The regulatory elements of the *Mycobacterium tuberculosis* gene Rv3881c function efficiently in *Escherichia coli*

Vijaya Satchidanandam †, Rama Rao Amara †, Pradeep Devappa Uchil, Varsha Singh

*Department of Microbiology and Cell Biology, Indian Institute of Science, Room 254A, Sir C.V. Raman Avenue, Bangalore 560012, India*

Received 26 July 2002; received in revised form 22 November 2002; accepted 29 November 2002

First published online 25 December 2002

**Abstract**

We report efficient expression of the *Mycobacterium tuberculosis* gene Rv3881c in *Escherichia coli* from its *M. tuberculosis* promoter, attributable to an *E. coli* consensus Pribnow box and ribosome binding site. The N-terminal sequence of the recombinant *E. coli*-generated protein was identical to the predicted open reading frame of Rv3881c and transcription of the Rv3881c gene initiated at the same nucleotide position in both bacteria. We demonstrate the utility of this promoter for rapid analysis of expression in *E. coli* of heterologous gene constructs, for subsequent expression from the genomes of slow-growing mycobacteria such as *Mycobacterium bovis*-BCG. *M. tuberculosis* Rv3881c homologues were present in other pathogenic mycobacteria such as *M. bovis*-BCG, *Mycobacterium szulgai* and *Mycobacterium kansasii*.

© 2002 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

**Keywords:** *Mycobacterium tuberculosis*; Rv3881c; Promoter; Ribosome binding site

1. Introduction

The genus *Mycobacterium* includes bacterial pathogens that take a toll of the largest number of human lives each year. The increasing incidence of tuberculosis worldwide in the past decade has thrown up many challenges including the need for successful genetic manipulation that would facilitate the generation of attenuated variants of the slow-growing pathogenic mycobacteria. The slow growth rate of the pathogenic members of this genus such as *Mycobacterium tuberculosis* (MTB) and *M. bovis* as well as the absence of a panel of plasmid vectors and strong promoters has made high-level expression of homologous and heterologous genes in these mycobacteria problematic. The non-pathogenic *Mycobacterium bovis*-BCG has been explored as a recombinant vaccine for several infectious diseases including HIV–AIDS and tuberculosis by the stable insertion of genes for ‘protective antigens’ into its genome [1–3]. Transcriptional and translational regulatory elements that can function efficiently in both *Escherichia coli* and mycobacteria are therefore of particular value as they would permit rapid and easy evaluation in the surrogate *E. coli* cells, of gene constructs for ultimate expression in mycobacteria.

Mycobacterial promoters with ability to function efficiently in *E. coli* are rare. A study to assess 125 MTB promoters [4] revealed that none of these was proficient to drive expression of the chloramphenicol acetyltransferase gene in *E. coli* [5]. The only mycobacterial regulatory elements known to function in *E. coli* belonged to the 65-kDa heat shock protein gene from *Mycobacterium leprae*, *M. bovis*-BCG and MTB [5–7]. While this promoter from MTB was shown to successfully drive the expression of β-galactosidase in *E. coli* [6], characterization of this promoter has not been carried out either in MTB or in *E. coli*.

One of the recombinant clones we had earlier identified in a screen of a lambda ZAP II::MTB genomic DNA expression library with pooled sera from TB patients [8] contained the Rv3881c gene of MTB [9], which shared homologues in the pathogenic mycobacteria *M. bovis*, *Mycobacterium kansasii* and *Mycobacterium szulgai*. In the
present report, we characterize and demonstrate that the promoter of this hitherto unreported MTB gene successfully directs the generation of transcripts in *E. coli*. We also demonstrated the stable maintenance in the genome of the slow-growing mycobacterium *M. bovis*-BCG, of a construct carrying an in-frame fusion of the Rv3881c gene to the green fluorescent protein (GFP) gene, driven by the Rv3881c promoter. Protein expression from this construct was verified in *E. coli*, thus demonstrating the utility of the Rv3881c promoter for rapid evaluation of protein expression from the MTB regulatory elements using *E. coli* as a substitute.

2. Materials and methods

2.1. Bacterial strains

MTB strain NTI-83949 obtained from the sputum of a patient with pulmonary TB was characterized as reported earlier [8]. All other mycobacterial strains were obtained from the Mycobacterial Repository Centre, Central JALMA Institute for Leprosy, Agra, India. Mycobacteria were grown in Middlebrook 7H9 medium supplemented with 0.5% albumin and 0.75% dextrose. *E. coli* strains SOLR and BL21(DE3) were obtained from Stratagene and DH10B was obtained from Life Technologies BRL.

2.2. Cloning procedures and DNA sequencing

Two clones, MTB213 and MTB202, representing the Rv3881c gene of MTB were obtained from the lambda ZAP II::MTB genomic DNA expression library and excised to obtain the pBluescript SK(+) [pBSSK(+), Stratagene] phagemids containing the MTB DNA inserts according to manufacturer’s instructions as described [8]. MTB213 contained a fusion of the Rv3881c gene at nucleotide (nt) position 192 of the open reading frame (ORF), to the lacZ ORF of the pBS vector. The insert from MTB213 was digested with *PstI*, which cuts in the vector upstream of the insert and *Sma*I which cuts 30 nt downstream of the stop codon of the Rv3881c ORF. The 2.1-kb fragment thus obtained was ligated to pRSETA between the *PstI* and Klenow-filled *EcoRI* sites to generate pRSETA213, which expressed the MTB213 protein fused to six histidine residues provided by the vector at its N-terminus. Recombinant protein obtained following induction of *E. coli* BL21(DE3) transformed with pRSETA213 using 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) was affinity purified by binding to and elution from Ni²⁺-NTA agarose matrix. Purified protein was used to generate antiserum in BALB/c mice as described [10] after purity was confirmed by N-terminal sequencing using Edman chemistry.

The complete 1383 nt long Rv3881c ORF along with its promoter contained within a 275 nt segment upstream of the start codon ATG was isolated from MTB202 as a 2.7-kb *Xba*I–*EcoRV* fragment that was Klenow-filled and cloned into the *EcoRV* site of the mycobacterial integration vector pDK20 [11] to give pDK-202. To construct the in-frame fusion of GFP with the ORF of Rv3881c in pDK20, the GFP gene [12] was obtained as a Klenow-filled 734-bp fragment after *Ban*I digestion of pGFP (Clontech) and ligated in frame to the Rv3881c ORF in MTB202 at the *Bsr*EI site 33 nt upstream of the stop codon to give rise to MTB202GFP. The 202GFP fusion gene was then excised with *Xba*I and *EcoRV*, Klenow-filled and ligated to *EcoRV*-digested pDK20 to generate pDK20GFP.

Recombinant pBS plasmids obtained by excision of clones from the lambda ZAP II library were sequenced using the T3 primer in an ABI Prism automated sequencing system using Taq DNA polymerase.

2.3. Protein analysis

Mycobacterial sonicates prepared as described [13] containing 50 µg protein and recombinant *E. coli* lysates containing 100–200 ng of the specific recombinant protein were electrophoresed per lane in SDS 10% polyacrylamide gels, Western blotted and immunostained with mouse anti-213 serum according to previously published procedures [13]. For N-terminal sequencing, the MTB202 protein or the purified 213 protein expressed from pRSETA213 was similarly electrophoresed, transferred to Immobilon-PVDF membranes, stained with 0.5% Ponceau S in 0.1% acetic acid and sequenced in a Precise model of Applied Biosystems, Foster City, CA, USA.

2.4. Southern analysis

Genomic DNA was isolated from mycobacteria grown in 7H9 broth as described previously [8], digested with *EcoRI* and *SalI* to completion and electrophoresed on 0.8% agarose gels. DNA transferred to nylon membranes was hybridized to a 1220-nt fragment obtained by *Bsm*I digestion of MTB213 that contained all but the first 192 nt of the Rv3881c ORF. Probes were labeled using the Gene Images-CDP detection kit from Amersham Life Science and membranes were processed according to manufacturer’s instructions.

2.5. Isolation of total RNA and identification of transcription starting points

MTB grown in Middlebrook 7H9 broth and *E. coli* grown in Luria–Bertani broth were harvested in mid growth phase and suspended in guanidinium isothiocyanate buffer. RNA was isolated according to published procedures [14], and treated with RNase-free DNase I.

Both RNase protection and primer extension methods were used to map the transcription start site of the
Rv3881c gene in MTB and *E. coli*. The probe for the former consisted of a 173-nt antisense transcript, whose last 149 nt were complementary to the region between the *NruI* and *Bsr*XI sites (Fig. 2A) of the Rv3881c gene. Labeling of this transcript by α<sup>32</sup>P-UTP and subsequent hybridization to total RNA from MTB NT183949 and *E. coli* followed by digestion with RNase A and RNase T1 (Roche Molecular Biochemicals) were performed according to standard procedures [15]. Products were analyzed on a 7 M urea–4% acrylamide gel.

Primer extension was carried out using a modified RNA sequencing protocol. 5 µg of total RNA from *E. coli* or MTB cells was annealed to 10 pmol of the specific antisense primer 5′-CCT CCA CCT CGT TGG CCC which maps between 67 and 50 nt downstream of the start of the ORF of Rv3881c (Fig. 2B). Reverse transcription reactions were performed using SUPERSCRIPT II RNaseH<sup>−</sup>-Reverse Transcriptase (Gibco BRL) in presence of [α<sup>32</sup>P]dATP (10 µCi, 3000 Ci/mmole) and 5 µM each of dCTP, dGTP and dTTP at 50°C for 5 min. This was followed by an extension reaction for 5 min in the presence of 200 µM each of the four dNTPs. The primer-extended products were analyzed by electrophoresis on a 7 M urea–6% acrylamide gel along with cycle sequencing reactions of MTB202 plasmid using the same primer mentioned above, performed using the *fmol* DNA cycle sequencing system (Promega Corporation, Madison, WI, USA) according to manufacturer’s instructions. Gels were scanned on a Fuji BAS1000 phosphorimager and analyzed using the FujiMacBAS V2.4 software.

3. Results

3.1. Rv3881c gene of MTB is expressed efficiently in *E. coli* from its own regulatory elements

The lambda ZAP II::MTB genomic DNA library [8] contained two clones MTB213 and MTB202 carrying the Rv3881c gene of MTB in the vector pBSK cloned at the *Eco*RI site. MTB202 carried a 2.9 kb insert and expressed a 48-kDa recombinant protein while MTB213 expressed a 45-kDa protein in *E. coli* (Fig. 1A, lanes 2 and 3). The mobility on polyacrylamide gels of the recombinant MTB202 protein expressed in *E. coli* corresponded to the predicted molecular mass of 47 594 Da for the 1383 nt ORF of the recombinant MTB202 protein in *E. coli* as determined from the nucleotide sequence at the beginning of the ORF. The start of the ORF is at the 213 bp downstream from the start of the *lacZ* ORF of the vector pBS vector used in MTB202 expression in *E. coli* revealed (for details) that the 48-kDa size of the recombinant 202 protein was identical to that of the authentic Rv3881c protein from MTB sonicates (Fig. 1B, lanes 1 and 2). This antiserum specifically detected only the MTB202 and MTB213 proteins (Fig. 1B, lane 3). It was therefore of interest to investigate the origins of the 48-kDa recombinant protein produced by MTB202 in *E. coli* cells carrying this plasmid.

The out of phase positioning of the start codons of the vector and insert as well as the size identity of the authentic MTB Rv3881c with *E. coli*-expressed MTB202 protein suggested that the recombinant protein in MTB202 initiated at the ATG of the Rv3881c ORF within the insert
and not from the lacZ start codon present upstream in the pBS vector. In keeping with this, N-terminal sequence analysis of the recombinant protein expressed by MTB202 in E. coli uncovered the sequence TQSQTVTVD, which was identical to that predicted from the Rv3881c ORF of MTB after removal of the initiator formyl methionine (Fig. 2B).

3.2. Transcription start site analysis of the MTB Rv3881c gene

We next asked whether the same start site was used to transcribe this gene in MTB and in E. coli. RNase protection analysis of equal quantities of total RNA isolated from E. coli cells carrying pDK-202 and MTB using a probe spanning from the NruI site located 94 nt upstream to the BstXI site at position 55 downstream of the start ATG (Fig. 2A and B) gave rise to identical protected fragments approximately 80 nt long (Fig. 3A, lanes E and M, respectively). The intensity of the protected fragment obtained from E. coli RNA (Fig. 3A, lane E) was 5.4 times greater than that obtained with MTB RNA (Fig. 3A, lane M). This might be attributed to the higher copy number of the Rv3881c gene in E. coli present extrachromosomally on pDK20, a high copy plasmid carrying the Co1EI origin of replication [11], as opposed to the single chromosomal copy in MTB and/or due to differences in message stability between the two bacterial species. Primer extension analysis of total RNA from MTB as well as E. coli with or without pDK-202 (Fig. 3B, lanes M, E and C, respectively) along with parallel sequence determination of the Rv3881c gene with the same primer revealed that nucleotide A on the template strand, 24 nt upstream of the start

---

Fig. 2. Scheme of the MTB202 plasmid. A: The vector pBS is shown as thin lines while the insert of MTB genomic DNA is shown as thick lines. The bent thick lines at the two ends represent the vector backbone. The Rv3881c ORF is shown as a hatched bar. The black box (Pws) in the vector upstream of the multiple cloning site (MCS) represents the lac promoter. The two bold bent arrows represent the transcription start sites of the lac promoter and the Rv3881c promoter, respectively. Relevant restriction enzyme sites are shown. The thick bar above the gene represents the probe used for RNase protection studies spanning between the NruI and BstXI sites. XbaI and EcoRI fall within the MCS. B: The sequence of the sense strand of the Rv3881c gene near the start codon ATG (bold and underlined) showing the putative SD sequence, Pribnow box, transcription start site (+1) and the sites for BstXI and NruI. The translated protein sequence is shown below. The sequence of the anti-sense primer used in primer extension/sequencing is shown in bold above the corresponding sense sequence.

Fig. 3. Analysis of the transcription start site of MTB Rv3881c. A: RNase protection analysis of total RNA from E. coli carrying pDK20 vector (lane C), pDK-202 (lane E) or MTB (lane M). Lane SM contains HinfI-digested pX174 DNA as size marker. The sizes of the bands in nt are shown on the right. The arrow points to the 80-nt protected fragment in lanes E and M. B: Primer extension analysis of total RNA from MTB (lane M), E. coli carrying pDK-202 (lane E), or pDK20 vector (lane C). CTAG represents the sequencing lanes of MTB202 DNA with the primer used for primer extension analysis as described in Section 2. The arrows point to the base A on the template strand that functions as the transcription start site.
of the ORF (arrow in Fig. 3B), was the start site for transcription of Rv3881c both in MTB and in E. coli. The intensity of the primer-extended product using E. coli RNA was 5.7 times greater than the MTB product. Thus, the promoter and RBS of the Rv3881c gene of MTB represent the second MTB regulatory element that can function efficiently in E. coli after the hsp65 promoter which was earlier reported to drive expression of heterologous genes in the E. coli system [6].

3.3. Rv3881c gene of MTB can be introduced into and maintained in slow-growing mycobacteria

The constructs pDK-202 and pDK202GFP gave rise to proteins of the expected sizes of 48 and 74 kDa in E. coli DH10B (Fig. 4A, lanes 2 and 3). In addition, the 74-kDa protein was also detected by immunostaining using anti-GFP antibodies (data not shown). We chose the non-pathogenic M. bovis-BCG as the recipient for the GFP fusion of the MTB Rv3881c gene. We transformed BCG to kanamycin resistance with the pDK202GFP fusion construct. Kanamycin resistant colonies of BCG revealed the expression of this 76-kDa fusion protein in addition to the endogenous homologue of Rv3881c (Fig. 4B, lanes 1 and 2). The level of protein expressed from the copy of the gene inserted into the genome of BCG was found to be independent of the orientation of the gene in the pDK20 vector (data not shown). While the E. coli cells carrying the pDK202GFP showed strong fluorescence that was easily visualized by exposing them to ultraviolet light, fluorescence emission of recombinant BCG expressing this protein could only be detected using a FACS analyzer (FACSCAN, BD, data not shown), in keeping with their levels of expression detected by immunostaining. Thus, our results show that the Rv3881c promoter can be successfully deployed to direct expression of both homologous and heterologous genes in the slow-growing mycobacteria.

3.4. Distribution of homologues of MTB Rv3881c in other mycobacteria

Western immunoblot analysis of protein lysates from different mycobacteria using antibodies specific to protein 213 (Fig. 5A) revealed the presence of a cross-reactive protein in M. bovis and BCG while weakly cross-reactive proteins of similar size were detected in M. kansasii, M. szulgai and M. scrofulaceum. All other virulent as well as avirulent mycobacteria tested were negative. Southern analysis also revealed the presence of hybridizing DNA segments only in M. bovis and BCG other than in MTB strains (Fig. 5B). The lack of hybridization of Rv3881c sequence to DNA from M. kansasii, M. szulgai and M. scrofulaceum may be due to the dissimilar GC content of the probe and target DNAs. We have observed similar differences between Western reactivity and Southern hybridization signals for other MTB genes [13], presumably due to differences in codon usage between MTB and these mycobacteria.

4. Discussion

This manuscript describes the characterization of the promoter of the MTB gene Rv3881c that is able to promote transcription in E. coli. We have identified a clone in
References


Acknowledgements

Our thanks are due to the National Tuberculosis Institute for the field strain NTI83949 of MTB. Thanks are due to Anil Tyagi of Delhi University for the gift of the plasmid pDK20. Priti Kumar is acknowledged for valuable discussions and help with manuscript preparation. We owe immense gratitude to K. Nagamohan for help with the sequencing of the Rv3881c gene. Thanks are due to Dinakar Salunke for N-terminal sequencing of the MTB202 protein. We thank V.M. Katoh, of the Central Jai Institute for Leprosy, Agra, and its Mycobacterial Repository Centre for providing the strains of mycobacteria used in this study. A.R.R. and P.D.U. were recipients of the senior research fellowship from the Council of Scientific and Industrial Research.

the genomic library containing Rv3881c, which also contained 275 nt upstream of the start codon of this gene. The size of the protein expressed from this clone, which was identical to that of the authentic MTB protein, alerted us to the activity of its promoter in E. coli. Initial inference of the activity of this promoter in E. coli was based on Western immunoblots, which are a measure of the combined efficiency of both transcriptional as well as translational signals. The presence of a good E. coli consensus RBS upstream of the Rv3881c ORF was the likely explanation for the high levels of recombinant protein produced. Subsequent analysis of the promoter revealed that it could indeed be transcribed in both these diverse species of bacteria.

The −35 region of mycobacterial promoters was shown to be highly variable in sequence [16], presumably reflecting the presence of multiple sigma factors with varying specificities. In keeping with this absence of an E. coli-like consensus sequence, the M. smegmatis principal sigma factor was found to differ significantly from the E. coli homologue in the 4.2 region which contacts the −35 region of the promoter. The Rv3881c promoter did not have a consensus −35 region nor did it reveal an extended −10 TGN motif which has been found in 22% of mycobacterial promoters analyzed and shown to be an important determinant of transcriptional strength in mycobacteria as in E. coli [17]. The Rv3881c gene introduced into the slow-growing mycobacteria was expressed as efficiently as the endogenous homologue, judging from the Western immunoblots. This second copy was also maintained stably for prolonged periods of growth in the slow-growing mycobacteria without perturbing the original gene. The lack of efficient homologous recombination in these mycobacteria may have contributed to this stability.