Cloning and sequencing of the gene which encodes the highly inducible acetamidase of *Mycobacterium smegmatis*

ESHWAR MAHENTHIRALINGAM,† PHILIP DRAPER,* ELAINE O. DAVIS and M. JOSEPH COLSTON

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

(Received 12 October 1992; revised 16 November 1992; accepted 23 November 1992)

The acetamidase of *Mycobacterium smegmatis* NCTC 8159 was purified, and the sequences of its amino-terminus and of two peptides obtained by proteolysis of the protein were obtained. A DNA fragment including the amidase structural gene was cloned in *Escherichia coli*, using oligonucleotide probes designed on the basis of the peptide sequences and a codon usage table calculated from published sequences of nine protein-antigen-encoding genes of the *Mycobacterium tuberculosis* complex. Sequence analysis of the cloned DNA revealed that the amidase gene encoded 406 amino acid residues. The nucleotide sequence close to and upstream of the amidase gene contained a probable ribosome-binding site but no identifiable promoter sequences. Three additional potential open-reading frames were found upstream of and very close to the amidase gene, with consensus ‘−35’ and ‘−10’ promoter sites between the first and second of these. It is hoped that the highly inducible expression of the acetamidase gene can be exploited to allow regulated expression of other genes cloned in mycobacteria.

**Introduction**

Disease caused by mycobacterial infections is a worldwide health problem. Leprosy and tuberculosis remain major diseases in the Third World while infections caused by *Mycobacterium tuberculosis* and members of the *M. avium-M. intracellulare* complex in individuals infected with human immunodeficiency virus are a serious health problem in many countries.

Over the last decade the study of the basic biology of mycobacterial pathogens has benefited greatly from a molecular biological approach. Gene transfer systems are now available for the manipulation of DNA in mycobacteria (Snapper *et al.*, 1988) which offer powerful means of investigating host–pathogen relationships in mycobacterial diseases. There is, however, little understanding of mechanisms of control of gene expression in mycobacteria at present. Better understanding of these mechanisms may give explanations of the comparatively slow growth rates of mycobacteria and their apparent ability to persist in a ‘resting’ phase, and may indicate the nature of the changes that occur in the mycobacterial cell after infection, during adaptation to life in vivo.

The availability of specific mycobacterial promoters, especially those which are readily regulated, would be useful for driving high-level, controllable expression of foreign genes in mycobacterial hosts. The acetamidase of *Mycobacterium smegmatis* NCTC 8159 was biochemically characterized and partially purified by Draper (1967). Assuming that control of expression of the amidase gene is at the transcriptional level, the promoter of this amidase is an excellent candidate regulatable mycobacterial promoter for experimental use. The enzyme has a high induction ratio, can be induced by non-substrate inducers, and is present in a rapidly-growing mycobacterial species. As a first step in isolating the promoter we have achieved further purification of the acetamidase of *M. smegmatis* NCTC 8159, the determination of partial amino acid sequences, the cloning of the acetamidase structural gene in *Escherichia coli* using oligonucleotide probes designed from these amino acid sequences, and the sequence analysis of the gene and adjacent upstream DNA where the promoter might be expected to occur.

* Author for correspondence. Tel. 081 959 3666; fax 081 906 4477.
† Present address: University of British Columbia, Faculty of Medicine, The Research Centre, 950 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada.

The nucleotide and amino acid sequences shown in Fig. 6 have been submitted to the EMBL Data Library and have been assigned the accession number X57175.
Methods

Bacterial strains and culture. Mycobacterium smegmatis NCTC 8159 was obtained from the National Collection of Type Cultures, Colindale, UK, and grown at 37 °C with shaking in Lenkom broth (Clarke & Meadow, 1959) containing 0.05% Tween 80 or minimal medium (Draper, 1967) containing 2 ml 1⁻¹ trace elements solution (Hopwood & Wright, 1978), 0.05% Tween 80, and carbon sources (acetamide, glutamate or succinate) at 2 g l⁻¹.

Escherichia coli strains NM522 (Gough & Murray, 1983) (supE, thi Δ(lac-proAB) hsdS (r' m') (F' proAB lacZAM15)) and JM101 (Messing et al., 1981) (supE thi Δ(lac-proAB) (F' traD36 proAB lacZAM15)) were grown in L-broth (LB) or 2×TY medium (Maniatis et al., 1982).

Preparation of cell-free extracts. Cell-free extracts were prepared from induced and uninduced M. smegmatis by sonication on ice of a 20% (wet weight) suspension of bacteria in 50 mM-Tris/HCl, pH 7.0, at 60 W for 15 x 1 min bursts with intervals of 1 min (Soniprobe; Dawe Instruments); phenylmethylsulphonyl fluoride was added to a final concentration of 1 mM immediately after cell disruption was completed. The soluble extracts were assayed for protein concentration by the Lowry method and analysed by SDS-PAGE (Laemmli, 1970).

Protein purification. The amidase was partially purified by ammonium sulphate precipitation as described previously (Draper, 1967) and the resulting crude extract (20 mg) fractionated by fast protein liquid chromatography (FPLC) using Mono S followed by Mono Q columns (Pharmacia) to obtain the purified protein. The system was operated at 1 ml min⁻¹, and linear gradients from zero to 1 M-NaCl were run over 20 min. The purified amidase was dialysed against 20 mM-Tris/HCl, pH 7.0, divided into aliquots and stored at −20 °C.

Acetamidase activity was assayed by the release of ammonia from acetamide, measured using glutamate dehydrogenase (Sigma diagnostic kit 170-A) according to the protocol provided with the reagents.

Prior to N-terminal amino acid sequence analysis the amidase was further purified by reverse-phase high-performance liquid chromatography (HPLC; Waters 600E system) using a C₄-Vydac column. The flow rate was 1 ml min⁻¹, and linear gradients over 40 min were run with water as the running buffer and acetonitrile as the elution buffer (both solvents contained 0.1% trifluoroacetic acid). Amidase was collected as a pure fraction at about 55% (v/v) acetonitrile.

Preparation and analysis of peptides. The amidase (100 μg) was denatured by heating at 100 °C for 2 min, then digested with 1 μg Staphylococcus aureus strain V8 protease (endoproteinase Glu-C; Sigma) in phosphate or ammonium carbonate buffers (Houmard & Draper, 1972) for 1 h at 25 °C in a total volume of 100 μl. Peptides were separated by SDS-PAGE on a 15% (w/v) polyacrylamide gel, transferred to polyvinylidene difluoride (PVDF) membranes using a semi-dry blotter (Millipore SDE system) and visualized by staining with Coomassie brilliant blue. Stained bands were excised from the gel, was present in cell-free extracts of induced cells; no such heavily staining band was apparent in extracts of cells grown in minimal medium with succinate or glutamate as carbon sources, or in complex medium. No other major differences were observed between these extracts.

The amidase was purified from crude extract by separation using a Mono S cation-exchange column (Pharmacia) with 50 mM-sodium acetate, pH 3.8, as the elution buffer. The amidase was not bound by the column under these conditions and was collected as an almost pure fraction shortly after the start of the run. Most other proteins were retained on the column until the NaCl concentration increased. The amidase fraction was then further purified using a Mono Q anion-exchange column (Pharmacia) with Tris/HCl, pH 7.0, as the elution buffer. The protein eluted as a pure fraction when the NaCl concentration reached 0.45 M. The level of purification achieved by each fractionation step is shown in Fig. 2(a), lanes A, N, S and Q. Acetamidase activity was present in the purified product, confirming DNA purification of acetamidase

DNA cloning. 3–5 kbp BamHI fragments of M. smegmatis DNA were isolated using a GeneClean kit (Bio 101) and cloned into pUC18 (Norrander et al., 1983). Competent cells were prepared by the CaCl₂ procedure (Cohen et al., 1972) and transformants were selected on LB-agar with 100 μg ampicillin ml⁻¹. The resultant ‘mini-library’ (about 1000 colonies) was screened with ³²P-labelled probe using a colony hybridization procedure (Grunstein & Hogness, 1975).

DNA sequencing. A 3 kbp HindIII fragment of the amidase clone containing the gene and about 1.5 kbp of upstream DNA was sequenced by a ‘shotgun’ DNA sequencing procedure (Bankier & Barrell, 1983) using random fragments cloned in the single-stranded bacteriophage vector M13mp10 (Messing, 1983) and the dideoxy chain-termination procedure (Sanger et al., 1977); the reactions were performed in a microtitre tray using reagents from the Sequenase Version 2 kit (US Biochemicals). Data from the random clones were assembled into a continuous sequence using the DB system programs (Staden, 1982). The complete sequence of both strands was determined.

Results

Analysis of cell-free extracts and purification of the acetamidase

The induction of acetamidase in M. smegmatis grown on acetamide is shown in Fig. 1. A massive band of a 47 kDa protein, much more conspicuous than any other band on the gel, was present in cell-free extracts of induced cells; no such heavily staining band was apparent in extracts of cells grown in minimal medium with succinate or glutamate as carbon sources, or in complex medium. No other major differences were observed between these extracts.

The acetamidase was purified from crude extract by separation using a Mono S cation-exchange column (Pharmacia) with 50 mM-sodium acetate, pH 3.8, as the elution buffer. The amidase was not bound by the column under these conditions and was collected as an almost pure fraction shortly after the start of the run. Most other proteins were retained on the column until the NaCl concentration increased. The amidase fraction was then further purified using a Mono Q anion-exchange column (Pharmacia) with Tris/HCl, pH 7.0, as the elution buffer. The protein eluted as a pure fraction when the NaCl concentration reached 0.45 M. The level of purification achieved by each fractionation step is shown in Fig. 2(a), lanes A, N, S and Q. Acetamidase activity was present in the purified product, confirming
M. smegmatis acetamidase gene

Fig. 1. Analysis of cell-free extracts of M. smegmatis grown on different carbon sources. Protein extracts (20 µg) were electrophoresed on an SDS-PAGE gel (12 %, w/v, acrylamide). The left-hand lane contains size markers as indicated; the remaining lanes are extracts of cells grown either in minimal medium containing acetamide (A), succinate (S) or glutamate (G), or in rich medium (Lemco broth, L). The massive increase of the 47 kDa acetamidase in cells grown on acetamide is evident; there are no other clearly visible differences between the four extracts. Two irrelevant lanes between the size markers and lane A have been blanked out.

that the 47 kDa protein was, indeed, the required enzyme.

Peptide sequences from the acetamidase and design of oligonucleotide probes

The native amidase was highly resistant to proteolysis (see Fig. 2a, lane T, which shows the lack of effect of overnight incubation with trypsin). Only nagarse and proteinase K degraded the native enzyme, but produced a large number of small peptides which were not useful for purification and sequencing. The amidase could be denatured by incubation at 100 °C for 2 min to allow proteolytic cleavage. Digestion with the sequence-specific Staphylococcus aureus strain V8 protease (endoproteinase Glu-C) in phosphate buffer (Houmard & Drapeau, 1972) yielded 10–15 peptides (Fig. 2b); few peptides were obtained using ammonium carbonate buffer. Three stretches of amino acid sequence were determined: the amino-terminus of the amidase and the derived peptides S36 and S24 (Fig. 2b); these are shown in Fig. 3. The yield of derivatized amino acids from the intact amidase was very low, suggesting that the N-terminus might normally be blocked and that the small amount of sample providing the sequence was a minority of unblocked amidase molecules.

The oligonucleotide probes were designed using a codon usage table assembled from the published sequences of nine cloned M. tuberculosis and M. bovis genes (Andersen & Hansen, 1989; Ashbridge et al., 1989; Baird et al., 1988; Borremans et al., 1989; Lathigra et al., 1988; Matsuo et al., 1988; Shinnick, 1987; Thole et al., 1987; Yamaguchi et al., 1989). Information was compiled from the sequence databases using sequence analysis programs (Devereux et al., 1984). DNAs of M. bovis and M. tuberculosis have almost the same molar G+C content as M. smegmatis (65–67 %; Bradley, 1973) and it was assumed that codon usage would be very similar in the three species. Two M. bovis BCG antigen-encoding genes had already been cloned (Matsuo et al., 1988; Yamaguchi et al., 1989) using single-sequence oligonucleotide probes derived from amino acid data, which were restricted to G or C in the third base positions.

Probe AM11 was a 41 base single-sequence oligonucleotide, incorporating the most frequently used codons encoding the first 14 amino acids of sequence AAS2 (Ala 2 to Ile 15), made as reverse-complement of the coding sequence:

5'-ATGTCCAGGATGTCGACGATCAGCAGGTCGCCC-GGCTCGGC-3'

Probe AM12 was a shorter (24 base) oligonucleotide of mixed sequence, incorporating both G and C at three sites most likely to be affected by the codon usage, and was derived from an eight amino acid stretch of AAS3 (Ala 2 to Asp 9) and made as the reverse-complement of the coding sequence:

5'-GTCCGGTGTCGAGCAGGGCCGAGG-3'

where S=G+C. Preliminary experiments with probes directed against the amidase N-terminus were unsuccessful (results not shown); this sequence was not used to design probes for subsequent experiments because of initial doubts over its authenticity.

Cloning the acetamidase gene

The results of the genomic Southern hybridization analysis with probes AM11 and AM12 are shown in Fig. 4. The hybridization pattern seen with probe AM11 was clear, with two major bands in each of the four restriction-endonuclease digestions of M. smegmatis DNA. Multiple banding was seen with probe AM12 (a
Fig. 2. SDS-PAGE gels showing purification of acetamidase and effect of protease digestion on native and denatured enzyme. Amidase-containing fractions or digests (approximately 10 µg protein) were analysed on 15% (w/v) acrylamide gels. Lanes are as follows. Both gels: M, molecular mass markers. Gel (a): A, cell-free extract of acetamide-grown *M. smegmatis*; N, soluble fraction after ammonium sulphate precipitation (crude amidase extract); T, crude amidase extract incubated overnight at 37°C with 0.1% trypsin; S, fraction not bound to Mono S column; Q, final pure amidase fraction eluted from Mono Q column. Gel (b): boiled amidase digested with endoprotease Glu-C at 25°C in ammonium carbonate (N) or in phosphate buffer (P); 0, material removed from digests immediately; 1, material removed after 1 h incubation. Peptides labelled S36 and S24 were used to design probes.

**AAS1**

\[\begin{align*}
\text{Pro} & \quad \text{Glu} & \quad \text{Val} & \quad \text{Val} & \quad \text{Phe} & \quad \text{Ser} & \quad \text{Val} & \quad \text{Asp} & \quad \text{His} & \quad \text{Ser} & \quad \text{Lys} & \quad \text{Ser} & \quad \text{Met} \\
& & & & & & & & & & & & \\
\text{Arg} & \quad \text{Asp} & \quad \text{Gln} & \quad \text{Ala} & \quad \text{Val} & \quad \text{Pro} & \quad \text{Gly} & \quad \text{His} & \quad \text{Asn} & \quad \text{Arg} \\
\end{align*}\]

**AAS2**

\[\begin{align*}
\text{(?)} & \quad \text{Ala} & \quad \text{Glu} & \quad \text{Pro} & \quad \text{Gly} & \quad \text{Asp} & \quad \text{Leu} & \quad \text{Leu} & \quad \text{Ile} & \quad \text{Val} & \quad \text{Asp} & \quad \text{Ile} & \quad \text{Leu} \\
& & & & & & & & & & & & \\
\text{Asp} & \quad \text{Ile} & \quad \text{Gly} & \quad \text{Ile} & \quad \text{Pro} & \quad \text{Ile} & \quad \text{Val} & \quad \text{Pro} & \quad \text{Gln} & \quad \text{Thr} \\
\end{align*}\]

**AAS3**

\[\begin{align*}
\text{(?)} & \quad \text{Ala} & \quad \text{Leu} & \quad \text{Ile} & \quad \text{Ala} & \quad \text{Thr} & \quad \text{Asp} & \quad \text{Pro} & \quad \text{Asp} & \quad \text{Arg} & \quad \text{(Val)} & \quad \text{(Val)} & \quad \text{Pro} \\
& & & & & & & & & & & & \\
\text{Pro} & \quad \text{Leu} & \quad \text{Ala} & \quad \text{Leu} \\
\end{align*}\]

Fig. 3. Amino acid sequence information from the amidase. Amino acid sequences of three stretches of the amidase are shown: the amidase N-terminal region AAS1 (obtained from HPLC-purified protein); sequence AAS2 derived from the second-largest amidase peptide S36; and sequence AAS3 originating from the third-largest peptide S24. The identities of the first amino acids of sequences AAS2 and AAS3 were not determined; amino acids whose identity was uncertain are shown in parentheses.

result of its mixed sequence design) and the signal intensity was weaker, probably due to its shorter length. Both probes recognized a band at about 4 kbp in a *BamHI* digest. Therefore *BamHI* fragments of about this size were cloned into pUC18, and 12 recombinant clones out of the 1000 colonies screened were found to be recognized by both oligonucleotide probes. One of these clones, designated pAM11, was selected for further study.

The position and orientation of the amidase gene were determined by hybridization analysis of restriction-digested fragments of pAM11 with probes AM11 and AM12, and are shown diagrammatically in Fig. 5. Digestion of the insert with *SalI* prevented probe AM11 hybridizing with the DNA; probe AM12 bound to a 600 bp fragment (data not shown). Probe AM11 encoded a recognition site for *SalI* (GTCGAC) near the middle of its sequence, so it appeared that the cloned DNA contained a corresponding *SalI* site. Therefore the *SalI* fragment recognized by AM12 was subcloned into pUC18 and sequenced from each end; the resulting sequence data obtained from one end of the insert matched probe AM11 from the *SalI* site, and translated correctly into the amino acid sequence AAS2 from the valine residue (position 10) onwards, confirming that the cloned DNA encoded the amidase.

**Nucleotide sequence**

A total of 2793 bp of nucleotide sequence of *M. smegmatis* DNA was obtained from a *HindIII* fragment
M. smegmatis acetamidase gene

Fig. 4. Southern blots of restriction-enzyme-digested M. smegmatis DNA hybridized with oligonucleotide probes AM11 and AM12. Lanes show M. smegmatis genomic DNA, digested to completion with the following restriction endonucleases: B, BamHI; Bg, BglII; E, EcoRI; K, KpnI. Blot (a) was hybridized with probe AM11, blot (b) with AM12. The stringency of hybridization and washing is given in Methods. Lanes M show 32P-end-labelled HindIII-digested λ DNA fragments to provide molecular size markers.

Fig. 5. Partial restriction map of the 4 kbp BamHI insert of pAM11. Restriction sites indicated are as follows: B, BamHI; E, EcoRI; H, HindIII; P, PstI; Sa, SstI; Sm, SmaI; Ss, SstI. Hybridization sites of AM11 and AM12 and the derived orientation of the acetamidase gene with respect to the lacZ gene of the vector are also indicated. The 3 kbp fragment between the HindIII (H*) sites was the region chosen for sequence analysis.

The base content of the fragment was as follows: 16.7% A, 17.3% T, 34.2% C and 31.8% G. The molar G+C content, 66%, was consistent with the global G+C ratio already determined for M. smegmatis DNA (Bradley, 1973). The oligonucleotide probes used were well matched: AM11 contained only two mismatched nucleotides within the 41 base sequence it recognized, while probe AM12 matched 23 out of 24 bases at its recognition site. The open-reading frame (ORF) of the amidase gene was easily found because amino acid sequences from three regions of the protein had been determined during the study, and could be used as markers of correct translation along the amidase encoding sequence. The amidase ORF was 1218 bp in length, with the stop codon 52 bp from one end of the BamHI fragment cloned.

The sequence of the coding strand, including 1521 bp of upstream sequence, and the protein translation for the amidase are shown in Fig. 6. The N-terminal amino acid sequence of the amidase was the same as sequence AAS1; the translation start codon was ATG, but the methionine residue encoded must have been removed from the native amidase as the residue was not found during microsequence analysis of the protein. The
amidase ORF also translated correctly into the two other amino acid sequences which had been determined, AAS2 and AAS3; these sequences began 81 and 180 amino acid residues, respectively, from the N-terminus of the protein. The ORF continued for a further 206 amino acids after the sequence of AAS3. The encoded amidase was 406 amino acids long and its molecular mass was calculated as 43965 Da. The sequence did not have any notable homology with published sequences available in the databases, and in particular did not resemble the inducible amidase of Pseudomonas aeruginosa (Brammar et al., 1987), a physiologically rather similar enzyme.

Discussion

The mechanisms which mycobacteria use to regulate gene expression are not well understood. Most mycobacterial promoters function poorly, if at all, in E. coli (Clark-Curtiss et al., 1985; Kieser et al., 1986), which suggests that mycobacteria use somewhat different systems of regulation. We have chosen to study the amidase-coding gene of M. smegmatis as a model of gene regulation in mycobacteria and as a potentially useful promoter for achieving overexpression of genes in these organisms.

In this report we present the cloning and complete nucleotide sequence of the gene, plus 1521 bp of upstream sequence where regulatory elements for the gene might be expected to occur. Directly upstream of the amidase gene was a recognizable purine-rich Shine-Dalgarno sequence (AAAGGGAG; see Fig. 6). This putative ribosome-binding site was 5 bases from the translation initiation codon ATG, a reasonable distance for its function. No sequences with obvious homology to ‘ – 35’ and ‘ – 10’ E. coli consensus promoter sequences were present immediately upstream of the amidase ORF, consistent with the lack of expression of amidase by E. coli clones carrying pAM1 as assayed by immunoblot analysis using an amidase antiserum (results not shown). In streptomycetes, promoters are quite diverse (Strohl, 1992), some resembling the E. coli type but others having little similarity. If a similar situation applies to mycobacteria, the failure to identify an amidase promoter by inspection of the sequence immediately upstream of the structural gene does not imply that none is present. Two direct repeat sequences, TCACCGCG and GCGCCGTC, upstream of the amidase gene are shown in Fig. 6; direct repeats have been found at the inducible promoter site of the galactose operon in Streptomyces lividans, but their role in induction is unknown (Fornwald et al., 1987).

Analysis of the DNA sequence from pAM1 revealed three additional probable ORFs (P1, P2 and P3) (Fig. 7) upstream of the ORF of the amidase, all showing a ‘mycobacterial’ codon usage, and coding for polypeptides of 15-7, 10 and 22 kDa, respectively. Analysis of the reverse-complementary strand did not show any substantial ORFs. All these ORFs possessed purine-rich ribosome-binding sites at appropriate distances from their putative ATG initiation codons. The spacing of ORF P2, which ends 8 bp from ORF P3, and ORF P3, which ends 20 bp from the amidase ORF, suggests that these genes may form part of an operon with the amidase. Further, appropriately spaced sequences almost identical to those of E. coli consensus promoters (Harley & Reynolds, 1987) were found upstream of and very close to the second (P2) of these ORFs (Fig. 6).

None of the putative polypeptides showed any close homology with published protein sequences (University of Leeds OWL composite protein sequence database, release 18.0, 9 October 1992), but ORF P2 contained two examples of the motif CXXC which is found in several metal-binding proteins, including some involved in regulation of gene expression in eukaryotes. Apart from this, none of the ORFs possessed conserved motifs associated with DNA-binding proteins (Pabo & Sauer, 1984). ORF P3 showed some homologies with a number of membrane-associated proteins; this may merely reflect its high content of hydrophobic amino acid residues (31% Ile + Leu + Val). However, no inducible polypeptides other than the amidase were seen in gels of cell-free extracts and any role the other ORFs may have in the expression of the amidase has yet to be elucidated.

The codon usage of the amidase gene closely matched the frequencies indicated by the cumulative table of codon usage assembled from the published M. tuberculosis and M. bovis gene sequences. The G or C bias in third-base position of the codons was in most cases the same as that indicated by the analysis of the published

Fig. 6. Sequence of the amidase gene of M. smegmatis. The complete nucleotide sequence of the amidase gene and 1521 bases of upstream DNA is shown. The protein translation of the gene is given above the nucleotide sequence, with the three regions of directly determined amino acid sequence described in Fig. 3 underlined. The potential Shine–Dalgarno sequence AAAGGGAG, preceding the ATG start codon, is underlined. Two pairs of direct repeat sequences, TCACCGCG and GCGCCGTC, are also indicated. Also shown (in bold) are the start and stop codons of the three other potential ORFs (see Fig. 7). The appropriately spaced E. coli consensus ‘ – 35’ and ‘ – 10’ sequences (Harley & Reynolds, 1987) preceding ORF P2 are indicated; an asterisk marks the predicted start site for transcription.
Fig. 7. ORF analysis of sequenced M. smegmatis DNA. Graphics output from the codon preference program (Gribskov et al., 1984) applied to the 2793 bp of M. smegmatis DNA sequenced, including the amidase structural gene, is shown. Use was made of the calculated cumulative codon-usage table described in the text. Analysis for each possible forward reading frame of the DNA strand coding the amidase is shown. The traces indicate the third-position G+C bias. Positions of rare codons are indicated by vertical hatches along the axes. Potential ORFs beginning with the start codon (ATG) are shown as open blocks beneath the traces for each reading frame; ORFs P1, P2 and P3 (and the amidase gene itself) show an appropriate codon preference and third-position base selection for mycobacterial genes.

sequences from other high G+C mycobacteria. However, some of the codon usages for the amidase were interestingly different from the predicted preferences. The codon used for isoleucine in the amidase (occurring 24 times) was exclusively ATC; from the cumulative analysis the codon ATT would have been expected to have been used about 20% of the time, so its absence would appear to be significant, taking into consideration the high occurrence of isoleucine in the amidase.

Of the mycobacterial gene sequences so far published, only those for rRNA contain groups of bases identified unequivocally as homologous to known E. coli consensus promoter sequences positioned at, or close to, the appropriate −10 and −35 regions (Sela & Clark-Curtiss, 1991; Suzuki et al., 1991). Apart from these genes, whose promoters are recognized by E. coli, the few genes so far identified which can be expressed from their own promoter sequences in E. coli are those for mycobacterial heat-shock proteins (Lamb et al., 1988; Shinnick, 1987; Thole et al., 1987), which are highly conserved among all organisms and so probably not representative of most mycobacterial genes. The cloning of the amidase gene reported here should permit the eventual isolation of a mycobacterial promoter with defined and experimentally useful properties.

We are grateful to A. Harris for carrying out the peptide sequencing and to P. Jenner for advice on protein purification. E. M. was supported by a Postgraduate Research Studentship from the Medical Research Council of Great Britain.

References


