The rRNA operons of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: comparison of promoter elements and of neighbouring upstream genes

Jorge A. Gonzalez-y-Merchand,¹ ¹² M. Joseph Colston¹ and Robert A. Cox¹

Author for correspondence: Robert A. Cox. Tel: + 44 181 959 3666. Fax: + 44 181 906 4477.

*Mycobacterium smegmatis* has two rRNA (rrn) operons designated rrnA, and rrnB. Appropriate restriction fragments of genomic DNA containing sequences immediately upstream from the 16S rRNA genes were cloned. We now report the nucleotide sequence of 552 bp upstream from the 5'-end of the Box A, antitermination element of the leader region of the rrnA, operon. The 5'-end of this segment of DNA was found to comprise 113 codons of an ORF encoding a protein which is significantly similar to UDP-N-acetylglucosamine 1-carboxyvinyl-transferase (EC 2.5.1.7), which is important to cell wall synthesis. A homologous ORF is located immediately upstream from the single rrn (rrnA,) operons of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. Primer-extension analysis of the RNA fraction of *M. smegmatis* revealed four products which were related to transcription start points; the rrnB, operon appears to have a single promoter whereas the rrnA, operon has three (P1, P2 and P3). Analysis of *M. tuberculosis* RNA revealed two products corresponding to transcripts directed by promoters homologous with P1 and P3 of the rrnA, of *M. smegmatis*. Thus, the promoter and upstream regions of the rrnA, operon of *M. smegmatis* and the rrnA, operon of *M. tuberculosis* are homologous. The presence of P2 in *M. smegmatis* and its absence from *M. tuberculosis* is attributable to insertions/deletions of 97 bp.

Keywords: pathogenic/non-pathogenic mycobacteria, rRNA operons, promoter activities, cell wall synthesis, UDP-N-acetylglucosamine 1-carboxyvinyl-transferase

INTRODUCTION

The two major mycobacterial pathogens, *Mycobacterium tuberculosis* and *Mycobacterium leprae* are characterized by the ability to grow slowly within host cells. The mechanisms which these organisms use to control their growth rate are not well understood. There is a clear relationship between rates of growth, protein biosynthesis and the rate of synthesis of ribosomes (Bremer & Dennis, 1987; Winder & Rooney, 1970). An understanding of the way in which ribosome synthesis is regulated in the slow-growing mycobacteria would further our understanding of cell proliferation. In addition, by comparing the regulation of ribosome synthesis in the pathogenic slow-growing species with that in fast-growing mycobacteria, most of which are non-pathogenic, we are likely to gain new insights into the evolution of growth-rate control in these two phylogenetically distinct groups.

*M. leprae* and *M. tuberculosis* each have a single rRNA (rrn) operon (Liesack et al., 1990; Sela & Clark-Curtiss, 1991; Kempsell et al., 1992; Suzuki et al., 1988). The single rrn operons of the slow-growers studied are sufficiently closely related in both their primary and secondary structures to be regarded as members of a family, described as rrnAₘ operons (the subscript 'ₘ' denotes slow-growers) by Ji et al. (1994c). In contrast to slow-growers, fast-growing mycobacteria have two rrn operons (Bercovier et al., 1986). The two rrn operons of a representative fast-grower, *Mycobacterium smegmatis*, were investigated previously (Ji et al., 1994c). Based on sequence analysis, one of the operons was found to be

---

Abbreviation: UNAcGCT, UDP-N-acetylglucosamine 1-carboxyvinyl-transferase.

The EMBL accession numbers for the sequences reported in this paper are X87943 and X87944.
closely related to the \( rrrA_f \) operon of slow-growers and was designated \( rrrA_f \) (the subscript 'f' denotes fast-growers). The other \( M. \) smegmatis operon was designated \( rrrB \) (Ji et al., 1994c); this operon appeared to be more distantly related to the \( rrrA_f \) operon of slow-growers.

We have now cloned and sequenced the regions of the \( rrrA_f \) and \( rrrB \) operons of \( M. \) smegmatis upstream from the putative antitermination element, Box \( A_f \), which is highly conserved between all of the mycobacterial \( rrr \) operons (Ji et al., 1994c). We find that the \( rrrA_f \) and the \( rrrA_s \) operons have similar sequences in this region, confirming our earlier hypothesis concerning the relatedness of these two operons. In addition, we have used RNA extracted from both \( M. \) smegmatis and \( M. \) tuberculosis to compare the promoter regions of the two species using primer extension. The RNA fraction of \( M. \) smegmatis revealed at least four transcription start points; the \( rrrB \) operon appears to have a single promoter, while the \( rrrA_f \) operon has at least three. The single \( M. \) tuberculosis \( rrrA_s \) operon appears to have two promoter elements. Thus the slow-growing mycobacteria not only have a reduced number of \( rrr \) operons, but also have a reduced number of regulatory elements involved in their transcription.

Intriguingly, we also found that both \( rrrA_f \) and \( rrrA_s \) have a promoter, \( P_1 \), which is located within a gene, which terminates 427 bp upstream of the 16s rRNA coding region. This gene encodes a protein that is significantly similar to an enzyme involved in bacterial cell wall synthesis suggesting that the production of the enzyme and transcription of the rRNA are linked.

**METHODS**

**Materials.** Restriction endonucleases and T4 ligase were from Boehringer Mannheim. Sequenase (USB) sequencing kit was supplied by Cambridge Biosciences. \(^{[35S]}\)dATP was from Amersham. GeneClean kit was from Biorad. The original TA cloning kit for PCR cloning was purchased from Invitrogen. The avian myeloblastosis virus (AMV) reverse transcriptase

---

**Fig. 1.** Scope of this investigation. (a) Part of the \( rrrA_f \) operon of \( M. \) smegmatis. The nucleotide sequence downstream from the Box \( A_f \) antitermination element was reported previously (Ji et al., 1994c). The nucleotide sequence of the V2 region of the 16s rRNA (for definition see Kempsell et al., 1992) serves as a signature sequence that is useful for identifying the species. This report concerns the nucleotide sequence of the region indicated by broken lines; that is, upstream from the Box \( A_f \) element to the Pstl site. Sequence data were obtained after cloning the BamHl/(downstream) Pstl fragment, and after amplifying and cloning the region between the upstream Pstl and position 357 of the 16s rRNA coding region. The primers used for amplification are shown by short horizontal arrows. The results reveal the identity of the promoters that correspond to transcription start sites \( ts(P_1) \), \( ts(P_2) \) and \( ts(P_3) \), and also the presence of an ORF in the position shown. (b) Sequencing strategy. Open circles indicate the binding sites of the primers used in the sequencing reaction, and the length of the arrows represent the region of sequence identified by means of the primer. (c) Part of the \( rrrA_f \) operon of \( M. \) tuberculosis. The data (see Kempsell et al., 1992) provide a frame of reference for (a). The segments of the operon shown in (a) and (c) are related by insertions/deletions of 97 bp in the region indicated by the open triangle (see text). \( ts(P_1) \) and \( ts(P_3) \) indicate the transcription start sites identified by Verma et al. (1994) and by primer-extension studies as described below. In both (a) and (c) the ORF corresponds to part of the gene encoding a protein significantly similar to UNAcGCT.
Fig. 2. Nucleotide sequences of the regions of the M. smegmatis (M.sm) genome upstream from the Box A elements of the rrnA and rrnB operons and comparison with the rna operon of M. tuberculosis (M.tu) [The sequence in Kemplell et al. (1992) was re-determined and several errors were identified. The principal correction is the deletion of positions 73-83 from their sequence. The corrected version is given above]. The deduced amino acid sequences encoded by the putative genes for UNAcGCT (see Fig. 1) are given using the one-letter code. A single letter is given where the same amino acid is encoded by both M.sm and M.tu, otherwise the amino acid for M.sm is given first; an italic letter indicates that the identity of the amino acid in the M. smegmatis gene is provisional because it is inferred from the sequence of the PCR primer. The PstI and BamHI sites are underlined and the TAG codons at the 3'-ends of coding regions are doubly underlined. Putative promoter elements (−35 and −10 boxes) are boxed; uncertainty in the identity of −35 boxes is indicated by broken lines. M.sm(A), etc., rna operon of M. smegmatis, etc.; G*, transcription start points inferred from primer-extension products (see Fig. 4). tA(P1), etc., transcription start site directed by the P1 promoter of the rna operon, etc. The sequence of the rnb operon of M. smegmatis was also established (see also Fig. 4) and is in agreement with the data of Predich et al. (1995). Deletion; N, nucleotide of unknown identity; *, amino acid of unknown identity; dotted lines indicate where the oligonucleotide primer, used for primer-extension studies, annealed.

primer-extension system was supplied by Promega. Oligonucleotides primers were prepared with an automated DNA synthesizer (model 370A, Applied Biosystems).

Bacterial strains, media and vector. M. smegmatis NCTC 8159 (National Collection of Type Cultures) was maintained on Löwenstein-Jensen slopes and grown to exponential phase in Lemco Broth containing 0.1% Tween 80 for use in primer-extension studies. M. tuberculosis H37 Rv was grown in Middlebrook medium. Epicurian Coli XL1-Blue competent cells (Stratagene) were used for transformation with the recombinant phagemid Bluescript SK (pSK) (Stratagene) and grown on LB medium. M. tuberculosis DNA was given by Dr E. Davis, National Institute for Medical Research, Mill Hill, London, UK.

Isolation of DNA. Plasmid DNA was isolated by standard methods exactly as described previously by Ji et al. (1994c). Genomic DNA from M. smegmatis was isolated by a novel method (Gonzalez-y-Merchand et al., 1996). Briefly, the cells were suspended in 6 M guanidinium chloride and then kept at 65°C for 10 min. The DNA was then precipitated and resuspended in TE buffer by conventional means.

Isolation of RNA. Exponential-phase cells (100 ml culture) were collected and resuspended in 1 ml guanidinium buffer (6 M guanidinium chloride, 0.1%, v/v, Tween 80, 10 mM EDTA, 1 mM 2-mercaptoethanol) and left at −20°C for 15 min. The suspension was added to a half the volume of heat-sterilized 0.15 M glass beads contained in a 2 ml screw-cap micro-centrifuge tube. Mycobacteria were ruptured by 3 pulses each of 1 min, on the Mini-BeadBeater device (Biospec Products). Debris and beads were sedimented by centrifugation (10000 g for 3 min), and the cleared lysate was retained. The pellet of beads and mycobacterial residues was briefly re-extracted on the BeadBeater (30 s pulse) with 300 μl fresh guanidinium buffer.
and the resulting extract was pooled with the first. The lysate was extracted three times with 2 vols chloroform/3-methyl-1-butanol (24:1, v/v). RNA was precipitated by the dropwise addition of 0.5 vol. ethanol and then redissolved in the appropriate volume of MOPS buffer.

**Primer extension.** The oligonucleotide primer 5'CACACTAT-TGAGTTCTC3' has a target site which is present in all three of the *rrn* operons studied (see Fig. 2). This primer was end-labelled with [32P]ATP by means of T4 polynucleotide kinase, and the primer extension was carried out using the AMV reverse transcriptase primer-extension system. Briefly, the 32P-labelled primer (100 fmol) was added to a mixture of 5 ml water containing 24 μg total RNA and 5 μl AMV primer extension (PE) 2× buffer (100 mM Tris, 100 mM KCl, 20 mM MgCl2, 20 mM DTT, 2 mM each dNTP, 1 mM spermidine). The mixture was annealed at 52 °C for 30 min and left to cool in air at room temperature for 10 min; then 5 μl PE 2× buffer, 1.4 μl 40 mM sodium pyrophosphate, and 1 unit AMV reverse transcriptase were added to each reaction. The reaction mixture was incubated at 42 °C for 1 h, extracted once with phenol/chloroform, ethanol-precipitated, and washed with 70% (v/v) ethanol. The extension products were separated on a 8% (w/v) polyacrylamide/urea gel and visualized by autoradiography.

**Genomic DNA cloning.** The products of the digestion with PstI, or double digestion with BamHI and PstI restriction endonucleases, of genomic DNA were separated by 1% (w/v) agarose gel electrophoresis. A 3.2 kbp PstI or 1.2 kbp BamHI/PstI fragment was recovered from the gel using glass milk (Vogelstein & Gillespie, 1979). The BamHI/PstI fraction was ligated into pSK (which was previously digested with PstI and BamHI). The PstI fraction was ligated into pSK previously digested with PstI. In each case the molar ratio of vector to insert was 1:1.

**PCR.** Bacterial DNA (1–100 ng) was subjected to PCR (Saiki et al., 1988) as described previously (Ji et al., 1994c). The relevant gene fragment encoding 114 amino acids of UDP-N-acetylgalactosamine-1-carboxyvinyl-transferase (UNAcGCT) (EC 2.5.1.7), the promoter region of the *rrn* operons, the 5'-ends of the leader regions of their precursor-RNA (pre-rRNA) and the 5'-region of the 16S ribosomal DNA was synthesized using the primer combination of JG7 (5'CACTGGTGCCTCCCGTAGG3') and promoter regions was achieved by means of the UK Human Genome Mapping Program (Altschul et al., 1990). The relevant databases (see Methods) revealed that the amino acid sequence inferred

**RESULTS AND DISCUSSION**

The strategy described in Fig. 1 were used to establish the nucleotide sequence upstream from the Box A1 motif of the *rrnA1* operon. As shown in Fig. 2, 342 bp of an ORF were identified. A search of the databases (see Methods) revealed that the nucleotide sequence upstream from the Box A1 motif of the *rrnA1* operon.

**Nucleotide sequences of the upstream regions**

The strategies described in Fig. 1 were used to establish the nucleotide sequence upstream from the Box A1 motif of the *rrnA1* operon. As shown in Fig. 2, 342 bp of an ORF were identified. A search of the databases (see Methods) revealed that the nucleotide sequence inferred

**Alignment of sequences.** Computer-aided analysis of the alignment of nucleotide sequences of the gene encoding UNAcGCT and promoter regions was achieved by means of the BLAST program (Altschul et al., 1990). The database searches were carried out using the UK Human Genome Mapping Project computer services (Rysavy et al., 1992).
from this gene sequence is significantly similar to 113 residues of the carboxyterminal region of UNAcGCT; the BLAST score was $P = 20 \times e^{-23}$ (the lower the score the more certain is the assignment). This enzyme comprises a single polypeptide of approximately 420 amino acids (see for example Ehrt & Hillen, 1994). As shown in Fig. 2 the gene lying upstream from the $rrnA_e$ operon of $M. tuberculosis$ is homologous with the gene which is located upstream from the $rrnA_e$ operon of $M. smegmatis$. Examination of $M. leprae$ sequences (Sela & Clark-Curtiss, 1991) reveals that a gene encoding UNAcGCT is also located upstream from the $rrnA_e$ operon in that species. In contrast, the gene lying upstream from the $rrnB_e$ operon of $M. smegmatis$ is believed to encode tyrosyl-tRNA synthetase (Predich et al., 1995).

These results confirm our original hypothesis (Ji et al., 1994c) that the $rrnA_e$ and $rrnA_e$ operons are homologous. The major difference between these two operons is a deletion of 97 bp in the $M. tuberculosis$ operon (75 bp in that of $M. lepra$) compared to that of $M. smegmatis$, as indicated in Fig. 2.

No intrinsic terminators are evident downstream from the UAG stop codons of the ORFs (Fig. 2) indicating that termination of transcription might be rho-dependent. Possible stem–loop structures that could be involved in termination are presented in Fig. 3. In each case a GC-rich stem–loop structure can be formed within 14–21 nucleotides from the UAG codon, suggesting that the trailer regions of the transcripts comprise a minimum of 14–21 nucleotides.

**Primer-extension studies**

The RNA fraction of $M. smegmatis$ was analysed as described in Methods and the products were identified using appropriate sequencing ladders for the $rrnA_e$ and $rrnB_e$ operons. Six major products were identified (Fig. 4). Four of the products had identifiable promoter consensus sequences immediately upstream. One of these, giving a product of 160 nucleotides, is on the $rrnB_e$ operon. Predich et al. (1995) also concluded that the $rrnB_e$ operon has a single promoter, on the basis of in vitro transcription.
studies. Three other putative promoters, giving products of 230, 136 and 60 nucleotides (designated P1, P2 and P3, respectively) are on the \( \text{rrnA}_t \) operon. Two other products, a and the more prominent b (125 and 98 nucleotides, respectively), have no promoter consensus of 230, 136 and 60 nucleotides (designated P1, P2 and P3, respectively) are on the \( \text{rrnA}_t \) operon, etc.; subscript numbers denote the nucleotide positions given in Fig. 2. Asterisks denote transcription start sites identified by primer extension (Fig. 4). The TAG codons at the 3'-ends of coding regions are underlined. Putative promoter elements (−35 and −10 boxes) are framed; uncertainty in the identity of −35 boxes is shown by broken lines; a, identified as a −35 box in promoters of \( \text{rrnB}_t \), \( \text{rrnD} \) and \( \text{rrnA}_s \) have four promoters which lie within other genes (Strohl, 1992).

The intervals separating the start sites (ts) of the promoters of the \( \text{rrnA}_t \) operon of \( M. \) \textit{smegmatis} are 94 bp between ts(P1) and ts(P2) and 76 bp between ts(P2) and ts(P3); the start sites ts(P1) and ts(P3) directed by promoters P1 and P3 of the \( \text{rrnA}_t \) operon of \( M. \) \textit{tuberculosis} are separated by 77 bp. These intervals are comparable to distances separating the two promoters of \( \text{rrnA}_t \) operons of \textit{Bacillus subtilis} [approx. 90 bp (Ogasawara et al., 1983)] and with the intervals between the four promoters of the \( \text{rrnA}_t \) operon of \textit{Streptomyces coelicolor} A3(2) (van Wezel et al., 1994); namely 180 bp separating P1 and P2, 91 bp separating P2 and P3, and 80 bp separating P3 and P4.

The differential usage of the three promoters of the \( \text{rrnA}_t \) operon of \( M. \) \textit{smegmatis} and the two promoters of the \( \text{rrnA}_s \) operon of \( M. \) \textit{tuberculosis} merits further study. Under the growth conditions used in this study the P2 promoter of the \( \text{rrnA}_t \) operon was more active than either the P1 or the P3 promoters. In contrast, transcription of the \( \text{rrnA}_s \) of \( M. \) \textit{tuberculosis} was governed mainly by the P3 promoter, the P2 promoter being absent from this species, with the P1 promoter playing the minor role. The presence of two or more promoters per operon is a strategy which offers the possibility of increasing the versatility of the operon, for example by enabling the operon to respond over a wide range of growth conditions to the cells’ need for rRNA. Another GC-rich Gram-positive bacterium, \textit{S. coelicolor} A3(2), has six \( \text{rrn} \) operons and at least two of them (\( \text{rrnA} \) and \( \text{rrnD} \)) have four promoters (Baylis & Bibb, 1988; van Wezel et al., 1994).
Each of the seven rrr operons of E. coli has two promoters (Gourse et al., 1986; Condon et al., 1992). The promoter elements identified by means of the primer-extension studies are compared in Fig. 5. Apart from the −35 boxes of the P1 promoters of the rrrA_p and rrrA_o operons the −35 and −10 motifs are very similar to those reported for promoters of rrr operons of B. subtilis (Ogasawara et al., 1983). The P1 promoters of rrrA_p and rrrA_o operons are very similar in sequence probably because they also form part of the coding region of a gene for UNAcGGT. A feature of the P2 and P3 promoters of the rrrA_p operon and the P3 promoter of the rrrA_o operon is that the −10 box is preceded by the sequence 5′ GTAT3′. There appear to be few features that distinguish elements used by the fast-grower from those used by the slow-grower. It should be noted however that a strongly expressed product was identified in M. smegmatis for which no consensus −35 or −10 boxes could be identified; whether this represents a novel promoter sequence or an RNA processing site cannot be distinguished at present.

These studies confirm and extend our previous observations on the relationship between the rrr operons of fast- and slow-growing mycobacteria. More importantly they demonstrate that the two groups have evolved differently not only in terms of the number of operons, but also in terms of their regulation. Experiments currently in progress are designed to investigate the environmental stimuli which affect the differential usage of the multiple promoters and hence should provide novel insights into mechanisms of growth control in these organisms.

ACKNOWLEDGEMENTS

J. A.G.-y-M. received financial support from the Mexican agencies CONACYT and COFAX & BDA (IPN). We thank Simon A. Cox for help in preparing the manuscript.

REFERENCES


Received 29 September 1995; accepted 2 November 1995.