Citrate synthase from *Mycobacterium smegmatis*

Cloning, sequence determination and expression in *Escherichia coli*

Michael DAVID,* Sharon LUBINSKY-MINK,* Aviva BEN-ZVI,* Mordechai SUISSA,* Shimon ULITZUR*† and Jonathan KUHN*‡

* Biolume Ltd., P.O. Box 9231, Haifa 31092, † Department of Food Engineering and Biotechnology, Technion–Israel Institute of Technology, Haifa 32000, and ‡ Faculty of Biology, Technion–Institute of Technology, Haifa 32000, Israel

A *Mycobacterium smegmatis* *Pst*I library was constructed by cloning these fragments downstream from the *lac* promoter of the expression vector pHG171. Three identically sized clones were isolated by complementation of an *Escherichia coli* strain (χ 2338) deficient in citrate synthase. One insert (pBL265) was used in hybridization experiments with DNA from *E. coli* and *M. smegmatis* and it was demonstrated that the clones were indeed from *M. smegmatis*. The transcription of the *M. smegmatis* citrate synthase gene in *E. coli* relied upon the *lac* promoter. In translation experiments performed *in vitro* pBL265 gave rise to a novel protein of about 42 kDa. This band was not seen in 'opposite-orientation' subclones. Various subclones in which the 5'-end was shortened nevertheless complement *E. coli* χ 2338 and produce the 42 kDa protein. This demonstrates that the *M. smegmatis* citrate synthase gene uses its own ribosome-binding site in *E. coli*. The relevant 1.8 kb of the 2.8 kb insert was sequenced. A consensus *E. coli* ribosome-binding site was found centred precisely 10 bp upstream of the methionine codon. Other interesting features revealed by the sequence are discussed. Citrate synthase activity was assayed *in vitro* and the mycobacterial enzyme was found to be similar to those of the Gram-positive bacteria.

INTRODUCTION

Mycobacteria are responsible for a number of diseases in man, and these include leprosy and tuberculosis. Members of this genus are also important opportunist pathogens in AIDS patients. Recently, tremendous progress has been made in elucidating the molecular biology of these organisms. A number of genes have been cloned and sequenced and their putative ribosome-binding sites and promoters have been indentified [1–6]. Our initial efforts to study the molecular genetics of this genus have concentrated on *Mycobacterium smegmatis*, which is a relatively fast-growing species and only a minor pathogen of man [7]. This organism is being used in a number of laboratories as a model system for the development of gene-transfer systems in this genus.

The citrate synthase gene from *Mycobacterium leprae* has been cloned and identified by its ability to complement a *gltA* mutant strain of *Escherichia coli* [8]. The corresponding gene from *M. smegmatis* should prove useful as a marker in genetic and evolutionary studies of these G+C-rich organisms. The present paper describes the cloning and expression of this gene in *E. coli*. The DNA sequence has been elucidated. There was no hybridization between the citrate synthase genes of *M. leprae* and *M. smegmatis*, in accord with previous studies that showed that *M. leprae* hybridizes only weakly to other mycobacterial species [9,10]. As found for other genes [11], the high G+C content of the coding region is maintained by excessive use of G+C at the third (wobble) base of the codons. The enzyme from *M. smegmatis* is similar to those from Gram-positive bacteria, and its sequence similarity to the enzymes from other organisms is compared.

EXPERIMENTAL

E. coli strains

E. coli χ 2338 (*gltA* and *hsdR*, among others [8]) was kindly given by J. Curtiss. *E. coli* JM101 [12] and *E. coli* DH5 [13] have

been described in the references cited. *E. coli* NM522 is a K12 strain.

Plasmids

Plasmids pHG165 and pHG171 have been described [14]. Plasmid pDB2 [15] was received from D. Bloxman and plasmid pYA1036 [8] from J. Curtiss. pKK232-8 is a promoter probe plasmid [16].

DNA

M. smegmatis DNA from strain A.T.C.C. 607 [17] was kindly given by H. Bercovier.

Enzymes

Restriction endonucleases, T4 DNA ligase, DNA polymerase I (Klenow fragment) and *Bal*31 were obtained from Biolabs, I.B.I. and Boehringer Mannheim and were used according to the manufacturers' instructions. Calf intestinal alkaline phosphatase was from Boehringer Mannheim and was used as described by Maniatis *et al.* [18].

Media

Luria broth has been described [19]. M9 minimal medium [19] also contained diaminopimelic acid (100 μ g/ml), thymidine (40 μ g/ml), tryptophan (20 μ g/ml), shikimic acid (60 μ g/ml) and glutamic acid (100 μ g/ml). LDTT medium is Luria broth containing diaminopimelic acid (100 μ g/ml), tryptophan (20 μ g/ml) and thymidine (40 μ g/ml).

Southern blotting

Blotting was performed on Amersham International Hybond-N nitrocellulose paper as advised by the manufacturer. Radioactive labelling of DNA fragments was done according to the method of Feinberg & Vogelstein [20] with Amersham International multiprimer labelling kit.

† To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases.

Transformation of *E. coli*

Protocol no. 1 of Hanahan [13] was used to prepare competent cells. The methods for detecting inserts in pHG165 and pHG171 have been described [14]. Clones producing *M. smegmatis* citrate synthase were detected in *E. coli* strain χ 2338 with the use of a selective medium [M9 lacking glutamic acid and containing diaminopimelic acid, thymidine, tryptophan, shikimic acid, ampicillin and 2% (w/v) glucose or lactose].

Transcription-translation of plasmid DNA *in vitro*

Amersham International DNA-directed transcription-translation kit was used. [³⁵S]Methionine-labelled proteins were analysed by SDS/PAGE [21] in gels containing 10% (w/v) polyacrylamide.

Bal31-derived subclones of pBL265

pBL265 was cut with *Hind*III and digested with 1 unit of *Bal*31 (slow)/ μ g of DNA. Samples containing approx. 4 μ g of DNA were removed after digestion for 50, 65, 80, 95 and 120 min and the *Bal*31 reaction was stopped with EGTA. A sample was analysed by 0.7% agarose-gel electrophoresis. The 80 min, 95 min and 120 min samples were considered to be overly digested and were discarded. The DNA of the 50 min and 65 min samples was extracted with phenol/chloroform (1:1, v/v) and precipitated with ethanol. The DNA was then cut with *Hpa*I and *Xmn*I. The ends produced by *Bal*31 were filled in by DNA polymerase I (Klenow fragment) [18]. The fragments produced by *Hpa*I-*Bal*31 digestions were purified by electrophoresis on agarose gels and extracted by the NaI/glass bead method, according to the GeneClean procedure of Bio101 Inc. pHG171 was digested with *Sma*I, treated with calf intestinal alkaline phosphatase and gel-purified as above. The *Bal*31-derived DNA fragments were ligated to pHG171 and transformed into *E. coli* JM101 as described above. The plasmid DNA of a few of the white colonies on MacConkey/AMP plates was analysed by *Sma*I-*Hind*III digestion and agarose-gel electrophoresis. The eight shortest inserts from the 50 min digestion and the two shortest inserts from the 65 min digestion were screened for proper orientation with respect to the *lac* promoter. Large-scale preparations of plasmid DNA were performed by the procedure of Birnboim & Doly [22] and purified by ethidium bromide/CsCl-equilibrium-density-gradient centrifugation [18]. The preparation of small amounts of plasmid DNA was done by the GeneClean method of Bio101 Inc.

Determination of promoter activity

pKK232-8 contains a promoterless chloramphenicol acetyltransferase (*CAT*) gene whose expression confers resistance to chloramphenicol. pBL265 was digested by *Sma*I and *Hind*III and the ends generated by the latter were filled. These fragments were inserted into pKK232-8 via the *Sma*I site, which lies upstream of the *CAT* gene. *E. coli* DH5 was transformed and plated on either CM plates (30 μ g of chloramphenicol/ml) to check for promoter activity or AMP plates (50 μ g of ampicillin/ml) to isolate insertions.

Two individual colonies from AMP plates that were shown to contain the *Hind*III-*Sma*I fragment in the proper orientation, as well as one colony that was obtained from a control transformation with pKK232-8, were grown overnight in Luria broth containing ampicillin (50 μ g/ml) and were tested in duplicate by a disc antibiotic assay using 13 mm Whatman AA discs preloaded with 100, 200, 400 or 600 μ g of chloramphenicol.

DNA sequencing of the citrate synthase gene

Sequencing was done by the dideoxy chain-termination method

[23]. Cloning into M13mp18 and M13mp19 and bacteriophage DNA preparation were performed according to the Biolabs M13 cloning manual. The Sequenase kit (U.S.B.) was used as suggested by the manufacturer. [γ -³⁵S]Thio[dATP] (> 1000 Ci/mmol) was from Amersham International. Reactions with both dGTP and dITP were run in parallel on the same gel. The entire 1800 bp *Eco*RI-*Xho*I fragment was sequenced on both strands with the exception of about 250 bp near the *Eco*RI site, which was sequenced on only one strand.

Citrate synthase activity *in vitro*

The citrate synthase [citrate oxaloacetate-lyase (CoA-acetylating), EC 4.1.3.7] activity was measured by the procedure of Srere [24]. Sonicated cell extract was centrifuged at 47800 *g* for 1 h in a Sorvall RC5C centrifuge with an SS34 rotor to remove debris. The assay mixture was incubated for 2 min in the absence of oxaloacetate to reduce background and the reaction was then started by the addition of oxaloacetate. A_{412} was recorded over the first few minutes and the increase in absorbance was nearly linear over this time period. Protein concentration was measured by the procedure of Lowry *et al.* [25] with BSA (Sigma Chemical Co.) as a standard.

RESULTS

Evidence that a *Pst*I library of *M. smegmatis* contains a gene that complements a citrate synthase-deficient *E. coli*

An *M. smegmatis* library was constructed by total cleavage with *Pst*I and ligation to pHG171 that had been treated with *Pst*I and calf intestinal alkaline phosphatase. pHG171 is similar to pUC8 but has a copy number similar to pBR322 [14]. Foreign DNA cloned into the multiple cloning site of pHG171 can be expressed from the *lac* promoter, sometimes as a fusion protein with the α -complementation peptide of β -galactosidase. The number of *AMP*^R transformants after ligation was $4 \times 10^5/\mu$ g and about 90% of the transformants contained inserts, as estimated by planting a portion on MacConkey AMP plates. Another portion was used to transform *E. coli* χ 2338, and the cells were plated on minimal medium (M9) without glutamic acid

Fig. 1. Size and orientation analysis of an *M. smegmatis* insert that complemented *E. coli* χ 2338

DNA was prepared and cleaved as described in the Experimental section. Lanes A and H contain known size markers. Lanes B, D and E are the plasmid vector without an insert. Lanes C, F and G are the vector containing an insert. Lane G shows that the insert contains an *Eco*RI site proximal to the *lac* promoter. Lane A, λ DNA cleaved with *Hind*III; lane B, pHG171 uncut; lane C, pBL265 uncut; lane D, pHG171 cut with *Pst*I; lane E, pHG171 cut with *Eco*RI; lane F, pBL265 cut with *Pst*I; lane G, pBL265 cut with *Eco*RI; lane H, λ DNA cut with *Bst*EII.

Fig. 2. Evidence that the insert in pBL265 is from *M. smegmatis*

(a) Gel electrophoresis of *E. coli* and *M. smegmatis* DNA cut with *Pst*I. λ DNA cleaved with *Hind*III was used as size marker (M). (b) Southern blots of half of the gel from (a). The probe was the insert of pDB2 (*E. coli* citrate synthase). (c) As in (b) but the probe was the insert of pBL265.

and containing ampicillin to select for *gltA*-complementing clones as described in the Experimental section. Three colonies appeared after 2 days at 37 °C. The transformants grew marginally better on glucose than on lactose. All three had plasmids with a *Pst*I insert of about 2.8 kb (Fig. 1). One of them, pBL265, was used to transform *E. coli* χ 2338; the transformation efficiency on selective plates was $10^4/\mu\text{g}$ of DNA.

To establish that the insert in pBL265 is from *M. smegmatis*, and to determine the extent of hybridization between the citrate synthase genes of *E. coli* and *M. smegmatis*, Southern-blot transfer experiments were performed. The insert of pBL265 and the insert of pDB2 [15], which contains the *E. coli* gene for citrate synthase, were each hybridized to *E. coli* and *M. smegmatis* chromosomal DNA digested by *Pst*I (Fig. 2a). As shown in Fig. 2(b), there was no cross-hybridization between the insert of pDB2 and *M. smegmatis* DNA or between the insert of pBL265 and *E. coli* DNA (Fig. 2c). pBL265 strongly hybridized to a 2.8 kb *M. smegmatis* DNA fragment. It seems clear that pBL265 contains an *M. smegmatis* DNA fragment that complements mutations in the citrate synthase gene of *E. coli*. In other experiments (results not shown) the insert of pYA1036 [8] containing the *M. leprae* citrate synthase gene failed to hybridize to *M. smegmatis* DNA. pYA1036 has previously been shown to hybridize only weakly to two other mycobacterial species, *Mycobacterium vaccae* and *Mycobacterium lufu*, and not to *E. coli* [8].

Evidence that the *gltA* clone is transcribed from the *plac* promoter

Restriction-enzyme mapping of pBL265 (Fig. 3) allowed us to compare the orientation of the three clones that originally complemented *E. coli* χ 2338. Digestion of pBL265 by *Eco*RI revealed that all three inserts are in the same orientation with respect to the *lac* promoter. The insert was removed by *Pst*I digestion, purified and ligated back into pHG171, cleaved similarly. The ligation was transformed into *E. coli* JM101, and plated on MacConkey AMP plates. Among the resulting white

Fig. 3. Restriction-enzyme map of pBL265

colonies, seven were isolated that had the insert in the opposite orientation (e.g. pBL266) to that of pBL265. Transformation with the DNA from these seven 'opposite-orientation' clones failed to complement the *gltA* mutation of *E. coli* χ 2338. These results show that expression of the mycobacterial *gltA* gene relies on transcription from a plasmid promoter, presumably the *lac* promoter. It cannot be excluded that there is also some tran-

Fig. 4. Transcription–translation *in vitro* of *M. smegmatis*-derived citrate synthase

A 2 µg portion of DNA was used in each experiment except for the ‘No DNA’ lane. About one-tenth of the reaction mixture was loaded in each case. The arrow indicates the novel protein seen from the translation of plasmids pBL265, pBL290, pBL292, pBL332 and pBL324 (see also Table 1). pHG171 is the vector by itself. Size markers are from [¹⁴C]methylated protein markers purchased from Amersham International.

Table 1. Summary of pBL265 and its subclones

Arrows in the *plac* columns indicate the location and direction of the promoter with respect to a given plasmid insert. The restriction map of the inserts is presented: P, *Pst*I; X, *Xho*I; B, *Bam*HI; E, *Eco*RI; Pv, *Pvu*II; S, *Sma*I; C, *Sca*I; H, *Hpa*I; K, blunt end produced by Klenow fragment after *Bal*31 treatment. Most pairs are the same insert in opposite orientations with respect to *plac* (e.g. pBL265 and pBL266) except pBL358 and pBL359, which both have the *plac* promoter to the left of the insert but differ by 50 bp. Dotted lines represent sequences outside the *M. smegmatis* insert that were used in subcloning. The last two columns summarize experiments described in the text.

Plasmid	<i>plac</i>	Map of <i>M. smegmatis</i> insert	<i>plac</i>	42 kDa protein	Complements <i>E. coli</i> χ2338
pBL265	→	P XBE Pv S C X P		+	+
pBL266			←	–	–
pBL288	→	P XBE Pv S		–	–
pBL287			←	–	–
pBL290	→	XBE Pv S C X		+	+
pBL289			←	–	–
pBL324	→	E Pv S C X PB		+	+
pBL323			←	–	–
pBL332	→	E Pv S C		+	±
pBL333			←	–	–
pBL358	→	K Pv S C X P H		+	±
pBL359	→	K Pv S C X P H		+	+

scription from within the inserts, but, if it occurs, the level of transcription is insufficient for complementation.

Evidence that no functional *E. coli*-like promoter is found upstream of the *M. smegmatis* *gl*A gene

The promoter probe plasmid pKK232–8 allows for the identification of putative promoters in DNA cloned upstream of

a promoterless *CAT* gene [16]. We have found that some pieces of *M. smegmatis* DNA can function as promoters in *E. coli* when cloned into pKK232–8. The 5′-end of the citrate synthase gene is distal to the *Hind*III site (position 3100 in pBL265), and the *Sma*I site (position 1200) is within the gene. The 1.9 kb *Hind*III–*Sma*I fragment when inserted in the correct orientation might contain a promoter but should not have a transcription termination signal. This fragment was cloned into pKK232–8. No colonies

were obtained from the transformation plated on CM plates. Plasmid DNA from a few transformants among those recovered on AMP plates was analysed and checked for orientation by *Bam*HI digestions. Inserts in both orientations were isolated. None of the individuals tested by the disc antibiotic assay method (see the Experimental section) appeared to be more resistant to chloramphenicol than the control, pKK232-8. There seems to be no *E. coli*-like promoter in this fragment. This could either be because the citrate synthase gene is part of an operon and its promoter is not included in the pBL265 insert or be because that the insert includes a promoter that is not recognized by *E. coli* RNA polymerase.

Evidence that translation of the *M. smegmatis gltA* clone relies on its own ribosome-binding site

Translation of the *gltA* gene could depend either on the ribosome-binding site of the α -peptide, resulting in a fusion protein with the first eight amino acid residues of β -galactosidase, or on the ribosome-binding site of the *M. smegmatis* citrate synthase gene. Experiments were carried out *in vitro* to examine the product of the *gltA* gene. The Amersham International transcription-translation coupled kit was incubated either in the presence or in the absence of 2 μ g of DNA from pHG171, pBL265 or pBL266. As shown in Fig. 4(a), pBL265 produces an additional protein band of about 42 kDa compared with pHG171. This band is not observed in pBL266. Although this experiment yields information about the size of citrate synthase, it cannot be excluded that this protein is a fusion product with the α -peptide. To answer this question, various fragments of the 2.8 kb insert were subcloned into pHG171 and assayed for expression in *E. coli* χ 2338 (see Table 1). These subclones were also used for translation experiments *in vitro* (see Figs 4a and 4b and Table 1). pBL290 and pBL324 remove 400 bp and 650 bp respectively of the insert from the end adjacent to the *lac* promoter. Both complement the mutation in *E. coli* χ 2338 (Table 1) and produce a protein of the same size as that produced by pBL265 (Fig. 4). Therefore citrate synthase from pBL265 is not a fusion protein and relies on its own ribosome-binding site.

pBL289 and pBL323 contain the same inserts as pBL290 and pBL324 respectively, but in opposite orientation with regards to the *lac* promoter. Both fail to complement *E. coli* χ 2338 and do not produce additional proteins when compared with pHG171 (Table 1; for pBL323 see also Fig. 4). The lack of complementation by pBL266, pBL289 and pBL323 indicates that the *M. smegmatis* citrate synthase gene does not contain a promoter that functions in *E. coli*.

Determination of the ends of the coding sequences of the citrate synthase gene

The transcription-translation experiment *in vitro* presented in Fig. 4 shows that the citrate synthase of *M. smegmatis* is about 42 kDa. This protein is similar in size to that of *M. leprae* (results not shown) but slightly smaller than *E. coli* citrate synthase (pDB2; Fig. 4b) and *Rickettsia prowazekii* citrate synthase [15,26,27]. A 42 kDa protein requires a structural gene of about 1.1 kb.

The approximate position of the 5'-end of the citrate synthase structural gene was determined by removing sequences adjacent to the *lac* promoter in pBL265. Subclones were made and these showed that the *Xho*I site (pBL290) and the *Eco*RI site (pBL324, pBL358 and pBL359) are upstream of the gene (Table 1). pBL358 and pBL359 were prepared by *Bal*31 digestion of pBL265 cleaved with *Hind*III. pBL358 removes about 300 bp beyond the *Eco*RI site, and pBL359 about 250 bp. Both produce a 42 kDa

protein and complement *E. coli* χ 2338. pBL358 only weakly complements (described below).

The 3'-end was located by truncation of sequences distal to *plac*. pBL290 complements and produces a 42 kDa protein; pBL332 weakly complements and synthesizes a 42 kDa protein; pBL288 neither complements nor synthesizes a detectable protein. Thus the 3'-end of the gene should be near the *Sca*I site. In agreement with the results obtained for pBL358 and pBL359, the region between the *Eco*RI and *Sca*I sites is sufficient for encoding a 42 kDa protein.

Sequence of the *M. smegmatis* citrate synthase gene

The 1800 bp *Eco*RI-*Xho*I segment is expected to cover the coding region of the gene and include some 250 bp flanking it on both sides. This fragment was sequenced by the dideoxy chain-termination method. The sequence and its translational product are given in Fig. 5. The 1.8 kb sequence contains only one large open reading frame, which is the structural gene for citrate synthase.

Other open reading frames can be seen. However, they code for not more than 140 amino acid residues. As pointed out by Schinnick [11], in an organism of high G+C content such as a mycobacterium, random stop codons are rarely expected and open reading frames of this size may not be significant. One of these open reading frames runs to the end of the sequenced region (residues 1403-1776) and overlaps the citrate synthase gene by 35 bp. No methionine codons are present in this region, but there are two GTG codons. Although it is not known whether GTG can serve as an initiation codon in mycobacteria, a sequence ACGAGCT exists 19 nucleotide residues upstream of one of the GTG codons. This might serve as a ribosome-binding site. The open reading frame of the citrate synthase gene also extends upstream of the ATG codon at position 313. This upstream region is at least 312 bp (*Eco*RI site) (see the Discussion section). The ATG codon is located 10 bp downstream of the centre of a Shine-Dalgarno sequence (AGGAGGT, starting at position 300). No other methionine codons are located upstream of that at position 312, and the predicted size of the polypeptide (41.5 kDa) is in good accord with the protein product seen in SDS/PAGE gels.

The area upstream of the *Pvu*II site was also sequenced from pBL358 and pBL359. pBL359 contains 70 bp upstream of the Shine-Dalgarno sequence and shows normal complementation. pBL358 is 58 bp shorter than pBL359, still retains 12 nucleotide residues upstream of the Shine-Dalgarno sequence but complements poorly. The reason for this is that a fusion protein of β -galactosidase-citrate synthase is being produced that has 14 amino acid residues of its N-terminus from β -galactosidase.

The sequence predicts that pBL332 is deleted for the last 10 amino acid residues of the citrate synthase C-terminus. The construction of pBL332 involved the ligation of the *Sca*I-produced end to the *Sma*I site of pHG165. This should result in the addition of 19 amino acid residues to the truncated end before a nonsense codon (TGA) is encountered.

The predicted citrate synthase amino acid sequence was compared with the published sequences for this enzyme from yeast [28], *E. coli* [29] and pig heart [30]. A significant degree of similarity (> 24%) exists between the *M. smegmatis* sequence and each of the others (Fig. 6). Furthermore, areas of strong similarity between the other sequences are those with a high degree of similarity to the mycobacterial enzyme. The histidine residues at positions 238, 274 and 320, the arginine residue at position 329 and the aspartic acid residue at position 375 of the pig heart enzyme are considered to be essential for catalysis [31]. It is noteworthy that these are conserved in the mycobacterial enzyme (see Fig. 6).

```

      10      20      30      40      50      60
      *      *      *      *      *      *
GAA TTC CCG CCT TCC GCG CCG CGT CGA CAC CCC GCT GCT GGC CAA TAT GAC CGA GTT CGG

      70      80      90      100      110      120
      *      *      *      *      *      *
CAA GTC CGA GCT GGT CAA TGC CCG CGA GCT GGC CGA CAT CGG CTA CAA CGT CGT CAT CTA

      130      140      150      160      170      180
      *      *      *      *      *      *
TCC CGT CAC CAC GCT GCG GCT GGC CAT GCA CGC CGT AGA GAT CGG TAT GCG CGA AGT CTT

      190      200      210      220      230      240
      *      *      *      *      *      *
CGA CGC TGG AAC GCA ATC CGG CCT ACT GGG CAG CAT GCA GCA CCG CAG CAG GCT CTA CGA
                                     pBL359

      250      260      270      280      290      300
      *      *      *      *      *      *
ACT GCT GCG CTA CGC CGA CTA CAA CCA GTT CGA CTC CGA CAT CTA CAA CTT CGC CGT GCA
                                     pBL358

      310      320      330      340      350      360
      *      *      *      *      *      *
GGA GGT GAG AGC ATG ACC ACA GCA ACA GAA TCC GAA GCT CCG CGC ATT CAC AAG GGA CTT
Met Thr Thr Ala Thr Glu Ser Glu Ala Pro Arg Ile His Lys Gly Leu

      370      380      390      400      410      420
      *      *      *      *      *      *
GCC GGT GTC GTC GTC GAC ACC ACC GCC ATC TCG AAG GTC GTG CCC GAG ACC AAC AGC CTG
Ala Gly Val Val Val Asp Thr Thr Ala Ile Ser Lys Val Val Pro Glu Thr Asn Ser Leu

      430      440      450      460      470      480
      *      *      *      *      *      *
ACC TAC CGG GGC TAC CCG GTG CAG GAC CTG GCC GCC CAG TGC AGC TTT GAA CAG GTC GCG
Thr Tyr Arg Gly Tyr Pro Val Gln Asp Leu Ala Ala Gln Cys Ser Phe Glu Gln Val Ala

      490      500      510      520      530      540
      *      *      *      *      *      *
TAT CTG TTG TGG CAC GGC GAA CTG CCG ACC GAT CAG CTG GCT TTG TTC AGC CAG CGT GAG
Tyr Leu Leu Trp His Gly Glu Leu Pro Thr Asp Gln Leu Ala Leu Phe Ser Gln Arg Glu

      550      560      570      580      590      600
      *      *      *      *      *      *
CGG GCG TCG CCG CCG ATC GAC CCG TCG ATG CAG GCG CTG CTG GCG AAG TTG CCC GAC AAC
Arg Ala Ser Arg Arg Ile Asp Arg Ser Met Gln Ala Leu Leu Ala Lys Leu Pro Asp Asn

      610      620      630      640      650      660
      *      *      *      *      *      *
TGT CAT CCC ATG GAC GTG GTG CGT ACG GCG ATC AGC TAT CTG GGA GCC GAG GAT CTA GAA
Cys His Pro Met Asp Val Val Arg Thr Ala Ile Ser Tyr Leu Gly Ala Glu Asp Leu Glu

      670      680      690      700      710      720
      *      *      *      *      *      *
GAG GAC GTC GAC ACG GCG GAG GCC AAC TAC GCC AAG TCG CTG CCG ATG TTC GCC GTG CTG
Glu Asp Val Asp Thr Ala Glu Ala Asn Tyr Ala Lys Ser Leu Arg Met Phe Ala Val Leu

      730      740      750      760      770      780
      *      *      *      *      *      *
CCG ACG ATC GTC GCC ACC GAC ATC CCG CGT CGT CAG GGA TTG ACC CCG ATC CCT CCG CAC
Pro Thr Ile Val Ala Thr Asp Ile Arg Arg Arg Gln Gly Leu Thr Pro Ile Pro Pro His

      790      800      810      820      830      840
      *      *      *      *      *      *
AGC CAG CTC GGT TAT GCG CAG AAC TTC TTG AAC ATG TGC TTC GGT GAG GTG CCG GAG CCC
Ser Gln Leu Gly Tyr Ala Gln Asn Phe Leu Asn Met Cys Phe Gly Glu Val Pro Glu Pro

      850      860      870      880      890      900
      *      *      *      *      *      *
GTC GTG GTG CCG GCG TTC GAA CAG TCG ATG GTG CTC TAC GCT GAG CAC AGC TTC AAC GCC
Val Val Val Arg Ala Phe Glu Gln Ser Met Val Leu Tyr Ala Glu His Ser Phe Asn Ala

      910      920      930      940      950      960
      *      *      *      *      *      *
TCC ACC TTC GCC GCC CCG GTC GTC ACC TCG ACC CAG TCA GAC ATC TAC AGC GCG GTG ACC
Ser Thr Phe Ala Ala Arg Val Val Thr Ser Thr Gln Ser Asp Ile Tyr Ser Ala Val Thr

      970      980      990      1000      1010      1020
      *      *      *      *      *      *
GCG GCG ATC GGC GCT CTC AAG GGT TCG CTG CAC GGC GGC GCC AAC GAA GCG GTC ATG CAC
Ala Ala Ile Gly Ala Leu Lys Gly Ser Leu His Gly Gly Ala Asn Glu Ala Val Met His

```

```

      1030      1040      1050      1060      1070      1080
      *      *      *      *      *      *
GAC ATG CTC GAA ATC GGT TCG GCC GAA AAG GCT CCG GAG TGG TTG CAC GGC AAA CTA TCC
Asp Met Leu Glu Ile Gly Ser Ala Glu Lys Ala Pro Glu Trp Leu His Gly Lys Leu Ser

      1090      1100      1110      1120      1130      1140
      *      *      *      *      *      *
CGC AAG GAG AAG GTC ATG GGA TTT GGC CAC CGG GTC TAC AAG AAC GGT GAT TCC CGG GTG
Arg Lys Glu Lys Val Met Gly Phe Gly His Arg Val Tyr Lys Asn Gly Asp Ser Arg Val

      1150      1160      1170      1180      1190      1200
      *      *      *      *      *      *
CCG ACC ATG AAG GTT GCT TTG GAG CAG GTT GCC CAG GTG CGT GAC GGA CAG CGT TGG CTC
Pro Thr Met Lys Val Ala Leu Glu Gln Val Ala Gln Val Arg Asp Gly Gln Arg Trp Leu

      1210      1220      1230      1240      1250      1260
      *      *      *      *      *      *
GAC ATC TAC AAC ACG CTG GAG AGC GCG ATG TTC GCA GCG ACG AGG ATC AAG CCG AAC CTC
Asp Ile Tyr Asn Thr Leu Glu Ser Ala Met Phe Ala Ala Thr Arg Ile Lys Pro Asn Leu

      1270      1280      1290      1300      1310      1320
      *      *      *      *      *      *
GAC TTT CCC ACC GGC CCT GCC TAT TAC CTG ATG GAC TTC CCG ATC GAG AGT TTC ACC CCG
Asp Phe Pro Thr Gly Pro Ala Tyr Tyr Leu Met Asp Phe Pro Ile Glu Ser Phe Thr Pro

      1330      1340      1350      1360      1370      1380
      *      *      *      *      *      *
CTG TTC GTG ATG AGC CGG ATC ACC GGT TGG ACC GCC CAC ATC ATG GAA CAG GCT GCC TCC
Leu Phe Val Met Ser Arg Ile Thr Gly Trp Thr Ala His Ile Met Glu Gln Ala Ala Ser

      1390      1400      1410      1420      1430      1440
      *      *      *      *      *      *
AAT GCT CTG ATC CGT CCC CTG AGT GAG TAC TCG GGN CAG CCG CAG CGG TCC CTT GTC TGA
Asn Ala Leu Ile Arg Pro Leu Ser Glu Tyr Ser Gly Gln Pro Gln Arg Ser Leu Val ---

      1450      1460      1470      1480      1490      1500
      *      *      *      *      *      *
GCG CGC GAT CAC CAA ACT GCT TGT CGC CAA TCG CGG CGA GAT CGC CAT CCG CGC GTT TCG

      1510      1520      1530      1540      1550      1560
      *      *      *      *      *      *
GGC CGC CTA CGA GCT CGG AAT CTC GAC GGT CGC GGT GTT TGC CTA CGA GGA CCG CAA TTC

      1570      1580      1590      1600      1610      1620
      *      *      *      *      *      *
GCC GCA CCG TCT CAA AGC GGA CGA GGC ATA CCA GAT CGG CGA AAC AGG TCA TCC GGT CCG

      1630      1640      1650      1660      1670      1680
      *      *      *      *      *      *
CGC ATA CCT GTC TGT CGA TGA GAT AAT CCG GGT AGC AAA GCA TTC CGG TGC TGA CGC AAT

      1690      1700      1710      1720      1730      1740
      *      *      *      *      *      *
CTA TCC GGG GTA CGG ATT CCT CTC GGA GAA TCC TGA ACT GGC TGC GGC GTG TCA CCA GGC

      1750      1760      1770
      *      *      *
CGG CAT CGT GTT CGT CGG GCC GCG TGC ACA CGT TCT

```

Fig. 5. Sequence of the *EcoRI*-*XhoI* 1.8 kb fragment of clone pBL265

The ribosome-binding-site sequence starting at position 300 is underlined. Arrows indicate the start of the *M. smegmatis* insert in pBL358 and pBL359.

Citrate synthase activity in *E. coli* χ 2338 carrying pBL265

Citrate synthases from Gram-positive bacteria and eukaryotes differ in their response to inhibitors from their counterparts in Gram-negative bacteria. The enzymes from Gram-positive bacteria are inhibited by ATP *in vitro* but not by 2-oxoglutarate, whereas citrate synthases from Gram-negative bacteria are insensitive to ATP but inhibited by 2-oxoglutarate.

The citrate synthase activity in various strains was assayed as described in the experimental section. *E. coli* NM522 contains,

as expected [15,26], an enzyme that is inhibited by 2-oxoglutarate but not by ATP. *E. coli* χ 2338, the *gltA* mutant strain, did not possess measurable activity. By contrast, *E. coli* χ 2338 bearing either pBL265 or pBL358 had citrate synthase activity, and this activity was inhibited by ATP but not by 2-oxoglutarate. As suggested by the complementation experiments *in vivo*, the specific activity of citrate synthase from pBL358 is lower than that of pBL265. *E. coli* χ 2338 containing pYA1036 is reported to grow even more slowly on selective plates (4 days; [8]) and its specific activity *in vitro* is even lower (Table 2).

DISCUSSION

The evidence presented above shows that the gene for citrate synthase from *M. smegmatis* can be expressed in *E. coli* when it is connected to an *E. coli* promoter and that the gene contains a functional ribosome-binding site. Indeed, an AGGAGGT sequence is found centred 10 bp upstream of what appears to be the codon for formylmethionine used by this gene. Deletions of the 5'-end and 3'-end of the *M. smegmatis* insert and the size of the enzyme make it unlikely that a codon other than the ATG at position 313 is used as a translational start site. Indeed, upstream of the ATG codon at position 313 neither ATG nor GTG codons are seen in the frame. Downstream of ATG, there is a GTG codon at position 400. However, initiation from this codon would result in a protein that is shorter by 29 amino acid residues, and this is not consistent with the size of the protein product. Furthermore, it is not known whether GTG codons can serve as initiators for translation in mycobacteria.

No promoter that functions in *E. coli* was found upstream of the gene, not were there any promoter-like sequences resembling those from *E. coli*. However, the 5'-terminal region might function as a promoter in mycobacteria.

Construction and analysis of mycobacterial genomic libraries [32,33] revealed that most mycobacterial genes lack the transcription and/or translation features required for their expression in *E. coli*. However the citrate synthase genes from both *M. leprae* [6] and *M. smegmatis* and the 65 kDa antigen gene from *Mycobacterium tuberculosis* [11] appear to rely on mycobacterial ribosome-binding sites in *E. coli*. The sequences of the *M. smegmatis* and *M. tuberculosis* genes have revealed *E. coli*-like ribosome-binding-site sequences upstream of ATG codons. One gene, the 65 kDa antigen, appears to possess *E. coli*-like transcription initiation sequences, but it is not known whether these function in mycobacteria.

Two potential palindromic structures were found downstream of the citrate synthase gene, and these were centred at positions 1557 and 1688. Upstream of the gene, a potential palindrome was found at positions 134-155. It is not known if any of these function in *M. smegmatis*. The palindrome upstream of citrate synthase clearly does not terminate transcription in *E. coli*.

The codon usage for citrate synthase resembles that found by Schinnick for two *M. tuberculosis* open reading frames. The overall G+C content is 62%. However, this is due to the G+C content of the third (wobble) base (82%); the first two positions of the codons have a G+C content of only 53%. This third base compensation eliminates the need for an unusual distribution of amino acid residues in proteins encoded by a G+C-rich genome (Table 3).

The reading frame containing the citrate synthase gene is open for at least 312 bp preceding the ATG codon. As mentioned above, the high G+C content of mycobacteria lowers the probability of randomly encountering a stop codon to about half that in an organism with equimolar G+C and A+T contents [11]. Long open reading frames upstream of the putative codons for formylmethionine can also be seen in the two open reading frames sequenced by Schinnick [11]. However, the sequence

Fig. 6. Comparison of the amino acid sequences of citrate synthase from various organisms

The species compared are *M. smegmatis* (MCS), *Saccharomyces cerevisiae* (YCS), *E. coli* (ECS) and pig heart (PCS). Similar or identical residues are boxed. The mycobacterial initiator methionine residue is marked with a +, and some of the region preceding it is also presented. The functionally important and conserved residues of the pig heart enzyme (see the text) are marked by asterisks (*).

Table 2. Citrate synthase activity and the effects of ATP and 2-oxoglutarate

Soluble protein extracts were prepared as described in the Experimental section. The resulting extracts contained between 3.4 and 7.8 mg of protein/ml. Citrate synthase activity was measured spectrophotometrically at A_{412} by the method of Srere [24]. The specific activities as presented are the means $\pm$ S.D. for three to nine assays. Abbreviation: N.D., not determined.

Source of soluble protein	Specific activity (munits/mg of protein)	Specific activity (% of control)	
		In the presence of 2-oxoglutarate	In the presence of ATP
<i>E. coli</i> NM522	10.9 $\pm$ 3.0	45.9	92
Clone pBL265	5.48 $\pm$ 0.78	120	64
Clone pBL358	2.54 $\pm$ 0.16	N.D.	N.D.
Clone pYA1036	0.50 $\pm$ 0.04	112	66.6
<i>E. coli</i> χ 2338	No activity detected		

Table 3. Codon usage for the citrate synthase gene of *M. smegmatis*

TTT	Phe	3	TCT	Ser	0	TAT	Tyr	4	TGT	Cys	1
TTC	Phe	11	TCC	Ser	6	TAC	Tyr	9	TGC	Cys	2
TTA	Leu	0	TCA	Ser	1	TAA	—	0	TGA	—	0
TTG	Leu	7	TCG	Ser	9	TAG	—	0	TGG	Trp	4
CTT	Leu	2	CCT	Pro	2	CAT	His	1	CGT	Arg	7
CTC	Leu	6	CCC	Pro	6	CAC	His	9	CGC	Arg	4
CTA	Leu	2	CCA	Pro	0	CAA	Gln	0	CGA	Arg	0
CTG	Leu	16	CCG	Pro	13	CAG	Gln	17	CGG	Arg	11
ATT	Ile	1	ACT	Thr	0	AAT	Asn	1	AGT	Ser	2
ATC	Ile	15	ACC	Thr	17	AAC	Asn	10	AGC	Ser	9
ATA	Ile	0	ACA	Thr	2	AAA	Lys	1	AGA	Arg	0
ATG	Met	14	ACG	Thr	5	AAG	Lys	11	AGG	Arg	1
GTT	Val	2	GCT	Ala	8	GAT	Asp	3	GGT	Gly	8
GTC	Val	14	GCC	Ala	18	GAC	Asp	14	GGC	Gly	8
GTA	Val	0	GCA	Ala	2	GAA	Glu	10	GGA	Gly	5
GTG	Val	13	GCG	Ala	14	GAG	Glu	14	GGG	Gly	0

Fig. 7. Sequence similarity between the N-terminus of citrate synthase from yeast and the protein encoded by part of the region preceding the ATG codon of the *M. smegmatis* *gltA* clone

Identical residues are boxed.

upstream of the codon for formylmethionine in the citrate synthase described here appears to be conserved. This is deduced from comparing the amino acid sequences from the other citrate synthases to the coding potential of the open reading frame upstream of the ATG codon at position 313 (Fig. 6). The yeast citrate synthase has a longer N-terminus than the others and shows significant similarity to the upstream region (Fig. 7). Similar to the yeast N-terminal sequence, the mycobacterial sequence upstream of the citrate synthase gene is rich in basic residues. Some of the amino acid residues at the N-terminus of

the yeast enzyme are responsible for targeting the protein to the mitochondria.

The underlying reason for this interesting similarity remains unknown.

We thank T. Bar On for plasmid preparations, E. Bruler for media preparation and J. Kenton for careful typing of the manuscript.

REFERENCES

- Thangaraj, H. S., Lamb, F. I., Davis, E. D., Jenner, P. J., Jeyakuma, L. H. & Colston, M. J. (1990) *Infect. Immun.* **58**, 1937–1942
- Baird, P. N., Hall, L. M. & Coates, A. R. (1989) *J. Gen. Microbiol.* **135**, 931–939
- Suzuki, Y., Nagata, A., Oni, Y. & Yamada, T. (1988) *J. Bacteriol.* **170**, 2886–2889
- Green, E. P., Tizard, M. L., Moss, M. T., Thompson, J., Winterbourne, D. J., McFadden, J. J. & Hermon-Taylor, J. (1989) *Nucleic Acids Res.* **17**, 9063–9073
- Sirakova, T. D., Bardarov, S. S., Kriakov, J. I. & Markov, K. I. (1989) *FEMS Microbiol. Lett.* **50**, 153–156
- Ashbridge, K. R., Booth, R. J., Watson, J. D. & Lathigra, R. B. (1989) *Nucleic Acids Res.* **17**, 1249
- Wallace, R. J., Jr., Nash, D. R., Tsukamuiira, M., Blacklock, Z. M. & Silcox, V. A. (1988) *J. Infect. Dis.* **158**, 52–59
- Jacobs, Jr., W. R., Docherty, M. A., Curtiss, R., III & Clark-Curtiss, J. E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1926–1930
- Athwal, R. S., Deo, S. S. & Imaeda, T. (1984) *Int. J. Syst. Bacteriol.* **34**, 371–375
- Imaeda, T. (1985) *Int. J. Syst. Bacteriol.* **35**, 147–150
- Schinnick, T. M. (1987) *J. Bacteriol.* **169**, 1080–1088
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119

13. Hanahan, D. (1985) in *DNA Cloning* (Glover, D. M., ed.), pp. 109–136, IRL Press, London and Washington
14. Stewart, G. S. A. B., Lubinsky-Mink, S., Jackson, C. G., Cassel, A. & Kuhn, J. (1986) *Plasmid* **15**, 172–181
15. Bloxham, D. P., Herbert, C. J., Ner, S. S. & Drabble, W. T. (1983) *J. Gen. Microbiol.* **129**, 1889–1897
16. Brosius, J. (1984) *Gene* **27**, 151–160
17. Bercovier, H., Kafri, O. & Sela, S. (1986) *Biochem. Biophys. Res. Commun.* **136**, 1136–1141
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
19. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 431–433, Cold Spring Harbor Laboratory, Cold Spring Harbor
20. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
21. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
22. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1518
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
24. Sreere, P. A. (1969) *Methods Enzymol.* **13**, 3–11
25. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
26. Wood, D. O., Atkinson, W. H., Sikorski, R. S. & Winkler, H. H. (1983) *J. Bacteriol.* **155**, 412–416
27. Wood, D. O., Williamson, L. R., Winkler, H. H. & Krause, D. C. (1987) *J. Bacteriol.* **169**, 3564–3572
28. Suissa, M., Suda, K. & Schatz, G. (1984) *EMBO J.* **3**, 1773–1781
29. Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W. & Bloxham, D. (1983) *Biochemistry* **22**, 5243–5249
30. Bloxham, D. P., Parmelee, D. C., Kumar, S., Wade, R. D., Ericsson, L. H., Neurath, H., Walsh, K. A. & Titani, K. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5381–5385
31. Bell, A. W., Bayena, V. & Duckworth, H. W. (1983) *Biochemistry* **22**, 3400–3405
32. Clark-Curtiss, J. E., Jacobs, W. R., Docherty, M. A., Ritchie, L. R. & Curtiss, R. (1985) *J. Bacteriol.* **161**, 1093–1102
33. Thole, J. E. R., Dauwerse, G., Das, P. K., Groothuis, D. G., Schouls, L. M. & van Embden, J. D. A. (1985) *Infect. Immun.* **50**, 800–806

Received 21 August 1990/14 December 1990; accepted 8 January 1991