Determination of a 15437 bp nucleotide sequence around the *inhA* gene of *Mycobacterium avium* and similarity analysis of the products of putative ORFs

Monica Labò, Laura Gusberti, Edda De Rossi, Pietro Speziale and Giovanna Riccardi

Author for correspondence: Giovanna Riccardi. Tel: +39 382 505 574. Fax: +39 382 528 496. e-mail: riccardi@ipvgen.unipv.it

A 15437 bp region encompassing the *inhA* locus from the *Mycobacterium avium* chromosome was cloned and sequenced. From the sequencing data generated and the results of homology searches, the primary structure of this region was determined. This region contains four known genes (*acnA*, *fabG*, *inhA* and *hemH*) and two genes, *invA* and *invB*, whose products display homology with p60 invasion protein of *Listeria monocytogenes*. Six proteins encoded by putative ORFs contained an RGD motif (often involved in binding to macrophage integrins), while ORF1 and MoxR are probably transcriptional regulators. The rest of the putative products encoded by ORFs in the sequenced region showed little homology with the proteins contained in the databases and were considered to be unknown proteins.

**Keywords:** inv genes, *inhA* gene, *Mycobacterium avium*, nucleotide sequence

INTRODUCTION

Mycobacteria, members of which cause tuberculosis and leprosy, produce cell walls of unusually low permeability which contribute to their resistance to therapeutic agents. Their cell walls contain large amounts of C\(_{06}\)–C\(_{30}\) fatty acids, mycolic acids, that are covalently linked to arabinogalactan. Differences in mycolic acid structure may affect the fluidity and permeability of the bilayer and may explain the different sensitivity levels of various mycobacterial species to lipophilic inhibitors (Liu et al., 1996). There are several indications that some existing antimycobacterial drugs may indeed act on the cell wall biosynthetic pathways (Blanchard, 1996). Identification of the precise targets of such drugs may allow design of novel inhibitors of the same enzyme or of related steps in the same pathway.

Several mechanisms seem to be involved in isoniazid (INH) resistance in *Mycobacterium tuberculosis* (Deretic et al., 1996). The simplest hypothesis is that the drug is converted inside the bacterium to an activated form by the catalase-peroxidase enzyme, encoded by the *katG* gene (Zhang et al., 1993), and that the activated drug then inhibits the product of the *inhA* gene, an enzyme probably involved in mycolic acid biosynthesis (Banerjee et al., 1994). InhA has been shown to catalyse the \(\beta\)-NADH-specific reduction of 2-trans-enoyl-acyl carrier protein, an essential step in fatty acid elongation (Dessen et al., 1995). The *M. tuberculosis* *inhA* gene appears to be in an operon with the gene encoding the 29 kDa ORF that exhibits sequence similarity to several other proteins involved in fatty acid biosynthesis, such as the 3-ketoacyl-acyl carrier protein reductase (*fabG* gene product) from *Escherichia coli* (Banerjee et al., 1994).

The *Mycobacterium avium* complex is an intracellular pathogen that predominantly infects and multiplies within macrophages. *M. avium* infections are frequently present in patients with AIDS (Inderlied et al., 1993). Treatment of infections caused by *M. avium* still remains a challenge since these organisms are resistant to the majority of antituberculous drugs, presumably because of their impermeability to these agents. The bacterial cell envelope architecture could act as a barrier to some of the drugs (Brennan & Nikaido, 1995). While the complex cell wall structure of mycobacteria prob-
ably confers the permeability barrier that underlies their resistance to many existing antibacterial agents, the same unique structure contains a series of potential targets for novel Mycobacterium-specific inhibitors. The mycolic acids are unique to the mycobacteria and, again, it can be envisaged that their synthesis and assembly into the cell wall entail a series of enzymes, each representing a potentially attractive target for antibacterial action.

Since the inhA gene seems to be directly involved in mycolic acid biosynthesis, we have sequenced and analysed 15437 bp around this region to establish if other related genes are present in M. avium.

METHODS

Isolation of the inhA locus from a λGEM-11 library of M. avium. The Mycobacterium smegmatis inhA gene (kindly provided by Professor W. R. Jacobs, Albert Einstein College of Medicine, NY, USA) was used as probe to screen a λGEM-11 library of M. avium GIR10 (Milano et al., 1996). Several positive phages were identified and one of them, λGEM-33, was chosen to map the M. avium inhA gene. DNA fragments from λGEM-33 were subcloned into pGEM-4Z (Promega) for sequencing.

Procedures such as plaque hybridization and synthesis of radioactive DNA probes with random hexamer primers were carried out using standard protocols (Sambrook et al., 1989).

Genomic and λGEM-33 DNAs were digested with appropriate restriction endonucleases, fractionated on a 1 % agarose gel and transferred to a nylon membrane (Schleicher & Schuell) according to the manufacturer's instructions. Hybridization was performed at 42 °C with 50 % formamide; the blot was washed at 60 °C in 1 x SSC/0.1 % SDS.

To isolate and clone the corresponding inv genes from M. tuberculosis, we used the homologous PCR product as a probe to screen an M. tuberculosis cosmid library (kindly provided by Dr S. Cole, Institut Pasteur, Paris, France). Several positive cosmids were isolated and the M. tuberculosis inv genes were cloned and sequenced by standard procedures.

DNA sequencing and analysis. The M. avium DNA fragment from λGEM-33 was digested with several restriction enzymes and the resulting fragments were subcloned into pGEM-4Z. Sequencing was performed by the dideoxy chain-termination method (Sanger et al., 1977), using the Pharmacia T7 Polymerase Sequencing kit according to the supplier's instructions with [35S]dATP as the radioactive label. The sequence was determined on both strands and all restriction sites used to subclone into pGEM-4Z were overlapped. The subclones were sequenced using the SP6 and T7 promoter primers for pGEM-4Z as well as synthetic oligonucleotide primers designed from the resulting sequence.

DNA and protein sequences were compared with non-redundant databases at the National Center for Biotechnology Information (NCBI), using the BLAST programs (Altschul et al., 1990, 1994). The putative ORFs were identified both with the NCBI ORF-FINDER and with the PC/GENE programs. Putative ORFs in the region were sought in all six possible translation frames. Further analysis was performed using the PROSITE catalogue (Bairoch, 1992). Calculations of the isoelectric point and of the molecular mass were determined by COMPUTE PI/Mw (Bjellqvist et al., 1993, 1994). Multiple alignment was performed with the CLUSTAL W program (Thompson et al., 1994).

PCR procedures. DNA for PCR analysis, from different mycobacterial species, was extracted by the following procedure. A loopful of organisms from a colony was suspended in 1 ml H2O and boiled for 10 min. Samples (10 μl) were then used directly in PCR reactions using the primers 5'-GGGGGTGCCGTACTCGTGGGG-3' (upper) and 5'-ACCCGCCGGGCGCTAGAAC-3' (lower).

RESULTS AND DISCUSSION

We used the M. smegmatis inhA gene (Banerjee et al., 1994) as a probe to isolate the corresponding gene from a λGEM-11 library of M. avium (Milano et al., 1996). Several positive recombinants were selected and one of them (λGEM-33) was further studied. This recombinant phage carried a total insert of 15437 bp and its physical map is shown in Fig. 1. Hybridization experiments revealed a fragment whose size was consistent with that predicted by the DNA sequence of the λ-insert (data not shown).

ORFs present in the resultant 15437 bp segment, which has a G + C content of 69 mol %, typical of M. avium, were identified on the basis of their codon usage and database search similarities (Tables 1 and 2). The end points of the ORFs in the sequence, the size of the protein encoded, isoelectric point, putative Shine-Dalgarno sequence and initiation codon are shown in Table 1. Many of the genes appear to comprise independent monocistronic transcriptional units although there are some discernible operons. Similarities of predicted ORF products with database protein sequences and matches to PROSITE motifs (Table 2) led to the identification of many of the genes and their biological functions.

Of our sequence, 13170 bp in total was found to encode protein. The remaining 2267 bp seems to be non-coding. The largest non-coding stretch, 1506 bp (from nt 3167 to nt 4673), was without apparent structural features, as well as the 644 bp second non-coding stretch.

M. avium inhA gene

From positions 10566 to 11333 and from 11368 to 12174, we found, respectively, the fabG and inhA genes, as previously identified in M. tuberculosis (Banerjee et al., 1994). The predicted InhA protein of M. avium shows strong sequence similarity (85.5 % identity in a 269-residue overlap) to the corresponding protein of M. tuberculosis. A serine-94 → alanine-94 mutation of the inhA gene of M. smegmatis and of Mycobacterium bovis has been shown to cause resistance to INH and ethionamide (Banerjee et al., 1994) by decreasing the affinity of the mutant protein for NADH (Dessen et al., 1995). The structure of the NADH binding site of InhA suggests that recognition of NADH is mediated by interactions with an array of polar amino acids and backbone atoms (Dessen et al., 1995). The residues which are involved in the NADH-binding site of InhA...
The \textit{inhA} region of the \textit{M. avium} chromosome

\begin{figure}
\centering
\includegraphics[scale=0.5]{fig1}
\caption{Restriction map and ORF organization of the \textit{M. avium} chromosome around the \textit{inhA} gene. B, \textit{BamHI} restriction sites; S, \textit{SacI} restriction sites; \textbf{R}, RGD motif.}
\end{figure}

\section*{Table 1. Putative ORFs and characteristics of their products in the 15437 bp sequence around the \textit{inhA} gene of the \textit{M. avium} chromosome}

<table>
<thead>
<tr>
<th>ORF</th>
<th>End points (nt)</th>
<th>Molecular mass (kDa)</th>
<th>pI</th>
<th>SD* consensus sequence (upper case) and initiation codon (bold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf1</td>
<td>246 &lt; 1</td>
<td>Partial</td>
<td>Partial</td>
<td>ATGcccAGctAGATG</td>
</tr>
<tr>
<td>orf2</td>
<td>3167 &lt; 282</td>
<td>104.0</td>
<td>5.2</td>
<td>GAcGGccGAcGacATG</td>
</tr>
<tr>
<td>orf4</td>
<td>4673 &gt; 5494</td>
<td>28.8</td>
<td>9.6</td>
<td>AGGAcGccGcGcGTG</td>
</tr>
<tr>
<td>orf5</td>
<td>5520 &gt; 6254</td>
<td>25.4</td>
<td>9.6</td>
<td>AAcGAGGAGccATG</td>
</tr>
<tr>
<td>orf6</td>
<td>6376 &gt; 7305</td>
<td>34.2</td>
<td>10.9</td>
<td>GAGGAGGttATcGATG</td>
</tr>
<tr>
<td>orf8</td>
<td>7949 &gt; 8438</td>
<td>34.6</td>
<td>10.7</td>
<td>AAcGcGcGccATG</td>
</tr>
<tr>
<td>orf9</td>
<td>8466 &gt; 9473</td>
<td>36.0</td>
<td>9.4</td>
<td>GAcGtGAATATG</td>
</tr>
<tr>
<td>orf10</td>
<td>10626 &lt; 9586</td>
<td>39.5</td>
<td>10.3</td>
<td>GGGAGAcGAATG</td>
</tr>
<tr>
<td>orf11</td>
<td>9733 &gt; 10569</td>
<td>30.6</td>
<td>11.5</td>
<td>AGcGtcGcGcATG</td>
</tr>
<tr>
<td>orf12</td>
<td>11414 &lt; 10509</td>
<td>30.8</td>
<td>5.4</td>
<td>GAcGAGGcGGTG</td>
</tr>
<tr>
<td>orf13</td>
<td>10566 &gt; 11333</td>
<td>26.7</td>
<td>5.6</td>
<td>AGGtGAcGcGcGTG</td>
</tr>
<tr>
<td>orf14</td>
<td>11368 &lt; 12174</td>
<td>28.5</td>
<td>5.0</td>
<td>AAGGAcGcGcATG</td>
</tr>
<tr>
<td>orf15</td>
<td>12182 &gt; 13189</td>
<td>29.8</td>
<td>11.8</td>
<td>AAGGcGcGcATG</td>
</tr>
<tr>
<td>orf16</td>
<td>13193 &lt; 14155</td>
<td>33.4</td>
<td>5.5</td>
<td>cAAGAAttAcGcGT</td>
</tr>
<tr>
<td>orf17</td>
<td>13211 &gt; 14128</td>
<td>33.7</td>
<td>11.9</td>
<td>AGGcGcGcATG</td>
</tr>
<tr>
<td>orf18</td>
<td>14171 &gt; 14608</td>
<td>15.6</td>
<td>5.6</td>
<td>ttcGcGcGcATG</td>
</tr>
<tr>
<td>orf19</td>
<td>14639 &gt; 15437</td>
<td>Partial</td>
<td>Partial</td>
<td>AAAGGAAAttccATG</td>
</tr>
</tbody>
</table>

* SD, Shine–Dalgarno.

from \textit{M. tuberculosis} wild-type are unchanged in the \textit{M. avium} \textit{InhA} sequence; this finding rules out a possible involvement of \textit{M. avium} \textit{InhA} in the natural resistance of the micro-organism to INH.

In \textit{M. avium} and in \textit{M. tuberculosis} DNA, the non-coding region between the two ORFs is short and may lack a promoter; consequently, the two genes may be organized in an operon. Immediately downstream of the \textit{inhA} gene (from nt 12182 to nt 13189) there is orf15, whose gene product shows similarity with several ferrochelatase enzymes. The last step in haem synthesis is the insertion of iron into the ring of protoporphyrin IX. The enzyme which catalyses this reaction is a ferrochelatase, encoded by the \textit{Rhodobacter capsulatus} \textit{hemH} gene (Kanazireva & Biel, 1995). The degree of sequence homology of \textit{M. avium} ORF15 compared with human ferrochelatase (Crouse \textit{et al.}, 1996) is 34% identity in a 132-residue overlap. The same gene organization is present in \textit{M. tuberculosis} (accession no. gnl/Z79701) and the \textit{fabG–inhA–hemH} genes could be organized in an operon in both micro-organisms (Fig. 2).

\section*{invA and invB genes}

Little is known about the bacterial factors that enable pathogenic mycobacteria to survive and multiply within the macrophages of the infected host. \textit{M. tuberculosis} bacilli are thought to enter the macrophage via specific binding to complement receptors and the macrophage mannose receptor (Fenton & Vermeulen, 1996). The neural tropism of \textit{Mycobacterium leprae} has been ascribed to the specific binding of \textit{M. leprae} to the laminin-\textit{\alpha} \textit{LN-\alpha} chain on Schwann cell axon units. It has been suggested that \textit{M. leprae} interacts with cells by binding to \textit{\beta} \textit{LN-\beta} through an \textit{LN-\alpha} bridge (Rambukkana \textit{et al.}, 1997). Recently, it has been demonstrated that the \textit{M. avium} 68 kDa protein mediates attachment of \textit{M. avium} to macrophages, by interacting with \textit{\alpha} \textit{\beta} \textit{LN-\alpha} on these cells (Hayashi \textit{et al.}, 1997).
**Table 2. Similarity of the putative products of ORFs identified on \( \lambda \)GEM-33 *M. avium* insert**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Length (aa)</th>
<th>Sequence similarity and/or PROSITE match</th>
<th>BLAST score (P-value)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf1</td>
<td>82</td>
<td>AcrR from <em>E. coli</em> (sp/P34000)</td>
<td>2.4e-08</td>
</tr>
<tr>
<td>acn</td>
<td>961</td>
<td>Aconitase from: <em>E. coli</em> (sp/P25516), <em>L. pneumophila</em> (sp/P37032), <em>Brad. japonicum</em> (U56817), <em>Bac. subtilis</em> (Z73234)</td>
<td>0.0</td>
</tr>
<tr>
<td>invA</td>
<td>273</td>
<td>Hypothetical invasion protein similar to: <em>M. tuberculosis</em> ORF MTCY336.37 (gb/Z95586), <em>M. leprae</em> Lepb1170..F3..153 (U00010), <em>Listeria ivanovii</em> protein p60 (sp/Q01837), PS00016: RGD cell attachment sequence</td>
<td>8.1e-53</td>
</tr>
<tr>
<td>invB</td>
<td>244</td>
<td>Hypothetical invasion protein similar to: <em>M. tuberculosis</em> ORF MTCY336.37 (gb/Z95586), <em>M. leprae</em> Lepb1170..F3..153 (U00010), <em>Listeria ivanovii</em> protein p60 (sp/Q01837), PS00016: RGD cell attachment sequence</td>
<td>6.4e-48</td>
</tr>
<tr>
<td>moxR</td>
<td>309</td>
<td>Hypothetical transcriptional regulator similar to: <em>Bor. burgdorferi</em> moxR gene product (X96434), <em>Bac. subtilis</em> YeaC (U51115), <em>Synechocystis</em> sp. MoxR (D64003)</td>
<td>3.6e-64</td>
</tr>
<tr>
<td>orf8</td>
<td>169</td>
<td>Unknown, High similarity to <em>M. tuberculosis</em> partial reading frame MTCY227.01 (Z79701), PS00016: RGD cell attachment sequence</td>
<td>1.2e-49</td>
</tr>
<tr>
<td>orf9</td>
<td>335</td>
<td>Unknown, High similarity to <em>M. tuberculosis</em> MTCY227.02 (Z79701)</td>
<td>1.5e-207</td>
</tr>
<tr>
<td>orf10</td>
<td>346</td>
<td>Unknown, High similarity to <em>M. tuberculosis</em> MTCY227.03 (Z79701)</td>
<td>5.3e-151</td>
</tr>
<tr>
<td>orf11</td>
<td>278</td>
<td>Unknown, PS00016: three RGD cell attachment sequences</td>
<td>1.6e-59</td>
</tr>
<tr>
<td>orf12</td>
<td>301</td>
<td>Unknown, PS00016: two RGD cell attachment sequences</td>
<td>1.6e-58</td>
</tr>
<tr>
<td>fabG</td>
<td>255</td>
<td>Ketoacyl-reductase, High similarity to <em>M. tuberculosis</em> FabG (Z79701)</td>
<td>1.5e-50</td>
</tr>
<tr>
<td>inhA</td>
<td>268</td>
<td>Enoyl-reductase, High similarity to <em>M. tuberculosis</em> InhA (Z79701)</td>
<td>4.0e-154</td>
</tr>
<tr>
<td>hemH</td>
<td>335</td>
<td>Gene encoding ferrochelatase, High similarity to corresponding protein of <em>M. tuberculosis</em> (Z79701), PS00016: RGD cell attachment sequence</td>
<td>9.2e-93</td>
</tr>
<tr>
<td>orf16</td>
<td>321</td>
<td>Unknown, High similarity to <em>M. tuberculosis</em> MTCY227.07 (Z79701)</td>
<td>1.3e-154</td>
</tr>
<tr>
<td>orf17</td>
<td>305</td>
<td>Unknown, PS00016: two RGD cell attachment sequences</td>
<td>1.1e-67</td>
</tr>
<tr>
<td>orf18</td>
<td>145</td>
<td>Unknown, Similarity to <em>M. tuberculosis</em> MTCY227.08 (Z79701)</td>
<td>2.6e-134</td>
</tr>
<tr>
<td>orf19</td>
<td>266</td>
<td>Integral membrane protein, High similarity to <em>M. tuberculosis</em> MTCY227.09 (Z79701), Similarity to integral membrane protein from: <em>Methanococcus jannaschii</em> (U67526), <em>Synechocystis</em> sp. (D90899)</td>
<td>4.9e-50</td>
</tr>
</tbody>
</table>

*The P-value, calculated with the BLAST program, represents the probability that two proteins are related. The lower the P-value, the greater the degree of similarity.*
The *inhA* region of the *M. avium* chromosome

![Diagram of *M. avium* and *M. tuberculosis* genetic regions](image)

**Fig. 2.** Physical organization of *fabG-inhA-hemH* in *M. avium* and *M. tuberculosis.*

To evaluate the presence of *inv* genes within the mycobacterial genus, we used PCR, utilizing the upper and lower primers (see Methods) from *M. avium* *invB*. As shown in Fig. 3, a band of amplified DNA, with the appropriate size, appeared in all mycobacteria strains tested. This result is not surprising since, for example, the *iap* gene (encoding p60) has been found in several tested *Listeria* spp. (pathogenic and non-pathogenic) (Kohler et al., 1990).

The *M. tuberculosis* PCR product was used as a probe to screen a corresponding cosmid library. The isolated *M. tuberculosis* clone was mapped and 2360 bp was sequenced. Sequence analysis revealed the presence of two *M. tuberculosis* *inv* genes and a physical organization very similar to that of *M. avium* (data not shown).

**Fig. 3.** Detection by PCR of the *inv* genes from mycobacterial species. Lanes: 1, *M. simiae*; 2, *M. chelonae*; 3, *M. gordonae*; 4, *M. marinum*; 5, *M. bovis* BCG; 6, *M. flaveescens*; 7, *M. fortuitum*; 8, *M. kansasii*; 9, *M. xenopi*; 10, *M. terrae*; 11, *M. smegmatis*; 12, *M. abscessus*; 13, *M. avium*; and 14, *M. tuberculosis* H37Rv. The fainter PCR amplification product observed in *M. bovis* BCG was due to the quality of the sample, and was brighter in other experiments.

orf4 (from nt 4673 to nt 5494) and orf5 (from nt 5520 to nt 6254) show a low but detectable level of homology with p60, a protein that is involved in the uptake of *Listeria monocytogenes* by non-professional phagocytic cells (Bubert et al., 1992), and that is probably essential for survival inside the phagosome (Bhunia, 1997). Further, p60 is a murein hydrolase that is secreted into the host cell cytosol, where it is degraded by proteasomes (Sijts et al., 1997). The I*-values for ORF4 and ORF5 compared with the p60 protein were 9.1-17 and 4.3e-15, respectively. The highest score, 8.1e-53 for ORF4 and 6.3e-48 for ORF5, was observed with the ORF MTCY336.37 from *M. tuberculosis* (accession no. gb/Z95586). Similar scores, 2.2e-43 for ORF4 and 1.4e-45 for ORF5, were with Lepb1170_F3_153 from *M. leprae* (accession no. gi/U00010). As *M. avium* ORF4 and ORF5 show homology with p60, have an RGD motif and are 59.4% identical in a 219-residue overlap, we decided to name the corresponding genes *invA* and *invB*, respectively. These genes appear to be in an operon as they are separated by only 26 bp.

Fig. 4 shows the alignment of amino acid sequences of *M. avium* InvA, *M. avium* InvB, *M. tuberculosis* InvA, *M. tuberculosis* InvB, *M. leprae* Inv, *M. tuberculosis* ORF MTCY336.37 and C-terminal *L. monocytogenes* p60. As shown in Fig. 4, several amino acid blocks are conserved between Inv proteins and p60 at the C-terminal domain. It has been demonstrated that the p60 cysteine residue at position 39 is essential for the hydrolytic activity of the protein (Wuenscher et al., 1993). The same residue is conserved in *M. avium* and *M. tuberculosis* Inv proteins. As shown in Fig. 4, five proteins contain an RGD motif, which is common to many extracellular proteins and which is thought to play a key role in cell adhesion (Rouslahti & Pierschbacher, 1986, 1987). RGD sequences have also been identified in bacterial proteins and have been shown to interact with host receptors (Frankel et al., 1996; Everest et al., 1996). However, non-RGD-mediated interactions between bacterial ligands and their host cell integrin receptors have also been reported (Leong et al., 1990). Whether the *M. avium* and *M. tuberculosis* Inv proteins are really involved in adhesion–invasion processes remains to be determined.
MTCY336.37 is 70% (in a 197-residue overlap). The amino acid sequences of M. *tuberculosis* compared with that of M. *leprae* are much closer to M. *tuberculosis* (35.6% identity in a 218-residue overlap). Furthermore, the amino acid sequences of M. *tuberculosis* InvA and InvB, compared with M. *leprae* InvA and InvB, are 76% identity (in a 246-residue overlap) to M. *tuberculosis* InvA and InvB for potential membrane proteins. For this purpose, the amino acid sequences were scanned with TMbase (Hofmann & Stoffel, 1993). The results indicated two strong transmembrane helices for InvA and one strong transmembrane helix for InvB (data not shown).

There is an immediate need to understand the molecular basis of virulence and the pathogenesis of *M. tuberculosis* before the current molecular biological and immunological tools can be used in the prevention and control of the disease. Pathogens entering the human body must quickly adapt to changes in temperature, pH, osmotic strength, oxygen availability and nutrient concentration. Successful pathogens have evolved the ability to regulate expression of virulence genes.

**invA/invB** gene expression has been exploited in *M. tuberculosis* extracellular and intracellular growth. From preliminary data, it seems that *M. tuberculosis* invB expression is induced only in response to growth within the macrophage (unpublished data). Other experiments are in progress to evaluate the function of InvA and InvB in mycobacterial pathogenesis.

Downstream of *M. avium* orf5 (*invA*) is orf6 (from nt 6376 to nt 7305), whose product showed high similarity to MoxR of *Borrelia burgdorferi* (P-value 3e-64; accession no. gi/X96434), MoxR of *Synechocystis* sp. (P-value 1e-59; accession no. gi/D64063) and YeaC of *Bacillus subtilis* (P-value 3e-61; accession no. gi/U51115). MoxR seems to be involved in the regulation of formation of active methanol dehydrogenase in *Paracoccus denitrificans* (van Spanning et al., 1991). Another transcriptional regulator could be ORF1, whose 82 amino acids showed similarity (35.6% identity in a 59-residue overlap) to *E. coli* AcrR, a repressor that plays a modulating role in the regulation of *acrAB* genes of *E. coli* (Ma et al., 1996).

**Aconitase**

Aconitases are monomeric proteins containing single labile [4Fe-4S] clusters that possess a versatile molecular structure adapted to catalyse reactions with citrate, isopropylmalate and homocitrate and to function as iron-responsive regulators of mRNA translation or stability (Klausner & Rouault, 1993; Rouault & Klausner, 1996; Hentze & Kuhn, 1996). Crystallographic studies with porcine heart mitochondrial aconitase have revealed the presence of three structural domains (1, 2 and 3) tightly packed around the [Fe-S] cluster and a fourth domain (4) that is connected by a long linker peptide in such a way as to create a deep active-site cleft (Robbins & Stout, 1989). This arrangement provides the molecular basis of virulence and the pathogenesis of *M. tuberculosis* before the current molecular biological and immunological tools can be used in the prevention and control of the disease.

Pathogens entering the human body must quickly adapt to changes in temperature, pH, osmotic strength, oxygen availability and nutrient concentration. Successful pathogens have evolved the ability to regulate expression of virulence genes.

**invA/invB** gene expression has been exploited in *M. tuberculosis* extracellular and intracellular growth. From preliminary data, it seems that *M. tuberculosis* invB expression is induced only in response to growth within the macrophage (unpublished data). Other experiments are in progress to evaluate the function of InvA and InvB in mycobacterial pathogenesis.

Downstream of *M. avium* orf5 (*invA*) is orf6 (from nt 6376 to nt 7305), whose product showed high similarity to MoxR of *Borrelia burgdorferi* (P-value 3e-64; accession no. gi/X96434), MoxR of *Synechocystis* sp. (P-value 1e-59; accession no. gi/D64063) and YeaC of *Bacillus subtilis* (P-value 3e-61; accession no. gi/U51115). MoxR seems to be involved in the regulation of formation of active methanol dehydrogenase in *Paracoccus denitrificans* (van Spanning et al., 1991). Another transcriptional regulator could be ORF1, whose 82 amino acids showed similarity (35.6% identity in a 59-residue overlap) to *E. coli* AcrR, a repressor that plays a modulating role in the regulation of *acrAB* genes of *E. coli* (Ma et al., 1996).

**Aconitase**

Aconitases are monomeric proteins containing single labile [4Fe-4S] clusters that possess a versatile molecular structure adapted to catalyse reactions with citrate, isopropylmalate and homocitrate and to function as iron-responsive regulators of mRNA translation or stability (Klausner & Rouault, 1993; Rouault & Klausner, 1996; Hentze & Kuhn, 1996). Crystallographic studies with porcine heart mitochondrial aconitase have revealed the presence of three structural domains (1, 2 and 3) tightly packed around the [Fe-S] cluster and a fourth domain (4) that is connected by a long linker peptide in such a way as to create a deep active-site cleft (Robbins & Stout, 1989). This arrangement provides the molecular basis of virulence and the pathogenesis of *M. tuberculosis* before the current molecular biological and immunological tools can be used in the prevention and control of the disease.

Pathogens entering the human body must quickly adapt to changes in temperature, pH, osmotic strength, oxygen availability and nutrient concentration. Successful pathogens have evolved the ability to regulate expression of virulence genes.

**invA/invB** gene expression has been exploited in *M. tuberculosis* extracellular and intracellular growth. From preliminary data, it seems that *M. tuberculosis* invB expression is induced only in response to growth within the macrophage (unpublished data). Other experiments are in progress to evaluate the function of InvA and InvB in mycobacterial pathogenesis.

Downstream of *M. avium* orf5 (*invA*) is orf6 (from nt 6376 to nt 7305), whose product showed high similarity to MoxR of *Borrelia burgdorferi* (P-value 3e-64; accession no. gi/X96434), MoxR of *Synechocystis* sp. (P-value 1e-59; accession no. gi/D64063) and YeaC of *Bacillus subtilis* (P-value 3e-61; accession no. gi/U51115). MoxR seems to be involved in the regulation of formation of active methanol dehydrogenase in *Paracoccus denitrificans* (van Spanning et al., 1991). Another transcriptional regulator could be ORF1, whose 82 amino acids showed similarity (35.6% identity in a 59-residue overlap) to *E. coli* AcrR, a repressor that plays a modulating role in the regulation of *acrAB* genes of *E. coli* (Ma et al., 1996).

**Aconitase**

Aconitases are monomeric proteins containing single labile [4Fe-4S] clusters that possess a versatile molecular structure adapted to catalyse reactions with citrate, isopropylmalate and homocitrate and to function as iron-responsive regulators of mRNA translation or stability (Klausner & Rouault, 1993; Rouault & Klausner, 1996; Hentze & Kuhn, 1996). Crystallographic studies with porcine heart mitochondrial aconitase have revealed the presence of three structural domains (1, 2 and 3) tightly packed around the [Fe-S] cluster and a fourth domain (4) that is connected by a long linker peptide in such a way as to create a deep active-site cleft (Robbins & Stout, 1989). This arrangement provides the molecular basis of virulence and the pathogenesis of *M. tuberculosis* before the current molecular biological and immunological tools can be used in the prevention and control of the disease.

Pathogens entering the human body must quickly adapt to changes in temperature, pH, osmotic strength, oxygen availability and nutrient concentration. Successful pathogens have evolved the ability to regulate expression of virulence genes.
structural prototype for the aconitase family (Gruer et al., 1997a).

Only a few bacterial aconitases have been studied so far. Two distinct aconitases, AcnA and AcnB, were discovered in E. coli. AcnB is the major citric acid cycle enzyme, subjected to anaerobic and catabolite repression, whereas AcnA is induced by redox stress and by the presence of iron (Prodromou et al., 1991; Bradbury et al., 1996). Aconitase has also been characterized in Bac. subtilis (Dingman & Sonenshein, 1987), Legionella pneumophila (Mengaud & Horwitz, 1993) and Bradyrhizobium japonicum (Thony-Meyer & Kunzler, 1996). E. coli, L. pneumophila and Bac. subtilis AcnA proteins, together with the iron-regulatory proteins and plant aconitases, form an iron-regulatory protein/AcnA subfamily whose members are more closely related to each other than to any other members of the aconitase protein family (Gruer et al., 1997b).

M. avium orf2 starts at position 3167 (complementary strand) with an ATG codon that is preceded by a Shine–Dalgarno-like sequence (5'-CGACG-3') at a distance of 7 nt and ends at position 282 with a TGA stop codon. The derived polypeptide has a predicted molecular mass of 104 kDa and an isoelectric point of 5.27. The deduced amino acid sequence shows high similarity with E. coli aconitase AcnA (80% similarity, 54.1% identity) as well as with the bacterial aconitases from L. pneumophila, Brad. japonicum and Bac. subtilis. The similarity between the M. avium and Cucurbita maxima (pumpkin seed) aconitases (Hayashi et al., 1995) was in the range of 76.7% (47.6% identity). By contrast, M. avium aconitase is less related (34.6% identity) to porcine mitochondrial aconitase (Zheng et al., 1990). A good degree of similarity was also observed with eukaryotic iron-response-element-binding proteins (P-value 5e–305).

It is noteworthy that the cytoplasmic aconitase holoenzyme containing the [Fe–S] cluster is in equilibrium with its apoprotein, the iron-response-element-binding protein, which has a regulatory function in iron transport and storage. Hence the function of this protein depends on the availability of intracellular iron for the formation of the [Fe–S] cofactor (Klausner & Rouault, 1993). Crystallographic structure analysis and mutational analysis have led to the identification of 23 residues contributing to the active site (Lauble et al., 1992; Zheng et al., 1992). The 23 residues, which are involved in substrate recognition, [Fe–S] cluster ligation and interaction, catalysis or hydrogen binding that support the active-site side chains are also present in the M. avium AcnA sequence (data not shown).

**ACKNOWLEDGEMENTS**

This work was supported by a grant from the Istituto Superiore di Sanità, Progetto AIDS-1996, N. 940429, Roma.

We thank S. T. Cole for providing the M. tuberculosis cosmid library and for critical reading of the manuscript. We also thank W. R. Jacobs for providing the inhA probe.

**REFERENCES**


Frankel, G., Hershkovic, R., Mould, A. P., Kachalsky, S. G., Candy,


Received 21 July 1997; revised 1 November 1997; accepted 19 November 1997.