

GENE 09144

Genomic organization of the mycobacterial sigma gene cluster

(*Mycobacterium tuberculosis*; *M. leprae*; *Corynebacterium*; σ factors; *dtxR*)

Laurence Doukhan^{a,b}, Mima Predich^c, Gopalan Nair^a, Olivier Dussurget^{a,b}, Ines Mandic-Mulec^a, Stewart T. Cole^d, Douglas R. Smith^e, Issar Smith^a

^aDepartment of Microbiology, The Public Health Research Institute, New York, NY 10016, USA. Tel. (1-212) 578-0868; ^bDépartement de Microbiologie, Université Paris VII Denis Diderot, 2 place Jussieu, 75251 Paris Cedex 05, France; ^cLaboratory of Cellular Physiology and Immunology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA. Tel. (1-212) 327-8107; ^dLaboratoire de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France. Tel. (33-1) 45-68-84-46; ^eCollaborative Research Inc., 1365 Main Street, Waltham, MA 01254, USA. Tel. (1-617) 893-5007

Received by R.E. Yasbin: 26 April 1995. Revised Accepted 19 May 22 May 1995; Received at publishers: 15 June 1995

SUMMARY

We have previously described σ^A and σ^B and their structural genes, *mysA* and *mysB*, respectively, in *Mycobacterium smegmatis*. We have now sequenced the corresponding regions in the *M. tuberculosis* and *M. leprae* chromosomes, and have found the two homologous genes. The chromosomal linkage and the deduced amino acid (aa) sequences of the two genes show very high similarity in the three species of mycobacteria. We also report the finding of two other open reading frames (ORF) in these clusters. *orfX*, which has an unknown function, is located between *mysA* and *mysB*. The other ORF, located downstream from *mysB*, encodes a homolog of DtxR, the iron regulatory protein from *Corynebacterium diphtheriae* (*Cd*).

INTRODUCTION

The diseases caused by *Mycobacterium tuberculosis* (*Mt*) and *M. leprae* (*Ml*) remain major causes of human morbidity. Despite the importance of these pathogens, our knowledge of their virulence determinants is still very limited (Young and Cole, 1993). For an understanding of the pathogenicity of mycobacteria, it is important to know how these organisms control expression of genes which allow survival and multiplication in the macro-

phages. To approach this problem, our first goal was to study the transcriptional apparatus in mycobacteria. We initially purified and characterized the *M. smegmatis* (*Ms*) RNA polymerase (Predich et al., 1995) and also cloned and sequenced two σ^{70} class genes, *mysA* and *mysB*, which encoded σ^A and σ^B , respectively. Both genes are co-localised on a 9-kb *EcoRI* fragment of the *Ms* chromosome. σ^A is thought to be the major mycobacterial sigma factor based on its stronger sequence similarity to σ^{HrdB} , the major sigma factor of *Streptomyces coelicolor* (Tanaka et al., 1991) whose structural gene was used as a probe for the cloning. In this communication we report on similar studies with the corresponding region of the *Mt* and *Ml* chromosomes.

Correspondence to: Dr. I. Smith, PHRI, 455 First Av., New York, NY 10016, USA. Tel. (212) 578-0867; Fax (212) 578-0804; e-mail: smitty@phri.nyu.edu.

Abbreviations: aa, amino acid(s); bp, base pair(s); *Cd*, *Corynebacterium diphtheriae*; DtxR, diphtheria toxin repressor; *dtxR*, gene encoding DtxR, kb, kilobases(s) or 1000 bp; *Mt*, *Mycobacterium tuberculosis*; *Ml*, *M. leprae*; *Ms*, *M. smegmatis*; *Mt*, *M. tuberculosis*; *mysA* and *mysB*, genes encoding σ^A and σ^B ; nt, nucleotides(s); ORF (*orf*), open reading frame; PCR, polymerase chain reaction; *S.*, *Streptomyces*; σ^A and σ^B , sigma factors A and B.

EXPERIMENTAL AND DISCUSSION

(a) Mycobacterial σ genes

In preliminary Southern experiments, DNA from *M. chelonii*, *M. goodii*, *M. kansasii* and *M. bovis* BCG were

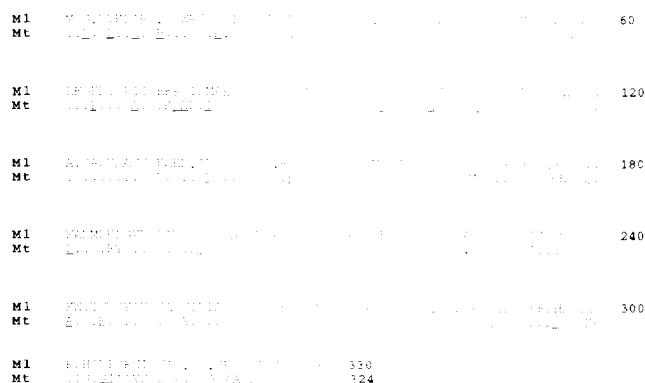


Fig. 3. The aa sequence of the mycobacterial *orfX*. The dots, letters and numbers have the same significance as in Fig. 1. The corresponding region of the *Ms* chromosome has not been sequenced, but PCR analysis indicates the presence of *orfX* in the same chromosomal location.

(b) Mycobacterial *orfX*

During the sequencing of the chromosomal region between *mysA* and *mysB* in *Mt* and *Ml* cosmids, we noted a highly conserved ORF (323 aa in *Mt* and 330 aa in *Ml*) approximately 1 kb upstream from *mysB* (Fig. 3). Its function is completely unknown as it shows no relationship to any protein product currently in existing data-banks but both ORFs have a very high content of hydrophobic aa. As expected from the deduced aa sequences, the hydrophobicity profiles are almost identical. More importantly, they suggest that the proteins are integral membrane components as they have seven putative transmembrane segments (data not shown). The corresponding region of the *Ms* genome has not been sequenced, but PCR and Southern analyses indicate that *orfX* is also found in this species, in the same location between *mysA* and *mysB* (data not shown).

(c) Mycobacterial *dtxR*

We also sequenced regions of the *Ms* and *Mt* chromosomes adjacent to the σ encoding genes. In both organisms, immediately downstream from *mysB* (134 nt in *Mt* and 223 nt in *Ms*), we found an ORF with extremely high similarity to DtxR (Fig. 4), the iron-binding repressor of iron uptake and the production of toxin in *Cd* (Boyd et al., 1990; Schmitt and Holmes, 1991). Overall identity between the 226-aa DtxR and the two mycobacterial ORFs was 58% and conservation was 70%. However, in the region defined by the first 140 aa, the comparable figures are 80 and 87%, respectively. We have not sequenced yet the region of the *Ml* chromosome that is expected to contain the *dtxR*, since cosmid B1764, originally used for sequencing of the *mysA*-*mysB* region terminates immediately downstream from *mysB*. We have analysed by PCR *Ml* cosmid B852, which contains sequences adjacent to *mysB* (data not shown). These studies show that nt sequences corresponding to the *dtxR*

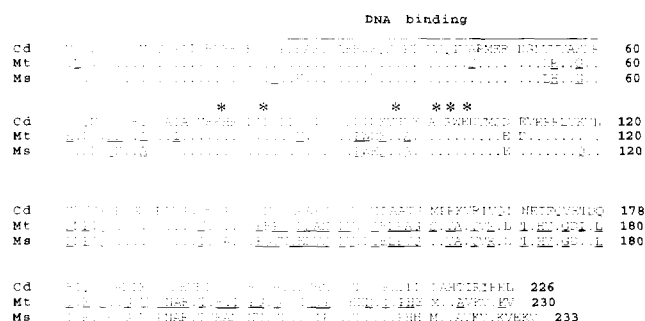


Fig. 4. The aa sequences of the *Cd* DtxR and the mycobacterial homologs. The dots, letters and numbers have the same significance as in Fig. 1. The region containing the DNA-binding domain (aa 26–52) of the corynebacterial DtxR is underlined. The asterisks indicate the residues His⁷⁹, Glu⁸³, His⁹⁸, Cys¹⁰², Trp¹⁰⁴, and His¹⁰⁶ which form part of the metal-binding pocket (Tao et al., 1994; Qiu et al., 1995). The GenBank access number for the *Mt* *dtxR* sequence is U14191 and for the corresponding *Ms* sequence is U14190.

gene are found in *Ml* and the intergenic space between this region and *mysB* is identical to that of *Mt*. We have done Southern analysis with genomic DNAs isolated from *M. avium*, *M. bovis* BCG, *Mt* and *M. hemophilum*. Using an internal fragment of the *Ms* *dtxR* gene as a probe, we were able to detect hybridizing bands in all samples tested except *M. hemophilum* (L.D., B. Kreiswirth and I.S., unpublished data). Interestingly, we have noted in the GenBank database (accession No. M34239, Oct 1991) that there is a uncharacterized ORF (115 aa) in the *Cd* chromosome 225 bp upstream from the *dtxR* gene that has 75% identity and 83% conservation with the C-terminal region of the mycobacteria *mysB* (Fig. 2). We believe this ORF actually is the C-terminal part of the corynebacterial σ^B . The organization of the gene cluster, as described in sections a–c is summarized schematically in Fig. 5.

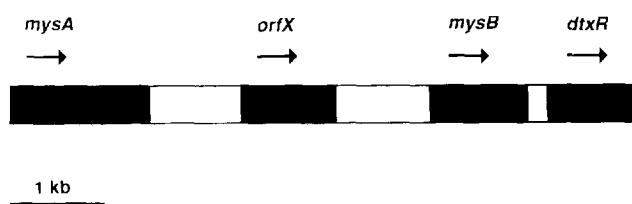


Fig. 5. Map of the mycobacterial gene cluster. This schematic diagram is a composite of data obtained from complete sequencing of a 10-kb chromosomal fragment from *Ml*, a 6-kb fragment from *Mt* and approximately 4 kb around *mysA* and around *mysB*-*dtxR* in *Ms* genome. The distances are roughly drawn to scale, but there are some differences when each organism is compared, e.g., the intergenic space between *mysA* and *mysB* is approximately 900 bp longer in *Ml* than in the other two mycobacteria. The region of the chromosome expected to contain *dtxR* in *Ml* has not been sequenced, as is the case with the *orfX* region of *Ms*, but PCR analyses indicate the presence of these genes in the respective organisms. Arrows indicate the orientation of the genes.

(d) Conclusions

(1) We have demonstrated that the genes, *mysA* and *mysB*, encoding sigma A and sigma B, are maintained in a contiguous cluster structure in *Ml* and *Mt*, 3 to 4 kb apart, as previously demonstrated in *Ms* (Predich et al., 1995). We have shown that this genomic structure is found in other mycobacterial species as well. The very high degree of identity we observed for the σ^A protein encoded by *mysA* in the three species, as well as the high sequence similarity to *hrdB* of *S. coelicolor* suggests that *mysA* codes for the major mycobacterial σ factor. We found highly variable aa sequences in the 1.1 region of the *mysA* gene. This region may be under less selection pressure than other domains of the protein.

(2) We have found *orfX*, an ORF located between *mysA* and *mysB* in *Ml*, *Mt* and *Ms*. Its conserved aa sequence does not show any resemblance to previously described proteins or ORFs, but the presence of several putative transmembrane hydrophobic domains suggests that it may be an integral membrane protein.

(3) We have found in *Mt*, *Ml* and *Ms* a gene coding for a protein that, on the basis of its structure, is the homolog of the DtxR protein of *Cd* (Tao et al., 1994). There is a very high similarity when the first 140 aa of the corynebacterial DtxR and the mycobacterial homologs are compared (Fig. 4). Mutational studies have shown that the N-terminal region of DtxR contains the DNA-binding domain at residues 26–52 and the metal-binding region at His⁷⁹, Glu⁸³, His⁹⁸, Cys¹⁰², Trp¹⁰³ and His¹⁰⁴ (Wang et al., 1994; Tao et al., 1994; Qiu et al., 1995). The C terminus is not as conserved, but all of the essential domains of *dtxR* are conserved in the mycobacterial homologs (Fig. 4). Recently, we have shown, in collaboration with M. Schmitt and R. Holmes, that the mycobacterial *dtxR* protein behaves identically to its corynebacterial counterpart. In an iron dependent manner, in vivo, it represses corynebacterial genes containing *dtxR* operator sites and also binds to these sites in vitro, giving footprints identical to those obtained with the corynebacterial DtxR (Schmitt et al., 1995).

(4) It is of interest that, as in mycobacteria, the *Cd* *dtxR* gene is located approximately 200 bp from the 3' end of a gene showing high similarity to *mysB*. In addition, it has recently been shown that a *dtxR* homolog in the Gram-positive coryneform bacteria *Brevibacterium lactofermentum* is closely linked to a σ factor gene homologous to *mysB* (Oguiza et al., 1995). These observations suggest a functional significance to this close linkage of *mysB* and *dtxR* in mycobacteria and corynebacteria.

ACKNOWLEDGMENTS

We would like to thank: E. Dubnau, D. Dubnau, B. Kreiswirth and L. Mindich for their helpful discussions. This work was supported by Grant GM 19693 from the National Institutes of Health (awarded to I.S.), funds from a grant from the Markey Foundation given to the Public Health Research Institute and by a fellowship from the Ministère de l'Éducation Supérieure et de la Recherche (awarded to O.D.).

REFERENCES

- Boyd, J., Oza, M.N. and Murphy, J.R.: Molecular cloning and DNA sequence analysis of a diphtheria toxin iron-dependent regulatory element (*dtxR*) from *Corynebacterium diphtheriae*. Proc. Natl. Acad. Sci. USA 87 (1990) 5968–5972.
- Dombroski, A.J., Walter, W.A. and Gross, C.A.: N-terminal amino acids modulate σ factor DNA-binding activity. Genes Dev. 7 (1993) 2446–2455.
- Eigelmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.: Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. Mol. Microbiol. 7 (1993) 197–206.
- Lonetto, M., Gribskov, M. and Gross, C.A.: The σ^{70} family: sequence conservation and evolutionary relationships. J. Bacteriol. 174 (1992) 3843–3849.
- Oguiza, J.A., Tao, X., Marcos, A.T., Martin, J.F. and Murphy, J.R.: Molecular cloning, DNA sequence analysis, and characterization of the *Corynebacterium diphtheriae* *dtxR* homolog from *Brevibacterium lactofermentum*. J. Bacteriol. 177 (1995) 465–467.
- Predich, M., Doukhan, L., Nair, G. and Smith, I.: Characterization of RNA polymerase and two σ factor genes from *Mycobacterium smegmatis*. Mol. Microbiol. 15 (1995) 355–366.
- Qiu, X., Verlinde, C.L., Zhang, L., Schmitt, M.P., Holmes, R.K. and Hol, W.G.: Three-dimensional structure of the diphtheria toxin repressor in complex with divalent cation co-repressors. Structure 3 (1995) 87–100.
- Schmitt, M.P. and Holmes, R.K.: Characterization of a defective diphtheria toxin repressor *dtxR* allele and analysis of *dtxR* transcription in wild-type and mutant strains of *Corynebacterium diphtheriae*. Infect. Immun. 59 (1991) 3903–3908.
- Schmitt, M.P., Predich, M., Doukhan, L., Smith, I. and Holmes, R.K.: Characterisation of the DtxR homolog from *Mycobacterium tuberculosis*. Infect. Immun. (1995) in press.
- Tanaka, K., Shiina, T. and Takahashi, H.: Nucleotide sequence of genes *hrdA*, *hrdC*, and *hrdD* from *Streptomyces coelicolor* A3(2) having similarity to *rpoD* genes. Mol. Gen. Genet. 229 (1991) 334–340.
- Tao, X., Schiering, N., Zeng, H.-Y., Ringe, D. and Murphy, J.R.: Iron, DtxR, and the regulation of diphtheria toxin expression. Mol. Microbiol. 14 (1994) 191–197.
- Wang, Z., Schmitt, M.P. and Holmes, R.K.: Characterization of mutations that inactivate the diphtheria toxin repressor gene (*dtxR*). Infect. Immun. 62 (1994) 1600–1608.
- Young, D.B. and Cole, S.T.: Leprosy, tuberculosis, and the new genetics. J. Bacteriol. 175 (1993) 1–6.