liesack *et al.* (1991); β , see Kempell *et al.* (1992). γ , the available sequence for the *rrn* operon of *M. bovis* (nucleotides 95–2152) (Suzuki *et al.*, 1988) is essentially the same as the *M. tuberculosis* sequence. δ , The nucleotide sequence of *M. simiae* serotype 1, strain 26110 Mycobacterium Reference Library, Cardiff, UK, was determined for positions 5127–5281, and is identical with that of *M. simiae* TMC 5135 (*M. habana*). Position 3 is the consensus site for the start of transcription (see Ji *et al.*, 1994). Positions 1753–1885 (S1-133) are spacer-1a sequences. Nucleotides 1886–2046 (S1-134–S1-294) inclusive comprise the spacer-1b region of pre-rRNA which extends from the 3'-end of the main stem of pre-

found in the leader region (Ji *et al.*, 1994). In particular, there is a high degree of homology between the putative elements Box A_L (5'-TGTGTTTGTGAGA-3') and Box A_S (5'-TGGTGTGTTGAG(T/A)-3') and between Box C_L (5'-GTGTGTTTGTG-3') and Box C_S (5'-GTGGTTG-3'), which differ by two insertions/deletions. There are also sequence similarities between the putative Box A_S of mycobacteria and Box A_S elements of other Gram-positive bacteria, for example, mycoplasma (5'-(A/G)G(C/T)TCTTT(C/G)AAA-3'; Taschke & Herrmann, 1986; Iwami *et al.*, 1984), *Bacillus subtilis* (5'-TGTCTCTTGAAGA-3'; Ogasawara *et al.*, 1983), and streptomycetes (5'-CGTGTGTTGTGAGA-3'; Baylis & Bibb, 1988; Pernodet *et al.*, 1989).

5'- and 3'-ends of mature 23S rRNA

The ends of mature 23S rRNA of mycobacteria have not been determined directly although sequences close to the termini have been identified by their homology with other published sequences. The 5'- and 3'-ends of mature 23S rRNA interact through base-pairing. Examination of 23S rRNA sequences (Gutell *et al.*, 1990) reveals that the 3'-end has the consensus sequence 5'-(G/A)(G/A)CUU(G/A) N_{5-8} -3'. The triplet 5'-CUU-3' is implicated in the interaction with sequences near to the 5'-end. For the mycobacterial species studied, the sequence 5'-UAAGU-3' located near to the 5'-end is complementary to five nucleotides of the consensus sequence. However, more extensive base-pairing is possible, as shown in Fig. 3, which also illustrates the possibility that the 5'-end of mature 23S rRNA is marked by a GU-rich sequence and that the 3'-end is marked by a CA-rich sequence. The 5'- and 3'-terminal residues indicated in

16S rRNA to nucleotide 1 of the 23S rRNA coding region. Box A_S , Box B_S and Box C_S are putative antitermination elements implicated in transcription of the gene. The possible secondary structures of the transcripts (see Fig. 6) are delineated by vertical divisions and horizontal arrows. Stem sequences have the potential to form base-pairs with spacer-2 sequences (and vice versa). 'Stem' marks a 45 nucleotide sequence (1921–1965) which is highly conserved in all six species and which may interact with the complementary, c (stem) sequence of spacer-2a. The 23S rRNA coding region (nucleotides 2047–5141) is boxed. The numbering within the box accords with the size of *M. leprae* 23S rRNA. The broken lines mark the hypothetical 5'- and 3'-ends of the 23S rRNA coding region (see Figs 3 and 6). The spacer-2 region comprises positions 5142–5246 (S2-1–S2-105); spacer 2a comprises positions 5142–5215 (S2-1–S2-74) and spacer 2b comprises positions 5216–5246 (S2-75–S2-105). The location of helix S2-1 is shown by (a), (b), etc; with the asterisk denoting the position of (d). The sequences of helix S2-1 (a)–(e) are presented as footnotes. Two forms of (b) are possible, as shown (see also Fig. 4). The 5S rRNA coding region is also boxed. *M. bo.* refers to *M. bovis* 5S rRNA sequences which are the same for *M. tuberculosis* except for the differences shown. (T) etc. refer to *M. bovis* BCG (Ohkubo *et al.*, 1986); [G] etc. refer to *M. bovis* ATCC 35737 (Dams *et al.*, 1987). RNase E cleavage site, either 5'-AAUUU-3' or 5'-AAUAA-3' (Ghora & Apirion, 1978). †, approximate 3'-end of pre-rRNA. Note that spacer-1 sequences were reported for strains of the *M. avium* complex by Fotheringham & Wilson (1993). n, unidentified nucleoside residue. ., deletion of nucleotide residue.

Fig. 3. Illustration of the possible interactions between the 5'- and 3'-ends of 23S rRNA in pre-23S rRNA. (a) *M. intracellulare*, (b) '*M. lufu*', (c) *M. simiae*. 23S rRNA sequences are boxed; the broken line indicates uncertainty in identifying the 5'- and 3'-terminal residues. The termini of mature 23S rRNA were assigned on the basis of the assumption that the 5'-end is generated by cleavage of a UpG bond located in a loop; and that the 3'-end is generated by cleavage of a CpA bond, also located in a loop. The arrows indicate these putative cleavage sites. Other cleavages in the vicinity of the arrows are possible, as indicated by the extent of the broken line. The putative 3'-end conforms to the consensus sequence (G/A)(G/A)CUU(G/A)N6 (see text). -, Hydrogen bonds formed between standard base-pairs; ·, hydrogen bonds formed between 'wobble' base-pairs.

Fig. 4. Helix S2-1, a variable feature of spacer-2, is located near the Box A₅ sequence in pre-rRNA. (a) *M. intracellulare*; (b) '*M. lufu*' [(i) and (ii) are alternative forms]; (c) *M. simiae*. The Box A₅ sequence is framed.

Fig. 2 are provisional, and could be in error by four nucleotides, depending on the specificity of the RNase involved in the maturation process (see Fig. 3).

Spacer-2 region

The spacer-2 region of the species studied ranges in length from 77 to 92 nucleotides (see Table 2). The 5'-end is marked by the conserved sequence (nucleotides 5142–5146) 5'-ACACA-3'. The following ten nucleotides (5149–5158) are highly conserved and are complementary to nucleotides 1956–1965 of the spacer-1 region, as shown in Fig. 4. Nucleotides 5164–5196 include a stem-loop structure, helix S2-1 (see Fig. 4), which is either very close to, or interrupts, sequences complementary to the Box A element of the spacer-1 region (nucleotides 1936–1947). The location of helix S2-1 is unusual since this feature is absent from other species such as *B. subtilis*, *Streptomyces*

and *Mycoplasma* for which data are available. Thus, sequences complementary to the highly conserved sequence (nucleotides 1921–1965) of spacer-1b are found in the region of nucleotides 5149–5216. Potential RNase E cleavage sites are found in the region nucleotides 5127–5242, as discussed below. The 5'-end of the 5S rRNA gene is preceded by the sequence 5'-ATAGA-3' (nucleotides 5242–5246).

5S rRNA gene sequences

The 5S rRNA genes of both *M. tuberculosis* and *M. leprae* (positions 5247–5363 inclusive) comprise 116 nucleotides, and the sequences differ in eight positions. *M. bovis* strain 35737 (Ohkubo *et al.*, 1986) and *M. bovis* BCG each comprise 117 nucleotides and the sequences differ in two and four positions respectively from the sequence inferred from the 5S rRNA gene of *M. tuberculosis*.

(a) *M. tuberculosis* (b) *M. leprae*

Fig. 5. Secondary structure of the trailer region of pre-rRNA based on possible steps in the termination of transcription. (a) *M. tuberculosis*; (b) *M. leprae*. (A) Possible structure of the DNA-RNA complex at the terminator site, before the release of pre-rRNA. The DNA sequence extends from the 3'-end of the 5S rRNA gene to a putative intrinsic termination site. During transcription approximately 12 base-pairs stabilize the interaction between the nascent RNA chain and the template strand of DNA. At the terminator this interaction is diminished because of the formation of a stem-loop structure near the 3'-end of the RNA transcript. The remaining dA.rU base-pairs are insufficient to maintain the complex, so that the RNA is released (see text). (B) Possible secondary structure of the trailer region of pre-rRNA. Further base-pair interactions may occur after the release of the completed pre-rRNA chain. RNA sequences are shown in bold italics.

The trailer region of the *rrn* operon

The trailer region extends from the 3'-end of the 5S rRNA gene to the 3'-end of the transcribed region of the operon, which is marked by the location of a putative intrinsic transcription terminator (for review see Richardson, 1993). The mechanism proposed for intrinsic termination is based on the assumption that the 3'-end of the nascent RNA is kept attached to the replicating complex by approximately 12 base-pairs formed between the 3'-terminal residues of nascent RNA and the template strand of DNA. Two factors may combine to diminish the interaction between the nascent RNA chain and template DNA so that the RNA chain is released and transcription is terminated (Von Hippel & Yager, 1991). The first factor is that residues which would otherwise interact with the template strand of DNA become involved in a competing reaction to form an intramolecular stem-loop

structure. This is apparent from the DNA sequence as a region of interrupted dyad symmetry. The second factor is the relative weakness of the dA.rU base-pair, so that a run of A residues on the template strand signifies a region of comparatively weak binding between nascent RNA and template DNA. Intrinsic termination takes place when both factors reinforce each other. A putative intrinsic terminator is located within residues 12–33 downstream from the 3'-end of *M. tuberculosis* 5S rRNA, suggesting a trailer sequence of 33 nucleotides (see Table 2, Figs 2 and 5). A homologous intrinsic terminator of transcription, allowing for C→U transitions, is located within the 32 bp sequence flanking the 3'-end of the 5S rRNA gene of *M. leprae* (see Figs 2 and 5). This element may have reduced efficiency as a terminator because the two G→U wobble base-pairs lower the stability of the competing stem-loop structure and so diminish its ability to destabilize the transcription complex.

Sequences flanking the 3'-end of the *rrn* operon

The sequence which was established for the 58 nucleotides downstream from the end of transcription of the *rrn* operon of *M. tuberculosis* (see Fig. 2) has no obvious similarities with the corresponding region of *M. leprae*.

Secondary structure of pre-rRNA

A scheme for the secondary structure of pre-rRNA extending from the start of transcription to the 3'-end of spacer-1a was described previously (Ji *et al.*, 1994). Sequences for pre-rRNA were inferred from the gene sequence data (Fig. 2) extending from the 5'-end of spacer-1b [nucleotide 1816 (S130)] to the end of transcription [nucleotide 5435 (T32)]. The inferred sequences were used to devise a scheme of pre-rRNA secondary structure which may apply generally to slow-growing mycobacteria. This scheme is illustrated in Fig. 6(a) for *M. tuberculosis* pre-rRNA and in Fig. 6(b) for *M. leprae* pre-rRNA. Both *M. tuberculosis* and *M. leprae* 5S-rRNAs fit the minimal model (Walters & Erdmann, 1988).

The principal features of the scheme are described below. The sequences from the potential Box B_s' and Box B_s'' elements each form a stem-loop structure. The long stem region comprises approximately 40 base-pairs which are formed between the highly conserved 40 nucleotide stretch of spacer-1b and two conserved regions of spacer-2. One of these two regions has a run of six consecutive C residues which forms base-pairs with a G-rich section of spacer-1b. Helix S2-1, formed between segments of spacer-2, is located towards the middle of the main stem. Following the main stem, spacer-1b sequences form helix S1-5, which is stabilized by 20 base-pairs; this is followed by the short A-U-rich helix S1-6.

The 5'-end of 23S rRNA is marked by the bulge S1-1 and the 3'-end is marked by a sequence of alternating A and C residues. These two regions contain complementary sequences with the potential to interact in the tertiary structure of pre-rRNA. The spacer-2 sequences form two stem-loop structures: one, helix S2-1, is close to the centre of the main stem and the other, helix S2-2, is

Fig. 6. A scheme for the secondary structure of mycobacterial pre-rRNA downstream from the 3'-end of spacer-1a. (a) *M. tuberculosis*; (b) *M. leprae*. The scheme is supported by sequence homologies (see Fig. 2). The sequences correspond to position 1880–5376 (Fig. 2). 51–130 etc. and 52–10 etc. refer to positions within spacer-1 and spacer-2 respectively (see Fig. 2). Stem-loop structures are numbered consecutively starting from the 5'-end; T indicates trailer sequences. The minimal model for 5S rRNA secondary structure (Walters & Erdmann, 1988) was used.

Table 3. Base-pair formation in pre-rRNA secondary structure (see Fig. 6)

Region	Species	No. of residues	No. of base-pairs				G-C % of base-pairs	% of residues forming base-pairs	Mismatched bases	
			A-U	G-U	G-C	Total			Bulge of 1N	Bulge of 2N
5S rRNA	<i>M. tuberculosis</i>	116	4	3	24	31	77	53	2	0
	<i>M. leprae</i>	116	5	6	20	31	65	53	2	0
Pre-rRNA excluding 5S rRNA	<i>M. tuberculosis</i>	238	30	10	46	86	54	72	9	1
	<i>M. leprae</i>	236	41	19	31	91	34	77	6	0
Stem excluding helix S2-1	<i>M. tuberculosis</i>	86	12	4	23	39	59	92	6	1
	<i>M. leprae</i>	90	17	7	18	42	43	94	6	0

located between the main stem and 5S rRNA. In *M. tuberculosis* pre-rRNA (Fig. 6a) helix S2-2, which is stabilized by two A-U base-pairs and 1 G-U base-pair, has marginal stability. The secondary structure of the trailer region was derived from an analysis of the putative termination process.

Details of the secondary structure for pre-rRNA of *M. leprae* and *M. tuberculosis* are compared in Table 3. The notable difference lies in the reduced number of G-C base-pairs and a corresponding increase in the wobble G-U base-pair in *M. leprae* compared with *M. tuberculosis*. It is known that G-C base-pairs are more stable than G-U

base-pairs, which neither stabilize nor destabilize a helix (see for example Ninio, 1979). Thus the secondary structure of *M. leprae* pre-rRNA is likely to be appreciably less stable than that of *M. tuberculosis* pre-rRNA. It is unclear whether the processing of *M. leprae* pre-rRNA is affected by the relatively low stability of its secondary structure.

DISCUSSION

Previously, we identified promoter elements, the site for the start of transcription, and antitermination elements in the leader region of the *rrn* operon of slow-growing mycobacteria (Kempsell *et al.*, 1992; Ji *et al.*, 1994). The results presented above provide information about spacer-1b, the 5'- and 3'-ends of the 23S rRNA gene, spacer-2, the 5S rRNA gene and the termination of transcription. The transcribed region of the operon comprises approximately 5.4 kbp (Fig. 2). The organization of the three rRNA genes into one operon is one way of ensuring that mature 16S rRNA, 23S rRNA and 5S rRNA are synthesized in equimolar amounts. This strategy requires that the transcription process is highly efficient; that is, a complete copy of the operon should be produced each time transcription is initiated, and the transcript should be efficiently processed.

The efficient transcription of the *rrn* operon requires that once the DNA-dependent RNA polymerase complex is formed at the initiation site it moves along, and remains attached to the template strand of DNA until the termination signal is reached. Antitermination elements assist in this process. As described above, the mycobacterial *rrn* operon has two sets of antitermination elements. One set is found in the leader region and the other is located in the spacer-1b region of the operon. This arrangement of antitermination elements is widespread among both Gram-positive and Gram-negative bacteria (Berg *et al.*, 1989). One example, the *rrnG* operon of *E. coli*, has been investigated more intensively than others (Berg *et al.*, 1989; Nodwell & Greenblatt, 1993) and the pattern reported for this operon is a motif which appears to have been conserved in evolution.

Box A_L element of the *rrn* operon

Comparatively little is known about the Box B and Box C elements of the *rrn* operon (Berg *et al.*, 1989; for review see Roberts, 1993) whereas the function of Box A is clearly defined. The Box A element of the leader region (Box A_L) of the *rrnG* operon of *E. coli* has been shown to modify the RNA polymerase complex (Albrechtsen *et al.*, 1990) by preventing rho-dependent termination. The Box A_L element can function independently of Box B_L and Box C_L. The Box A_L RNA sequence of 12 nucleotides binds a complex of two proteins, namely the antitermination factor NusB and ribosomal protein S10 (Nodwell & Greenblatt, 1993). With one exception, mutations in the Box A element considerably reduce or eliminate antitermination activity (Nodwell & Greenblatt, 1993).

The sequence required for function is 5'-TGCTCTTTAA(C/A)A-3'. One form Box A_L (5'-TGCTCTTTAACA-5') is present in the leader region, and the other form Box A_S (5'-TGCTCTTTAAAA-3') is located in the spacer-1b region of the *rrnG* operon. Although the mycobacterial Box A_L sequence is unlikely to bind the *E. coli* proteins NusB and S10 it is likely to interact with the homologous mycobacterial factors. We infer that, as in *E. coli*, a ribosomal protein, for example S10, takes part in the regulation of rRNA biosynthesis, through interaction with the Box A_L antitermination element.

The Box A_S and Box A_L elements of the *rrnG* operon differ in the penultimate nucleotide position: an A residue replaces a C residue. As mentioned above, this change does not diminish the capacity of Box A_S RNA to bind the proteins NusB and S10. Hence, protein S10 influences the transcription of the *rrnG* operon at two sites, one of which is located in the leader and the other in the spacer-1b region of the operon.

The mycobacterial Box A_S element has the sequence 5'-TGGTGTTTGAGT-3' in five mycobacteria and the sequence 5'-TGGTGTTTGAGA-3' in *M. tuberculosis*. The significance of this change in one nucleotide position has yet to be established. It is possible that both sequences recognize the same combination of protein factors.

The mycobacterial Box A_S sequence differs from the Box A_L sequence in position 3: a G residue replaces a T residue. One or two differences in the nucleotide sequence of Box A_L and Box A_S appear to be the rule (Berg *et al.*, 1989). On the basis of the assumption that the *rrnG* operon of *E. coli* is representative of bacterial *rrn* operons in general, it is likely that antitermination function of the mycobacterial Box A_S element is regulated by a ribosomal protein (probably S10), which may also interact with Box A_L. It is clear that the *rrn* operons of the slow-growing mycobacteria which we have studied, including *M. leprae*, share the same mechanisms for antitermination.

The involvement of the ribosomal protein S10 in the control of the expression of the *rrn* operon is significant because it links rRNA biosynthesis with ribosomal protein synthesis. The gene for S10 is one of the genes of the S10 operon of *E. coli* which encodes 11 ribosomal (both S- and L-) proteins. This operon is autogenously controlled: protein L4, the product of the third gene of the operon, represses the expression of all eleven genes when its accumulation exceeds the available binding sites on 23S rRNA (Lindahl & Zengel, 1979; Zengel *et al.*, 1980). The details of the mechanism operating in mycobacteria need to be elucidated.

Processing pre-rRNA

Mature rRNA species are produced from pre-rRNA by the action of a number of different RNases (for review see Apirion & Miczak, 1993). Early cleavages produce precursors of the mature species: namely, pre-16S rRNA pre-23S rRNA and pre-5S rRNA (see Fig. 1). It is not feasible to study the maturation process directly because

the ratio pre-rRNA/rRNA in the RNA fraction is unfavourable owing to the slow growth of the bacteria and the rapid processing of pre-rRNA (Ji *et al.*, 1994). However, some insight into processing may be gained from the scheme for secondary structure (Fig. 5) which completes the putative secondary structure of pre-rRNA based on phylogenetic considerations.

Potential RNase III recognition sites were identified by their homology with known sites. These sites are located in the stem formed by the interaction between complementary sequences of the spacer-1b and spacer-2 regions (Fig. 5). Cleavage of *M. tuberculosis* pre-rRNA at this site would generate pre-23S rRNA comprising mature 23S rRNA extended by 84 nucleotides upstream from the 5'-end and by 22 nucleotides downstream from the 3'-end. Little is known about further RNase reactions which generate mature 23S rRNA. The proposed secondary structure suggests that bulge-S1, and the sequence 5'-ACACACA-3', are implicated in the generation of the 5'- and 3'-ends, respectively, of mature 23S rRNA.

RNase E is implicated in the production of pre-5S rRNA and its recognition site has the consensus sequence 5'-(A/G)AUU(A/U)-3' (Ehresmann *et al.*, 1992). One potential RNase E site is located in spacer-2 downstream from the main stem and upstream from the 5'-end of mature 5S rRNA (Figs 2 and 6). Cleavage at this site would generate *M. tuberculosis* pre-5S rRNA comprising mature 5S rRNA extended by 11 nucleotide residues upstream from the 5'-end and 33 nucleotide residues downstream from the 3'-end of mature 5S rRNA to the 3'-end of the pre-rRNA transcript. The *rrnO* operon of *B. subtilis* appears to be cleaved in this way (Ogasawara *et al.*, 1983), generating pre-5S rRNA which has homologies with the putative product of the mycobacterial *rrn* operon. *M. leprae* has two potential RNase E cleavage sites within close proximity (Figs 2 and 6b).

Comparison of mycobacteria with other species

Pre-rRNAs have several features in common but they are also characterized by much diversity, as may be seen from the size of spacer-1 and its organization into spacer-1a and spacer-1b sections (see Table 2). For example, the mycobacterial *rrn* operon has many features in common with the *rrnO* operon of *B. subtilis* (Ji *et al.*, 1994). However, the latter has two tRNA genes located between spacer-1a and spacer-1b (Table 2). As shown previously for pre-16S rRNA (Ji *et al.*, 1994; and see Figs 2 and 6 above) the mycobacterial *rrn* operon has a number of distinctive features that are apparent in pre-rRNA and which are helpful in identifying members of the family of slow-growing mycobacteria, and in distinguishing one mycobacterial species from another. For example, the leader and spacer-1a sequences suggest that '*M. lufu*' is a close relative of *M. avium* and *M. intracellulare* (Ji *et al.*, 1994) and the data for spacer-1b and spacer-2 (see Fig. 2) provide further evidence for this relationship.

Finally, the progress made in identifying the principal features of the *rrn* operon and its transcript pre-rRNA of *M. tuberculosis*, *M. leprae* and other slow-growers provides

a basis for comparison with the *rrn* operons and the pre-rRNA transcript of fast-growers. A comparison of this kind may reveal features of the *rrn* operon which are associated with the slow growth of the pathogens.

ACKNOWLEDGEMENTS

This investigation received support from the UWDP/WHO special programme for Research Training in Tropical Diseases and from the British Leprosy Relief Association (LEPRA). Dr Yuan-en Ji was supported by a grant from LEPRA.

REFERENCES

- Albrechtsen, B., Squires, C. L. & Squires, C. (1990). Antitermination of characterized transcriptional terminators by the *Escherichia coli* *rrnG* leader region. *J Mol Biol* **213**, 123-134.
- Apirion, D. & Miczak, A. (1993). RNA Processing in prokaryotic cells. *BioEssays* **15**, 113-120.
- Baylis, H. A. & Bibb, M. J. (1988). Transcriptional analysis of the 16S rRNA gene of the *rrnD* gene set of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **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 microorganisms. *J Mol Biol* **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. *Biochem Biophys Res Commun* **136**, 1136-1141.
- Bruckner, D. A. & Colonna, P. (1993). Nomenclature for aerobic and facultative bacteria. *Clin Infect Dis* **16**, 598-605.
- Dams, E., Yamada, T., De Baene, R., Huysmans, E., Vandenberghe, A. & De Wachter, R. (1987). Structure of 5S rRNA in actinomycetes and relatives and evolution of eubacteria. *J Mol Evol* **25**, 255-260.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387-395.
- Dubos, R. J. & Davis, B. D. (1946). Factors affecting the growth of tubercle bacilli in liquid media. *J Exp Med* **83**, 409-423.
- Ehresmann, C. P., Carpousis, A. J. & Krisch, H. M. (1992). Specificity of *Escherichia coli* endoribonuclease RNaseE: *in vivo* and *in vitro* analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes & Dev* **6**, 149-159.
- Fotheringham, R. & Wilson, K. (1993). Sequence-based differentiation of strains in the *Mycobacterium avium* complex. *J Bacteriol* **175**, 2818-2825.
- Ghora, B. K. & Apirion, D. (1978). Structural analysis and *in vitro* processing to P5 rRNA of a 9S RNA molecule isolated from an *rrn* mutant of *E. coli*. *Cell* **15**, 1055-1066.
- Gutell, R. R., Schnare, M. N. & Gray, M. W. (1990). A compilation of large subunit (23S-like) ribosomal RNA sequences presented in a secondary structure format. *Nucleic Acids Res* **18**, Sequences Supplement, 2319-2331.
- Iwami, M., Muto, A., Yamao, F. & Osawa, S. (1984). Nucleotide sequence of the *rrnB* 16S ribosomal RNA gene from *Mycoplasma capricolum*. *Mol & Gen Genet* **196**, 317-322.
- Ji, Y.-E., Colton, M. J., Cox, R. A. (1994). Nucleotide sequence and secondary structures of precursor 16S rRNA of slow-growing mycobacteria. *Microbiology* **140**, 123-132.
- Kempell, K. E., Ji, Y.-E., Estrada-G, I. C. E., Colton, M. J. & Cox, R. A. (1992). 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. *J Gen Microbiol* **138**, 1717–1727.

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 Lett* **281**, 114–118.

Lindahl, L. & Zengel, J. M. (1979). Operon-specific regulation of ribosomal protein synthesis in *Escherichia coli*. *Proc Natl Acad Sci USA* **76**, 6542–6546.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Meissner, G. & Schroder, K.-H. (1975). Relationship between *Mycobacterium simiae* and *Mycobacterium habana*. *Am Rev Respir Dis* **111**, 196–200.

Ninio, J. (1979). Prediction of pairing schemes in RNA molecules: loop contributions and energy of wobble and non-wobble base-pairs. *Biochemistry* **61**, 1133–1150.

Nodwell, J. R. & Greenblatt, J. (1993). Recognition of Box A antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell* **72**, 261–268.

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 Res* **11**, 6301–6318.

Ohkubo, S., Lwasaki, H., Hori, H. & Osowa, S. (1986). Evolutionary relationship of denitrifying bacteria as deduced from 5S rRNA sequences. *J Biochem* **100**, 1261–1267.

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* **79**, 33–46.

Portaels, F. (1980). Unclassified mycobacterial strains susceptible to dapsone isolated from the environment in central Africa. *Int J Lepr* **48**, 330.

Richardson, J. P. (1993). Transcription termination. *Crit Rev Biochem Mol Biol* **28**, 1–30.

Roberts, J. W. (1993). RNA and protein elements of *E. coli* and λ transcription antitermination complexes. *Cell* **72**, 653–655.

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. *J Exp Med* **112**, 445–454.

Singh, N. B., Lowe, A. C. R. E., Rees, R. J. W. & Colston, M. J. (1989). Vaccination of mice against *Mycobacterium leprae* infection. *Infect Immun* **57**, 653–655.

Suzuki, Y., Nagata, A., Ono, Y. & Yamada, I. (1988). Complete nucleotide sequence of the 16S rRNA gene of *Mycobacterium bovis* BCG. *J Bacteriol* **170**, 1631–1636.

Taschke, C. & Herrmann, R. (1986). Analysis of transcription and processing signals of the 16S–23S rRNA operon of *Mycoplasma hyopneumoniae*. *Mol & Gen Genet* **205**, 434–441.

Valdivia Alvarez, F., Suarez Mendez, R. & Echemendia Font, M. (1971). *Mycobacterium habana*: probable especie dentro de las micobacterias no clasificadas. *Bol Hig Epidemiol* **9**, 65–73.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci USA* **76**, 615–619.

Von Hippel, P. H. & Yager, T. D. (1991). Transcript elongation and termination are competitive kinetic processes. *Proc Natl Acad Sci USA* **88**, 2307–2311.

Walters, J. & Erdmann, V. A. (1988). Compilation of 5S rRNA and 5S rRNA gene sequences. *Nucleic Acids Res* **18**, Supplement, 2215–2236.

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. *J Gen Microbiol* **63**, 29–39.

Zengel, J. M., Muecki, D. & Lindahl, L. (1980). Protein L4 of the *E. coli* ribosome regulates an eleven gene rprotein operon. *Cell* **21**, 523–535.

Received 1 November 1993; revised 12 January 1994; accepted 18 January 1994.