The *Mycobacterium tuberculosis* purine biosynthetic pathway: isolation and characterization of the *purC* and *purL* genes

Mary Jackson,¹ François-Xavier Berthet,¹ Isabel Otal,² Jean Rauzier,¹ Carlos Martin,² Brigitte Gicquel¹ and Christophe Guilhot¹

Author for correspondence: Mary Jackson. Tel: +33 1 45 68 88 77. Fax: +33 1 45 68 88 43.
e-mail: mjackson@pasteur.fr

¹ Unité de Génétique Mycobactérienne (CNRS, URA 1300), Institut Pasteur, F-75724 Paris, France

² Departamento de Microbiología, Facultad de Medicina, Universidad de Zaragoza, Domingo Miral s/n 50009, Zaragoza, Spain

Genes from the *Mycobacterium tuberculosis* purine biosynthetic pathway were identified using purine auxotrophic mutants of *Mycobacterium smegmatis* obtained by Tn611 transposon mutagenesis. Two approaches were followed in parallel. The first consisted of the complementation of the *M. smegmatis* purine auxotrophs using a *M. tuberculosis* H37Rv shuttle cosmid library. In the second approach, specific probes corresponding to the regions adjacent to the insertion sites of Tn611 in the *M. smegmatis* genome were used to screen a *M. tuberculosis* plasmid library by colony hybridization for inserts carrying homologous DNA fragments. Nucleotide sequence analysis of two *M. tuberculosis* genes isolated by these methods revealed high similarities with *purC* and *purL* genes from other bacterial and fungal sources. Transcriptional start sites were mapped for both genes, which revealed similar –10 boxes but with a higher GC content than the *Escherichia coli* σ^{70} consensus.

Keywords: *Mycobacterium tuberculosis*, auxotrophy, complementation, purine biosynthetic pathway

INTRODUCTION

Mycobacterium tuberculosis, the aetiological agent of tuberculosis, continues to be an important cause of human morbidity and mortality worldwide (Dolin *et al.*, 1994). Chemotherapy and prophylactic vaccination using *Mycobacterium bovis* BCG (bacillus Calmette-Guérin) have been employed against mycobacterial infection (Collins, 1993). However, BCG was shown to have a protective effect against the adult form of tuberculosis varying from 0% in trials in the south of India to 77% in the UK (Bloom & Murray, 1992). The reasons for such differences in the protection BCG provides and the mechanisms whereby this vaccine induces protective immunity remain unclear.

A major area of mycobacterial research over the last 10 years has been directed towards the identification of major antigens involved in the interaction with the immune system, with the aim of developing 'subunit' vaccines that are effective against mycobacterial infections (Young *et al.*, 1990; Andersen, 1994; Averill *et al.*, 1993). The

recent development of molecular genetic methods for mycobacteria (Hatfull, 1993; Guilhot *et al.*, 1994; Husson *et al.*, 1990; Jacobs *et al.*, 1987) provides alternative approaches to the design of new antimycobacterial vaccines by the construction of mutant strains of *M. tuberculosis* that are rationally attenuated for virulence or maintenance in host cells. This approach has been widely applied among bacterial pathogens. Studies with *Salmonella typhimurium* (Hoiseth & Stocker, 1981; Levine *et al.*, 1987; Fields *et al.*, 1986; Curtiss, 1990), *Vibrio cholerae* (Baselky *et al.*, 1979), *Brucella melitensis* (Drazek *et al.*, 1995) and *Rhizobium* (Newman *et al.*, 1994) have demonstrated that strains carrying mutations in their purine or aromatic amino acid biosynthetic pathways show attenuated virulence; in the case of *S. typhimurium*, such strains induced a protective immunity in mice (O'Callaghan *et al.*, 1988).

With the aim of constructing mutant strains of *M. tuberculosis* that are defective in purine biosynthesis, and evaluating their virulence and immunogenicity, we have characterized two purine biosynthetic loci: *purL* and *purC*. We report here the cloning and the characterization of these two *pur* genes involved in the *M. tuberculosis de novo* purine biosynthetic pathway.

The GenBank accession numbers for the sequences reported in this paper are U34956 (*purL*) and U34957 (*purC*).

METHODS

Bacteria, media and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. *Escherichia coli* cultures were routinely grown in Luria B medium (Difco) at 37 °C. *Mycobacterium* cultures were grown in Middlebrook 7H9 medium (Difco) supplemented with glycerol (0.2%) and ADC (glucose, 0.2%; BSA fraction V, 0.5%; and NaCl, 0.085%) at 37 °C with the exception of the strains carrying the thermo-sensitive pCG63 plasmid, which were grown at 30 °C. Minimal medium (M63) was used as the purine-limiting medium for mycobacteria and contained, per litre: 10 g (NH₄)₂SO₄; 68 g KH₂PO₄; 2.5 mg FeSO₄·7H₂O; 1 mM MgSO₄·7H₂O and 2 g glucose. Antibiotics were added where required at the following concentrations: ampicillin, 100 µg ml⁻¹; kanamycin, 20 µg ml⁻¹; lividomycin, 20 µg ml⁻¹; butyrosin, 10 µg ml⁻¹.

Complementation analysis. The cosmid library of *M. tuberculosis* H37Rv in the cosmid vector pYUB18 has been described previously (Jacobs *et al.*, 1991), as has the *M. smegmatis* mutant library containing purine auxotrophic mutants (Guilhot *et al.*, 1994). The cosmid library was transformed into *E. coli* and DNAs from the resulting colonies were pooled and electroducted into the *M. smegmatis* *auxC* mutant strain. Electroporations were performed as described previously (Guilhot *et al.*, 1992). A cosmid library clone was isolated which was able to complement the *M. smegmatis* *auxC* strain grown on minimal medium. Cosmids selected for further analyses were electroporated into *E. coli*, as described by Baulard *et al.* (1992). Complementation analysis was carried out the same way, using the shuttle plasmids pCG63 and pOMK.

Subcloning procedures. Restriction enzymes and T4 DNA ligase were purchased from Gibco BRL, Boehringer Mannheim and New England Biolabs. All enzymes were used in accordance with the manufacturer's recommendations. DNA fragments used in the cloning procedures were gel purified using the GeneClean II kit (BIO 101). Cosmids and plasmids were isolated by alkaline lysis (Sambrook *et al.*, 1989).

Southern blot analysis and colony hybridization. DNA fragments for radiolabelling were separated on 0.7% agarose gels (Gibco BRL) in a Tris/borate/EDTA buffer system (Sambrook *et al.*, 1989) and isolated from the gel by using GeneClean II (BIO 101). Radiolabelling was carried out with a random primed labelling kit (Megaprime, Amersham) with 5 µCi (185 kBq) [α -³²P]dCTP, and unincorporated label was removed by passing through a Nick Column (Pharmacia). Southern blotting (Southern, 1975) was carried out in 0.4 M NaOH with nylon membranes (Hybond-N+, Amersham); prehybridization and hybridization were carried out as recommended by the manufacturer using RHB buffer (Amersham). Washing at 65 °C was as follows: two washes with 2 × SSPE (SSPE is 150 mM NaCl, 8.8 mM NaH₂PO₄, 1 mM EDTA pH 7.4)/0.1% SDS, each for 15 min; one wash with 1 × SSPE/0.1% SDS for 10 min; two washes with 0.7 × SSPE/0.1% SDS, each for 15 min. Autoradiographs were prepared by exposure with X-ray film (Kodak X-Omat AR) at -80 °C overnight.

Colony hybridization was carried out using nylon membrane discs (Hybond-N+ 0.45 µm, Amersham). *E. coli* colonies adsorbed on the membranes were lysed in 0.5 M NaOH/1.5 M NaCl before being placed for 1 min in a microwave oven to fix the DNA. Hybridization and washings were as described for the Southern blotting analysis. The plasmid library that was used to clone the *purL* gene by colony hybridization consisted of *Pst*I restriction fragments sized between 2 and 7 kb after complete digestion of *M. tuberculosis* genomic DNA, which were ligated into pBluescript KS⁻ (Stratagene). This library therefore partially represented the *M. tuberculosis* genome.

Oligonucleotide and PCR amplifications. PCR amplifications were carried out in a DNA thermal Cycler (Perkin Elmer), using *Taq* polymerase (Cetus) according to the manufacturer's recommendations. Primers used for amplifying the regions adjacent to the insertion sites of the transposon Tn611 in the *M. smegmatis* *auxM* auxotroph were derived from the IS6100 insertion sequence and were A (5'-GCATTGGGCCCTCGGCGCTG-

Table 1. Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Source or reference
<i>E. coli</i> XL1 blue	<i>supE44 bsdR17 recA1 gyrA46 thi relA1 lac F'</i> [<i>proAB</i> + <i>lacI</i> ^a <i>lacZ</i> ΔM15 Tn10 (Tet ^R)]	Sambrook <i>et al.</i> (1989)
<i>M. smegmatis</i> mc ² 155	High-transformation mutant of <i>M. smegmatis</i> ATCC 607	Snapper <i>et al.</i> (1990)
<i>M. tuberculosis</i> H37Rv	Virulent strain of mycobacterium originally isolated from a tuberculosis patient	Steenken & Gardner (1946)
<i>M. bovis</i> BCG 1173 P2	Vaccine strain	Institut Pasteur
pYUB18	(Km-But) ^R shuttle vector	Jacobs <i>et al.</i> (1991)
pBluescriptII KS ⁻	Phagemid derived from pUC19 cloning vector	Stratagene
pCG63	Temperature-sensitive derivative of pAL5000; (Liv-Km) ^R	Guilhot <i>et al.</i> (1992)
pCG79	Derivative of pCG63 containing Tn611; (Str-Spc) ^R , (Liv-Km) ^R	Guilhot <i>et al.</i> (1994)
pIPJ46	pSL1180 cloning vector derivative carrying the ORIM, <i>aph</i> cassette; Km ^R	Timm <i>et al.</i> (1994b)
pJEM12	Promoter probe plasmid vector containing the <i>E. coli lacZ</i> gene devoid of its promoter, ribosome-binding site and initiation codon; Km ^R	Timm <i>et al.</i> (1994a)
pOMK	pBluescript KS ⁻ derivative carrying the ORIM, <i>aph</i> cassette in the <i>ScaI</i> restriction site; Km ^R	This work
pMJ1	pCG63 derivative carrying the 2.5 kb insert containing the <i>purC</i> gene	This work
pMJ2	pOMK derivative carrying the 5.5 kb insert containing the <i>purL</i> gene	This work

3') and E (5'-AAGGCGCACCTCGCCCCTGATG-3'). The PCR reaction was carried out on the *M. smegmatis auxM* mutant chromosomal DNA, digested by *Pst*I and religated using T4 DNA ligase (Gibco BRL).

DNA sequencing and analysis. Sequences of double-stranded plasmid DNA were determined by the dideoxy chain-termination method (Sanger *et al.*, 1977) using the *Taq* Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems), on a GeneAmp PCR System 9600 (Perkin Elmer), and run on a DNA Analysis System model 373 stretch (Applied Biosystems). The sequence was assembled and processed using DNA Strider (CEA, France) and the University of Wisconsin Genetics Computer Group (UWCGC) packages. The BLAST algorithm (Altschul *et al.*, 1990) was used to search protein databases for similarity. The sequencing strategy consisted of initially using the M13 universal primers followed by designing primers using the newly acquired sequences.

RNA isolation and primer extension analysis. Total RNA was isolated using the extraction procedure described by Bashyam & Tyagi (1994). The *M. tuberculosis purC* transcriptional start site was mapped by primer extension using the JC6 primer (5'-GGTGCTCGTCATCGACACGGTAGATCTCGC-3'), complementary to nucleotides 147-176. The *purL* transcriptional start site was mapped using the JL3 primer (5'-CCGGGCTCACCATACGGTTGTGGTTGGGCGGGTGTGG-3'), complementary to nucleotides 200-232. The experimental procedure used was as described by Berthet *et al.* (1995).

β -Galactosidase gene fusion and assays. A *purC* *Bam*HI/*Kpn*I PCR amplification product corresponding to nucleotides 1-690 was ligated into the *Bam*HI/*Kpn*I sites of pJEM12, a shuttle plasmid containing the truncated *lacZ* gene devoid of the promoter, ribosome-binding site and initiation codon. The pJEM12C vector was transformed into *M. smegmatis* mc²155 and cells were grown in minimal medium (M63), M63 plus adenine (20 μ g ml⁻¹) and guanine (20 μ g ml⁻¹) or 7H9 + ADC + Tween (TW, 0.05%). β -Galactosidase activity was assayed in sonicated extracts of *M. smegmatis*, after 24 h growth at 37 °C, by the method of Pardee *et al.* (1959). Experiments were carried out in triplicate. Units of β -galactosidase activity were calculated by the following formula: $200 \times A_{420} \text{ (mg protein)}^{-1} \text{ min}^{-1}$. Soluble protein in bacterial extracts was measured by the Bio-Rad assay.

RESULTS

Isolation of the *purC* gene

M. smegmatis mc²155 mutants deficient in purine biosynthesis were isolated from a mutant library by transposon mutagenesis using Tn611 (Guilhot *et al.*, 1994). Two *pur* mutants (designated *auxC* and *auxM*) were plated onto minimal agar plates (M63 medium) to check the stability of the mutations. In both cases reversion frequencies were very low: 3×10^{-7} per cell per generation for *auxC* and 1×10^{-6} per cell per generation for *auxM*. Since both mutants grew on M63 plus adenine and guanine, as well as on M63 plus hypoxanthine, we hypothesized that the *auxC* and *auxM* mutants carry mutations in the common part of the *de novo* purine biosynthetic pathway. A genomic library, constructed by inserting *San*3A partially digested *M. tuberculosis* H37Rv DNA into the unique *Bam*HI site of the mycobacterial cosmid shuttle vector pYUB18 (Jacobs *et al.*, 1991), was used to transform the *M. smegmatis auxC* auxotroph by electro-

Fig. 1. *Pst*I digestion of ten cosmids conferring a Pur⁺ phenotype upon the *auxC* mutant. A, B and J, cosmid C₃; C, D, E and I, cosmid C₇; F, G, H, cosmid C₉. MT, size markers (1 kb ladder, Gibco BRL). The digests were electrophoresed in a 0.7% agarose gel and stained with ethidium bromide.

poration. After 4 h growth in 7H9 + ADC + 0.05% TW medium at 37 °C, transformants were plated onto minimal medium plates and onto rich medium plates (L medium) containing butyrosin (10 μ g ml⁻¹) to select for the presence of a cosmid. After 6 d incubation at 37 °C, colonies were detected on both media. The number of colonies on the M63 plates was approximately 0.6% of that on the rich medium. To check whether the colonies obtained on minimal media corresponded to complemented mutants and not to revertants, colonies were plated onto L medium containing lividomycin (20 μ g ml⁻¹) and butyrosin (10 μ g ml⁻¹). These assays confirmed the presence of the Tn611 transposon, which harbours a gene conferring resistance to lividomycin, and of a complementing cosmid, in 97% of the colonies grown on minimal medium. Cosmid DNA from ten prototrophic transformants was reintroduced into *E. coli* XL1 blue by electroporation for further analyses.

*Pst*I digestion of the ten cosmids revealed that the clones contained three different types of cosmid inserts (designated C₃, C₇ and C₉) sharing common bands, thus suggesting that the clones all overlapped (Fig. 1). Electrotransformation of each of the three types of cosmids into the *auxC* mutant in each case restored a Pur⁺ phenotype, confirming the presence on the cosmid inserts of a gene complementing the deficient function of the auxotrophic mutant.

Fig. 2. Nucleotide and derived amino acid sequences of the *M. tuberculosis purC* gene. The mapped transcriptional start site with the predicted -10 and -35 boxes are represented in bold italic letters. Arrows indicate the positions of inverted repeats between nucleotides 43 and 58 and nucleotides 71 and 85.

To isolate the complementing gene, we used a subcloning strategy involving deletion of DNA by restriction digestion and subcloning of common restriction fragments in the shuttle plasmid pCG63 (Guilhot *et al.*, 1992). Each subclone was assayed for complementation in the *auxC* strain on minimal medium. A 2.5 kb *Bam*HI fragment, carried by pMJ1, retained the complementation activity. This fragment was sequenced. It contained one complete ORF of 893 bp starting with a potential ATG initiation codon at position 96 (Fig. 2). The G+C content of the DNA was 64 mol %, which is similar to the estimates for *M. tuberculosis* DNA (Clark-Curtiss, 1990). The predicted amino acid sequence derived from this ORF showed high homology with the PurC protein of *Candida maltosa* (45% identity, 66% including conserved residues) (Sasnauskas *et al.*, 1991) and *Saccharomyces cerevisiae* (44% identity, 66% similarity) (Myasnikov *et al.*, 1991), and lower homology with the PurC protein of *Bacillus subtilis* (25% identity, 51% similarity) (Ebbole & Zalkin, 1987) and *E. coli* (26% identity, 47% similarity) (Tiedemann *et al.*, 1990). We therefore suggest that we have cloned the *M. tuberculosis purC* gene encoding phosphoribosylaminoimidazole-succinocarboxamide synthase, which catalyses the conversion of 5'-phosphoribosyl-5-aminoimidazole-

Fig. 3. Identification of the *purL* gene by Southern blot analysis. The C_3 cosmid DNA was digested with different restriction enzymes, as indicated above the lanes, and each preparation was then loaded on a 0.7% agarose gel and electrophoresed overnight at 30 V. The gel was blotted and hybridized with the 32 P-labelled 2.3 kb *Pst*I restriction fragment.

carboxylic acid to 5'-phosphorybosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole in the purine biosynthetic pathway.

As genes involved in the *de novo* purine biosynthetic pathway from other organisms such as *B. subtilis* can be organized in operons (Ebbole & Zalkin, 1987), DNA sequencing was also undertaken to characterize the neighbouring genes. However, analysis of 0.6 kb DNA located upstream and of 1 kb DNA located downstream of the *purC* ORF showed no homology with any other purine biosynthetic gene.

Isolation of the *purL* gene

In a second approach to isolate *M. tuberculosis* purine biosynthetic genes, we amplified by PCR the regions adjacent to the insertion sites of Tn611 in the *M. smegmatis auxM* mutant. We then used the amplified region as a 32 P-radiolabelled probe to screen a *M. tuberculosis* H37Rv plasmid library, by colony hybridization, for inserts containing the probe. The probe was constructed as follows. The *auxM* mutant genomic DNA was digested with *Pst*I, and the restriction fragments were circularized by ligation. As the insertion sequence IS6100 from Tn611 contains an internal *Pst*I site, *Pst*I digestion followed by religation gives two IS-containing circular fragments, one

1	CAACCTGAACCCCCCCCCGATTTCAACCCGACGGCTGATTCCCGGGGGTCTTTCCGC	2101	GTGTCCCGCGCGACGATCTGTCCGAGGGTG99CTGGCCCAAGCCATCGTGAATCGCG
61	GGGCAACCCAAACCGGGCTTGTCCGTTTC -35 -10 TAGACTG CTCCCTCCGTGATGTCTCCGCTCGCCGGACCCCGCGCAAAACGTCCGTGCTGGAC	2161	TTAGCGGGTGAAACCGGTTCCCGCATAGTCTCCCGAAGCGGCTGACCCGTTTGTGCTG
121	+1 TAGACTG CTCCCTCCGTGATGTCTCCGCTCGCCGGACCCCGCGCAAAACGTCCGTGCTGGAC M L D	2221	CTGTCTCCGAGTCCGGCGGTTCGGTCTGGTTCGGGTGCCACGCCAGGAGAGCGCG
181	ACCGTGAACACCGCGGACACACCGGACCAACCAACCGTATGGTGAAGTGGGCTC	2281	TTTCGGGGATGTGTGAGGCGCGGGACTTCCCGCGTCCGATCGCGCTTGTGATCAA
241	AAAGACGACGAGTACCGCGGATTCGCCAGATCCTGGACCCCGGCCACCGACACCGAG	2341	GGTTCGGACCGGTTGAGGTGCAGGGCTTGTTCGGGTGTCTTGGCGGAAGTCCGTGG
301	CTGGCGATGACTCGGTGATGTGGAGCGAAGCTGTTCGTACAAGTCTCCGCGTGCAC	2401	ACATCCGAGCGGTGTTCGGCGGATACTTCGGATGAGTCCGCTTCGCGCCCTGTCTTGT
361	CTGGCTACTTCGGTGAGACACCTCCGACGAGATCGCGCGGTTCATGCTGGTCCGATC	2461	TTG
421	GGCGAGAAGCGCGGCTCGTGCATCGCGACGCGTGGCGGTACCTTCAAGGTGGAG		
481	TCACACAACACCGCTCTACGTCGAGCGCTACCGAGGCGCGCCACCGGGTGGCGGC		
541	ATCGTCCGGACATCATGGCCATGGCGCGCGACCGGTTCGGTGTGAGACGCTTCGG		
601	TTCCGCGCGCGGACCGCCCGATACCGCGCGGTGCTCGACCGCTGTTCGCGCGATC		
661	CGCGATACGGCAACTCCCTGGCGCTGCCAACATTTGGCGGAGACCGCTTTCGACCCG		
721	TGCTACCGCGCAACCCCTTAGTGAACCGGTGTGTGTTCGGGTATTACGCCAGGAGGAC		
781	CTGCATTGGCGGTTCGGTTCGGCGCGCGCAACAGATCATCTGTTTGGCGCGCGCAC		
841	GGGCTCGACGATATCGCGCGGTGTTCGGTGTTCGGTTCGGTTCGGTTCGGTTCGGT		
901	TCCCGCAAGAGCTGCGCTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
961	GAATGCTGTCTCGAGCTTCACGCGCGCTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1021	gccccattatcttggccacatcgaggttagcatccgccccggtgatggcggaatgacgac		
1081	CAGCTGGACAGCGTCCCGCTCGCGCGCGAGGAGATGACCGCGCGCGAGGTCTTCGACG		
1141	GAATCGCAGGAGCGAGTGTTCGGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1201	GTGTCCCGCAAGTGGGAGGTGTTCGGCGCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1261	CTGCAGATCACTGGCGCGCGAGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1321	GAAGTTCGGTATATCAGCGCGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1381	GACCGTTCGGCAAGCTGTACGGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1441	CGTTTACTTGGCAGTCCGACCTGTGTCAGCGCGGTTCGGTTCGGTTCGGTTCGGT		
1501	TATGTTCGGCGCAACCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1561	GAGTGCAGCGCGCGGATCGCGGTATCGACCGCGGTTCGGTTCGGTTCGGTTCGGT		
1621	GATCCCTACGCTGGCGCGCACTCCGCTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1681	GGCGCCACCCCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1741	GTGATGTGGCAGTTCACGCGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1801	ATTCCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1861	CCCACCGCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGGT		
1921	GGCTTCGGCGCGCGCGCGCAACCTTCGATCTGTTCGGCGACACCGCGCGAGTTC		
1981	GACGGTTCGGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG		
2041	GATCTGGCG		

Fig. 4. Nucleotide and derived amino acid sequences of the *M. tuberculosis purL* gene. The mapped transcriptional start site, together with the predicted -10 and -35 boxes are represented in bold italic letters.

corresponding to the region located immediately downstream from the transposon insertion site and one corresponding to the region located immediately upstream from the transposon insertion site. A PCR reaction was then carried out on the ligation products using internal outward IS6100 oligonucleotides as primers (see Methods). An amplified fragment of 220 bp corresponding to the region upstream of the transposon insertion site was obtained and used as a radiolabelled probe to hybridize to *M. tuberculosis* H37Rv genomic DNA digested with *Pst*I. The probe hybridized to a single 2.3 kb *Pst*I fragment, and therefore seemed to be specific to a particular chromosomal region. The same 220 bp probe was used to screen, by colony hybridization, a plasmid genomic library containing *Pst*I-digested genomic DNA inserts of *M. tuberculosis* H37Rv. Among 4000 colonies that were screened, two contained a plasmid which hybridized to the probe. Analysis of these two clones revealed they contained identical plasmids carrying a 2.3 kb insert, which corresponded to the expected size according to the previous Southern blotting experiment with *M. tuberculosis* genomic DNA. This 2.3 kb *Pst*I insert was subcloned into the shuttle plasmid pOMK to assay for complementation of the *M. smegmatis auxM* mutant. Its failure to restore the *Pur*⁺ phenotype suggested that the complementing gene was not entirely contained in the *Pst*I fragment or that it was not expressed. Further hybridization experiments using the *Pst*I 2.3 kb insert as a radiolabelled probe revealed that this insert hybridized to one of the three cosmids (*C*₉) already carrying the *purC* gene. Furthermore, the *C*₉ cosmid electroporated into the *auxM* mutant restored the *Pur*⁺ phenotype.

Hybridization of the *Pst*I 2.3 kb probe on different restriction digestions of the *C*₉ cosmid enabled the isolation of two *Not*I restriction fragments of 5.5 kb and 4.5 kb (Fig. 3), one of which (the 5.5 kb fragment), once subcloned into pOMK shuttle plasmid (resultant plasmid: pMJ2), was able to complement the *auxM* mutant. With the *Bam*HI digestion of the *C*₉ cosmid, the probe hybridized to a fragment distinct from that carrying the *purC* gene, suggesting that we had cloned two different

Fig. 5. Mapping of the *purC* (a) and *purL* (b) transcriptional start sites by primer extension using the JC6 primer and JL3 primer respectively. Lanes 1 correspond to primer extension reactions performed with RNA from *M. smegmatis* carrying pMJ1 (a) or pMJ2 (b). Lanes 2 correspond to reactions performed with no RNA (negative control). The end of the primer extension product is indicated by an asterisk.

genes involved in the *M. tuberculosis* purine biosynthetic pathway.

The complementing gene carried by the 5.5 kb *NotI* fragment was sequenced. Sequence analysis of the insert revealed an ORF starting with a potential GTG initiation codon at position 172, and encoding a 755-amino-acid polypeptide (Fig. 4). This shares extensive sequence similarity with previously characterized bacterial phosphoribosylformylglycinamide synthases. The *M. tuberculosis* sequence shares 90% identity at the amino acid level and 82% identity at the nucleotide level with the putative *purL* gene from *M. leprae* (located on cosmid B2266 deposited in GenBank as part of the *M. leprae* genome sequencing project; GenBank accession number U15182). The amino acid sequence shares 43% identity (65% including conserved residues) with the corresponding enzyme in *B. subtilis* (Ebbole & Zalkin, 1987) and 28% identity (52% including conserved residues) with the *E. coli purL* gene product (Schendel *et al.*, 1989). Homology with the corresponding enzyme in yeast was lower than for the *purC* gene (25% identity with the *S. cerevisiae* enzyme) (unpublished; GenBank accession number U14566). Interestingly, the very high homology at the nucleotide level between the *M. tuberculosis* and *M. leprae purL* genes is much lower upstream of the GTG potential initiation codon (data not shown). Here again, no other gene involved in the purine biosynthetic pathway was identified within the 1 kb regions flanking the *purL* gene.

Characterization of the *M. tuberculosis purC* and *purL* transcriptional start sites

The *M. tuberculosis purC* and *purL* transcriptional start sites were investigated by primer extension analysis using RNA extracted from *M. tuberculosis* or from *M. smegmatis* mc²155 carrying pMJ1 or pMJ2, grown in 7H9 + ADC + TW medium. Three very close trans-

criptional start sites were identified for the *purC* gene when RNA from *M. smegmatis* carrying pMJ1 was used (Fig. 5a). If the same primer extension analysis was performed using *M. tuberculosis* RNA, only the start corresponding to the A from the ATG start codon was identified (data not shown). This result is probably linked to differences between the *M. tuberculosis* and the *M. smegmatis* transcriptional machinery, which might have different affinities for the same -10 and -35 boxes. The *purL* transcriptional start site was determined using RNA from *M. smegmatis* carrying pMJ2. A dominant start point located at nucleotide 131 was identified (Fig. 5b). The predicted *purC* and *purL* -10 boxes, TAGGCT and TAGACT respectively, share some sequence similarities with the *E. coli* (σ^{70}) and *B. subtilis* (σ^A) -10 consensus TATAAT. The *purC* and *purL* -35 boxes, TACCAG and CACGCG respectively, are more atypical in this respect. The -10 boxes of the two genes are very similar. This is not the case for the -35 boxes. Interestingly, the *M. tuberculosis purL* -10 and -35 motifs are similar to sequences found in the upstream region of *M. leprae purL* gene although no other sequence similarities between the two *purL* genes are found in this region. This suggests a similar promoter region in both species.

Although no inverted repeat motifs were found in the *purL* promoter region, two repeats were found around the *purC* promoter (Fig. 2). Within these regions, residues 43-58 and 71-85 have the potential to form loop structures with a stem containing in both cases exact matches of 5 bp. Such stem-and-loop structures could be elements of a transcriptional regulation mechanism.

Regulation of *purC* expression

To investigate whether the expression of this purine biosynthetic gene could be regulated by the presence or absence of purine bases in the culture medium we

Table 2. β -Galactosidase activities of *M. smegmatis* clones expressing the various fusions to *lacZ* in different media

Plasmid (promoter)	β -Galactosidase activity (U) in:		
	7H9 + ADC	M63	M63 + A + G
pJEM12 (-)	0†	0†	0†
pJEM17 (pblaF*)	2050	3580	3920
pJEM12C (<i>purC</i>)	1470	1613	1563

† A_{420} values below the linear range for β -galactosidase assay.

constructed a fusion between *purC* and the *E. coli lacZ* gene. A 690 bp *purC* fragment from the *Bam*HI site of the 2.5 kb *Bam*HI insert to nucleotide 170 was cloned into the pJEM12 vector (Timm *et al.*, 1994a) and the construction, designated pJEM12C, was electroporated into *M. smegmatis* mc²155 cells. *M. smegmatis* transformants were grown for 24 h at 37 °C in rich medium (7H9 + ADC + TW), minimal medium (M63) or minimal medium complemented with adenine and guanine (20 μ g ml⁻¹ each). β -Galactosidase activity was then assayed in sonicated extracts of *M. smegmatis*. β -Galactosidase assays were carried out simultaneously in *M. smegmatis* mc²155 having received pJEM12 (negative control) or pJEM17, which contains the strong *M. fortuitum* promoter for β -lactamase: pblaF* (positive control) (Timm *et al.*, 1994b). The results are shown in Table 2. The *purC* promoter was shown to have a strong activity as compared to pblaF*, in all three media. No differences in β -galactosidase activity were observed between the bacteria grown in minimal medium and bacteria grown in minimal medium complemented with purine bases.

DISCUSSION

Two genes from the *M. tuberculosis* purine biosynthetic pathway have been isolated and characterized using two different approaches. In the first approach, a *M. smegmatis* purine auxotroph was complemented using a *M. tuberculosis* H37Rv shuttle cosmid library; in the second approach, probes corresponding to Tn611 insertion sites in *M. smegmatis* were used to screen a *M. tuberculosis* plasmid library.

The use of shuttle cosmids to complement a mutant offers two major advantages because of the large size of DNA inserts (30–40 kb) they contain. First, relatively few transformants need to be screened (of the order of 300) for a particular phenotype. Secondly, the inserts are large enough to carry an entire operon; this may be of particular interest where studying biosynthetic pathways for which genes are sometimes organized into an operon, as is the case for the *B. subtilis* purine biosynthetic pathway (Ebbole & Zalkin, 1987). Performing complementation experiments between two mycobacterial species avoids the problem of expression signal recognition that authors

have often encountered when expressing mycobacterial genes in *E. coli* (Gupta *et al.*, 1993). Furthermore, this method enables the study of species-specific processes such as cell wall synthesis, for which there are no corresponding mutants in *E. coli*.

The two genes we isolated do not seem to be organized into a cluster. Furthermore, a shotgun experiment with the C₉ cosmid insert did not reveal the presence of other genes involved in purine nucleotide synthesis (data not shown). Probably the genetic organization of the purine biosynthetic pathway in *M. tuberculosis* more closely resembles that in *S. typhimurium* or *E. coli*. *purC* and *purL* are not linked on the chromosome in the latter bacteria, but are concentrated in certain chromosomal regions, with *purC* being located 2 min from *purL* (Bachmann, 1987; Sanderson & Hurley, 1987).

Differences concerning the transcriptional start sites of *M. tuberculosis purL* and *purC* genes were observed. The *purC* transcriptional start site is located at nucleotide 96 (A from the ATG potential start codon) and *purL* +1 is located 39 bp upstream of the GTG potential start codon. However, the -10 boxes of *purC* (TAGGCT) and *purL* (TAGACT) are very similar and the *purL* -10 box is identical to the putative *rrn* 16S -10 box identified by Ji *et al.* (1994). In contrast the *purC* and *purL* putative -35 boxes show no homology between each other nor to previously characterized mycobacterial -35 boxes.

No regulation of the *purC* gene expression could be demonstrated when the *purC* promoter region was fused to the *E. coli lacZ* gene on the shuttle plasmid pJEM12 and expressed in *M. smegmatis*, in the presence or absence of purine bases. Possibly *purC* is deregulated when carried on a plasmid, or perhaps the *purC* promoter region segment fused to β -galactosidase is too short to contain the regulatory region in its entirety. This could explain its very high activity compared to the strong promoter of the *M. fortuitum* β -lactamase gene. Regulatory mechanisms might also be different between *M. tuberculosis* H37Rv and *M. smegmatis*. Or, if the mechanisms are conserved between the two mycobacterial species, they might be insufficient in *M. smegmatis* to regulate the expression of *purC* on a multicopy plasmid such as pJEM12 (a pAL5000 derivative) (Martin *et al.*, 1990).

In *E. coli*, the purine biosynthetic genes are scattered throughout the chromosome. The expression of each gene of the pathway is repressed at the transcription initiation level by a common repressor which binds to a 16 bp palindromic sequence (5'-aCGCAAACGTTTtC-nT-3', where the lower-case letters signify less-conserved positions) present in the promoter region of each gene (Rolfes & Zalkin, 1990). In *B. subtilis*, the genes are organized as an operon. The expression of purine biosynthetic genes is regulated independently at the transcription initiation level by adenine and by a termination/anti-termination system of transcription by guanine. Adenine induces the fixation of a regulatory protein in the -35 region, thus blocking the access of the -35 box to RNA polymerase (Ebbole & Zalkin, 1989b). Guanine facilitates binding of a regulatory protein to

nascent mRNA at a site that prevents formation of the antiterminator (Ebbole & Zalkin, 1989a).

The *M. tuberculosis* purine biosynthetic genes identified here show no sequence similarities to the *B. subtilis* promoter region nor to the palindromic regulatory sequence in *E. coli*. Comparison of the upstream regions of the *M. tuberculosis* and *M. leprae* purine biosynthetic genes did not reveal any conserved motif, except for the *purL* gene promoter region, which seems to be conserved between these species, suggesting a common regulatory mechanism. These data raise the question of how purine nucleotide synthesis is regulated in *M. tuberculosis*. However, two sets of inverted repeats located between the -10 and -35 boxes and upstream of the -35 box of *purC* were detected. Although absent from the promoter region of *purL*, such structures could be the targets for a regulatory mechanism which remains to be determined.

This work represents the first steps towards the understanding of the purine biosynthetic pathways in *M. tuberculosis*. Its importance for the maintenance of *M. tuberculosis* in macrophages will be investigated in further genetic experiments using allelic exchange methods recently developed for *M. tuberculosis* complex strains (Reyrat *et al.*, 1995).

ACKNOWLEDGEMENTS

We are grateful to S. Cole for the kind gift of the *M. tuberculosis* H37Rv shuttle cosmid library. We would like to thank M. Debarbouillé and G. Rapoport for helpful discussions.

M. Jackson and F. X. Berthet are recipients of the Fondation Mérieux and of the Agence Nationale de la Recherche sur le SIDA (ANRS), respectively. This work was supported by a European Economic Community (EEC) Biotech program grant (BIO-CT92-0520).

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. M. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Andersen, P. (1994). Effective vaccination of mice against *Mycobacterium tuberculosis* infection with a soluble mixture of secreted mycobacterial proteins. *Infect Immun* **62**, 2536–2544.
- Averill, L. E., Cavallo, U., Wallis, R. S., Boom, W. H., Bona, M., Mincek, M., Pascopella, L., Jacobs, W. R. & Ellner, J. J. (1993). Screening of a cosmid library of *Mycobacterium bovis* BCG in *Mycobacterium smegmatis* for novel T-cell stimulatory antigens. *Res Microbiol* **144**, 349–362.
- Bachmann, B. J. (1987). Linkage map of *Escherichia coli* K-12. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 807–876. Edited by F. C. Neidhart and others. Washington, DC: American Society for Microbiology.
- Baselky, V. S., Medina, R. A. & Parker, C. D. (1979). *In vivo* and *in vitro* characterization of virulence deficient mutants of *Vibrio cholerae*. *Infect Immun* **24**, 111–116.
- Bashyam, M. D. & Tyagi, A. K. (1994). An efficient and high yielding method for isolation of RNA from mycobacteria. *Bio-Techniques* **17**, 834–836.
- Baulard, A., Jourdan, C., Mercenier, A. & Loch, C. (1992). Rapid mycobacterial plasmid analysis by electrotransformation between *Mycobacterium* spp. and *Escherichia coli*. *Nucleic Acids Res* **20**, 4105.
- Berthet, F. X., Rauzier, J., Lim, E. M., Philipp, W., Gicquel, B. & Portnoï, D. (1995). Characterization of the *M. tuberculosis* *erp* gene encoding a potential cell surface protein with repetitive structures. *Microbiology* **141**, 2123–2130.
- Bloom, B. R. & Murray, C. J. L. (1992). Tuberculosis: commentary on a reemergent killer. *Science* **257**, 1055–1064.
- Clark-Curtiss, J. E. (1990). Genome structure of mycobacteria. In *Molecular Biology of the Mycobacteria*, pp. 77–96. Edited by J. J. McFadden. London: Surrey University Press.
- Collins, F. M. (1993). Tuberculosis: the return of an old enemy. *Crit Rev Microbiol* **19**, 1–16.
- Curtiss, R. (1990). Attenuated *Salmonella* strains as live vectors for the expression of foreign antigens. In *New Generation Vaccines*, pp. 161–188. Edited by G. C. Woodrow & M. M. Levine. New York & Basel: Marcel Dekker.
- Dolin, P. J., Raviglione, M. C. & Kochi, A. (1994). Global tuberculosis incidence and mortality during 1990–2000. *Bull WHO* **72**, 213–220.
- Drazek, E. S., Hough, H. S. H., Crawford, R. M., Hadfield, T. L., Hoover, D. L. & Warren, R. L. (1995). Deletion of *purE* attenuates *Brucella melitensis* 16M for growth in human monocyte-derived macrophages. *Infect Immun* **63**, 3297–3301.
- Ebbole, D. J. & Zalkin, H. (1987). Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. *J Biol Chem* **262**, 8274–8287.
- Ebbole, D. J. & Zalkin, H. (1989a). *Bacillus subtilis* *pur* operon expression and regulation. *J Bacteriol* **171**, 2136–2141.
- Ebbole, D. J. & Zalkin, H. (1989b). Interaction of a putative repressor protein with an extended control region of the *Bacillus subtilis* *pur* operon. *J Biol Chem* **264**, 3553–3561.
- Fields, P. I., Swanson, R. V., Haidaris, C. G. & Heffron, F. (1986). Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *Proc Natl Acad Sci USA* **83**, 5189–5193.
- Guilhot, C., Gicquel, B. & Martin, C. (1992). Temperature-sensitive mutants of the *Mycobacterium* plasmid pAI.5000. *FEMS Microbiol Lett* **98**, 181–186.
- Guilhot, C., Otal, I., Van Rompaey, I., Martin, C. & Gicquel, B. (1994). Efficient transposition in mycobacteria: construction of *Mycobacterium smegmatis* insertional mutant libraries. *J Bacteriol* **176**, 535–539.
- Gupta, S. K. D., Bashyam, M. D. & Tyagi, A. K. (1993). Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. *J Bacteriol* **175**, 5186–5192.
- Hatfull, G. F. (1993). Genetic transformation of mycobacteria. *Trends Microbiol* **1**, 310–314.
- Hoiseth, S. K. & Stocker, B. A. D. (1981). Aromatic dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* **291**, 238–239.
- Husson, R. N., James, B. E. & Young, R. A. (1990). Gene replacement and expression of foreign DNA in mycobacteria. *J Bacteriol* **172**, 519–524.
- Jacobs, W. R., Tuckman, R. & Bloom, B. R. (1987). Introduction of foreign DNA into mycobacteria using a shuttle plasmid. *Nature* **327**, 532–535.
- Jacobs, W. R., Kalpana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G. & Bloom, B. R. (1991). Genetic systems for mycobacteria. *Methods Enzymol* **204**, 537–555.
- Ji, Y. E., Colston, M. J. & Cox, R. A. (1994). Nucleotide sequence and secondary structures of precursor 16S rRNA of slow-growing mycobacteria. *Microbiology* **140**, 123–132.
- Levine, M. M., Herrington, D., Murphy, J. R., Morris, J. G.,

- Losonsky, G., Tall, B., Lindberg, A. A., Svenson, S., Baqar, S., Edwards, M. F. & Stocker, B. (1987). Safety, infectivity, immunogenicity, and *in vivo* stability of two attenuated strains of *Salmonella typhi*, 541Ty and 543Ty, as live vaccines in humans. *J Clin Invest* **79**, 888–902.
- Martin, C., Ranes, M. & Gicquel, B. (1990). Plasmids, antibiotic resistance and mobile genetic elements in mycobacteria. In *Molecular Biology of the Mycobacteria*, pp. 121–138. Edited by J. J. McFadden. London: Surrey University Press.
- Myasnikov, A. N., Sasnauskas, K. V., Janulaitis, A. A. & Smirnov, M. N. (1991). The *Saccharomyces cerevisiae ADE1* gene: structure, overexpression and possible regulation by general amino acid control. *Gene* **109**, 143–147.
- Newman, J. D., Diebold, R. J., Schultz, B. W. & Noel, K. D. D. (1994). Infection of soybean and pea nodules by *Rhizobium* spp. purine auxotrophs in the presence of 5-aminoimidazole-4-carboxamide riboside. *J Bacteriol* **176**, 3286–3294.
- O'Callaghan, D., Maskell, D., Liew, F. Y., Easmon, C. S. F. & Dougan, G. (1988). Characterization of aromatic- and purine-dependent *Salmonella typhimurium*: attenuation, persistence, and ability to induce protective immunity in BALB/c mice. *Infect Immun* **56**, 419–423.
- Pardee, A. P., Jacob, F. & Monod, J. (1959). The genetic control and cytoplasmic expression of 'inducibility' in the synthesis of β -galactosidase and tryptophanase induction in *Escherichia coli*. *J Mol Biol* **1**, 656–664.
- Reyrat, J. M., Berthet, F. X. & Gicquel, B. (1995). The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guérin. *Proc Natl Acad Sci USA* **92**, 8768–8772.
- Rolfes, R. J. & Zalkin, H. (1990). Purification of the *Escherichia coli* purine regulon repressor and identification of corepressors. *J Bacteriol* **172**, 5637–5642.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanderson, K. E. & Hurley, J. A. (1987). Linkage map of *Salmonella typhimurium*. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 877–918. Edited by F. C. Neidhart and others. Washington, DC: American Society for Microbiology.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Sasnauskas, K., Jomantiene, R., Geneviciute, E., Januska, A. & Lebedys, J. (1991). Molecular cloning of the *Candida maltosa ADE1* gene. *Gene* **107**, 161–164.
- Schendel, F. J., Mueller, E., Stubbe, J., Schiau, A. & Smith, J. M. (1989). Formylglycinamide ribonucleotide synthetase from *Escherichia coli*: cloning, sequencing, overproduction, isolation and characterization. *Biochemistry* **28**, 2459–2471.
- Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. (1990). Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**, 1911–1919.
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.
- Steenken, W. & Gardner, L. U. (1946). History of H37 strain of tubercle bacillus. *Am Rev Tuberc* **54**, 62–66.
- Tiedemann, A. A., Demarini, D. J., Parker, J. & Smith, J. M. (1990). DNA sequence of the *purC* gene encoding 5'-phosphoribosyl-5-aminoimidazole-4-N-succinocarboxamide synthetase and organization of the *dapA-purC* region of *Escherichia coli* K-12. *J Bacteriol* **172**, 6035–6041.
- Timm, J., Lim, E. M. & Gicquel, B. (1994a). *Escherichia coli*–mycobacteria shuttle vectors for operon and gene fusions to *lacZ*: the pJEM series. *J Bacteriol* **176**, 6749–6753.
- Timm, J., Perilli, M. G., Duez, C., Trias, J., Orefici, G., Fattorini, L., Amicosante, G., Oratore, A., Joris, B., Frère, J. M., Pugsley, A. P. & Gicquel, B. (1994b). Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* β -lactamase genes cloned from a natural isolate and a high-level β -lactamase producer. *Mol Microbiol* **12**, 491–504.
- Young, D., Garbe, T., Lathigra, R. & Abou-Zeid, C. (1990). Protein antigens: structure, function and regulation. In *Molecular Biology of the Mycobacteria*, pp. 1–35. Edited by J. J. McFadden. London: Surrey University Press.

Received 23 February 1996; revised 1 May 1996; accepted 16 May 1996.