Published online 18 August 2006 *Nucleic Acids Research, 2006, Vol. 34, No. 15* **4245–4257**
doi:10.1093/nar/gkl521

Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity

Nisheeth Agarwal and Anil K. Tyagi*

Department of Biochemistry, University of Delhi South Campus, Benito Juarez Road, New Delhi 110 021, India

Received April 17, 2006; Revised and Accepted July 6, 2006

ABSTRACT

Majority of the promoter elements of mycobacteria do not function well in other eubacterial systems and analysis of their sequences has established the presence of only single conserved sequence located at the −10 position. Additional sequences for the appropriate functioning of these promoters have been proposed but not characterized, probably due to the absence of sufficient number of strong mycobacterial promoters. In the current study, we have isolated functional promoter-like sequences of mycobacteria from the pool of random DNA sequences. Based on the promoter activity in *Mycobacterium smegmatis* and score assigned by neural network promoter prediction program, we selected one of these promoter sequences, namely A_{37} for characterization in order to understand the structure of housekeeping promoters of mycobacteria. A_{37}–RNAP complexes were subjected to DNase I footprinting and subsequent mutagenesis. Our results demonstrate that in addition to −10 sequences, DNA sequence at −35 site can also influence the activity of mycobacterial promoters by modulating the promoter recognition by RNA polymerase and subsequent formation of open complex. We also provide evidence that despite exhibiting similarities in −10 and −35 sequences, promoter regions of mycobacteria and *Escherichia coli* differ from each other due to differences in their requirement of spacer sequences between the two positions.

INTRODUCTION

Success of a pathogen lies in its ability to control gene expression under varying environmental conditions. In prokaryotes gene expression is mostly regulated at the initiation of transcription. *Mycobacterium tuberculosis*, one of the most successful and dreaded organisms is responsible for the loss of ∼40 000 human lives every week, globally (1). The genome sequence of *M.tuberculosis* has revealed the presence of 13 different sigma factors (2); some of these have been implicated in the control of pathogenesis of the organism (3–10). Hence, one of the critical factors for the success of *M.tuberculosis* as a pathogen appears to be its ability for differential expression of genes required for its survival under various environmental conditions, it faces during the course of infection. Despite several elegant studies on the expression of genes in mycobacteria, we still lack a clear understanding about the structure of mycobacterial transcriptional elements.

In an earlier study, we had characterized promoter elements of *M.tuberculosis* and *Mycobacterium smegmatis* and showed that fast growing *M.smegmatis* and slow growing *M.tuberculosis* and *Mycobacterium bovis* BCG recognize mycobacterial promoters with similar efficiencies and that the basic transcriptional machineries of these mycobacterial species have common determinants of transcriptional specificity (11). Our study revealed that at −10/extended −10 positions the mycobacterial promoters have sequences similar to their counterparts in *Escherichia coli*; most significantly, the mycobacterial promoters appeared to lack any other canonical sequence(s) except the Pribnow Box. However, deletion of entire promoter sequence, upstream to −10 hexamer in a few of the mycobacterial promoters and consequent loss of their transcriptional activities (∼90%) clearly suggested the requirement of additional upstream sequences for the optimal functioning of these promoters (11). Although, a number of mycobacterial promoters have been reported to posses *E.coli*-like −35 sequences (12), only a few of these exhibit comparable activity in both the organisms (13–23). In fact, in a systematic study on the transcriptional signals of mycobacteria, we had shown that most of the mycobacterial promoters function very inefficiently in *E.coli* (11).

In this study, we have characterized the important regions in the housekeeping mycobacterial promoters. The availability of abundant strong promoter sequences marks the primary requirement for the characterization of important promoter

*To whom correspondence should be addressed. Tel: +91 11 24110970; Fax: +91 11 24115270; Email: akt1003@rediffmail.com
Present address:
Nisheeth Agarwal, Johns Hopkins University, 1550 Orleans Street, Cancer Research Building II, Room No. 1.76, Baltimore, Maryland 21231 (USA)

© 2006 The Author(s).
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

regions. However, in an earlier study it was demonstrated that most of the mycobacterial promoters are weak (24). Hence, in the current study, we first selected sequences similar to those of functional promoters of mycobacteria from a random DNA sequence library. The selection was based on the promoter activity of these DNA sequences in *M.smegmatis*. Out of 475 promoter clones, 40 clones exhibiting the varying promoter activities were sequenced and 26 unique clones were subsequently analyzed by using neural network promoter prediction (NNPP) program, (version 2.2) (http://www.fruitfly.org/seq_tools/nnppAbst.html). One of these promoter-like sequences namely A_{37}, based on its high activity and score was subsequently selected for detailed characterization. Our results demonstrate that both the −10 and −35 conservations are required by mycobacterial promoters for the initiation of transcription and optimal activity. We also provide evidence for the functional differences in the promoter regions of mycobacteria and *E.coli*.

MATERIALS AND METHODS

Strains, plasmids and enzymes

In this study we used *E.coli* strain XL1Blue (*F′*::Tn*10 proA*$^+$*B*$^+$ *lacI*^{q} Δ[*lacZ*] *M15*/*recA1 endA1 gyrA96* [Nal^r] *thi hsdR17* [$r_k^- m_k^+$] *glnV44 relA1 lac*), procured from Stratagene, La Jolla, CA, USA and BL21 (DE3) pLysS (*F- ompT hsdS*$_B$ [$r_B^- m_B^-$] *gal dcm* [DE3] pLysS [Cm^R]), purchased from Novagen, Madison, WI, USA. *M.smegmatis* MC^{2}155 (avirulent, fast growing saprophytic strain of mycobacterium) was kindly provided by Dr S. Bardarov, Albert Einstein College of Medicine, NY, USA. Plasmid pSD5B (mycobacteria-*E.coli* shuttle vector carrying promoter-less *lacZ* gene) was constructed in our laboratory (25) and pET-23b (expression vector with pBR322 plasmid origin of DNA replication and T7 promoter) was obtained from Novagen, Madison, WI, USA. Enzymes and kits were as follows: restriction enzymes, T4 DNA ligase and T4 polynucleotide kinase from New England Biolabs (Beverly, MA, USA); Sequenase version 2.0 kit from United States Biochemical Corp. (Cleveland, OH, USA); *Pfu* DNA polymerase from Stratgene (La Jolla, CA, USA) and DNase I (Grade II) from Roche Applied Science (Indianapolis, IN, USA).

DNA manipulations

PCR amplifications were carried out according to the manufacturer's specifications (Applied Biosystems, Foster City, CA, USA). Each of the 30 cycles was carried out at 94°C for 1 min, 60°C for 1 min and 72°C for 1 min, followed by final extension at 72°C for 10 min. DNA fragments used for cloning and labeling reactions were purified by using GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences, Piscataway, NJ, USA) as per the manufacturer's specifications.

Construction of promoter fragments with random sequences

Promoter fragments were selected by partially annealing oligomers, oligos 1 and 2 (Figure 1). The oligo1 contained a 29 bp long random sequence, flanked by constant sequences, at its 5′ and 3′ ends. The flanking sequence at the 5′ end comprised of a partial TER3 terminator sequence of pSD5B and the flanking sequence at the 3′ end carried the sequence of coding strand of *sigA* gene from position −142 to −163 (with respect to translational start codon). The oligo2 contained the sequence of template strand of *sigA* gene from position −88 to −163 (with respect to translational start codon). Thus, these oligomers had a 22 bp long complementary sequence for hybridization. The template strand of *sigA* gene (in oligo2) carried mutations at positions, −141 and −140 (5′-TA-3′ to 5′-GC-3′), which abolished inherent transcriptional activity of this sequence (data not shown) (Figure 1). Promoter fragments used in the first round of selection were generated *in vitro* by hybridizing these two partially complementary oligonucleotides, oligo1 and oligo2 (Figure 1). Oligonucleotides (1 μM each) were incubated in 40 mM Tris–HCl (pH 7.5), 20 mM $MgCl_2$ and 50 mM NaCl for 5 min at 95°C and then annealed by slow cooling to 30°C for 1 h. The 3′ ends were extended with 0.5 mM dNTPs and 3 U of Sequenase DNA polymerase at 37°C for 1 h. The resulting fragments were extracted with phenol, followed by chloroform. A total of 20 randomly chosen fragments were sequenced after cloning in pGEM3Z vector at XbaI–SphI sites; the frequency of base pairs at each position in the randomized region was approximately equal (data not shown). DNA fragments were end-labeled by PCR amplification by using a set of primers, PrF′ and oligo2, one of which (oligo2) was labeled with 10 μCi [γ-^{32}P]ATP (4000 Ci/mMol) and polynucleotide kinase, according to the manufacturer's recommendations.

In vitro selection

Labeled DNA fragments (~0.5 μM) were incubated with ~1 μM RNA polymerase (purified from *M.smegmatis*, as described earlier (26)). After binding for 10 min at 37°C, heparin was added at a final concentration of 10 μg/ml. After incubation for 10 min at 37°C, RNAP-promoter complexes were separated from unbound DNA on 5% non-denaturing polyacrylamide gels. These complexes were eluted by Tris–HCl (10 mM)/EDTA (1 mM), pH 8.0 and the purified DNA was amplified by PCR by using *Pfu* DNA polymerase. Primers, PrF′ and oligo2 contained all non-randomized promoter positions to reduce the frequency of PCR-generated mutations that might increase promoter binding by RNAP. Amplicons were monitored on gels and the PCR was terminated at appropriate stage to avoid formation of heteroduplexes (27). In the third and subsequent round of selection, the PCR-amplified DNA fragments from the previous round were used and the RNAP-binding reactions were carried out under progressively more stringent conditions [by reducing RNAP concentration to 0.1 μM after second round of binding; 10 nM after fourth round of binding and 1 nM after sixth round of binding (keeping reaction time 10 min), and reducing reaction time to 5 min after sixth round of binding and 2 min after eight round of binding (keeping RNAP concentration 1 nM)]. DNA fragments obtained in each round of selection were subsequently PCR amplified and cloned in pSD5B at its promoter cloning sites. The recombinant vectors were employed to determine the transcriptional

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

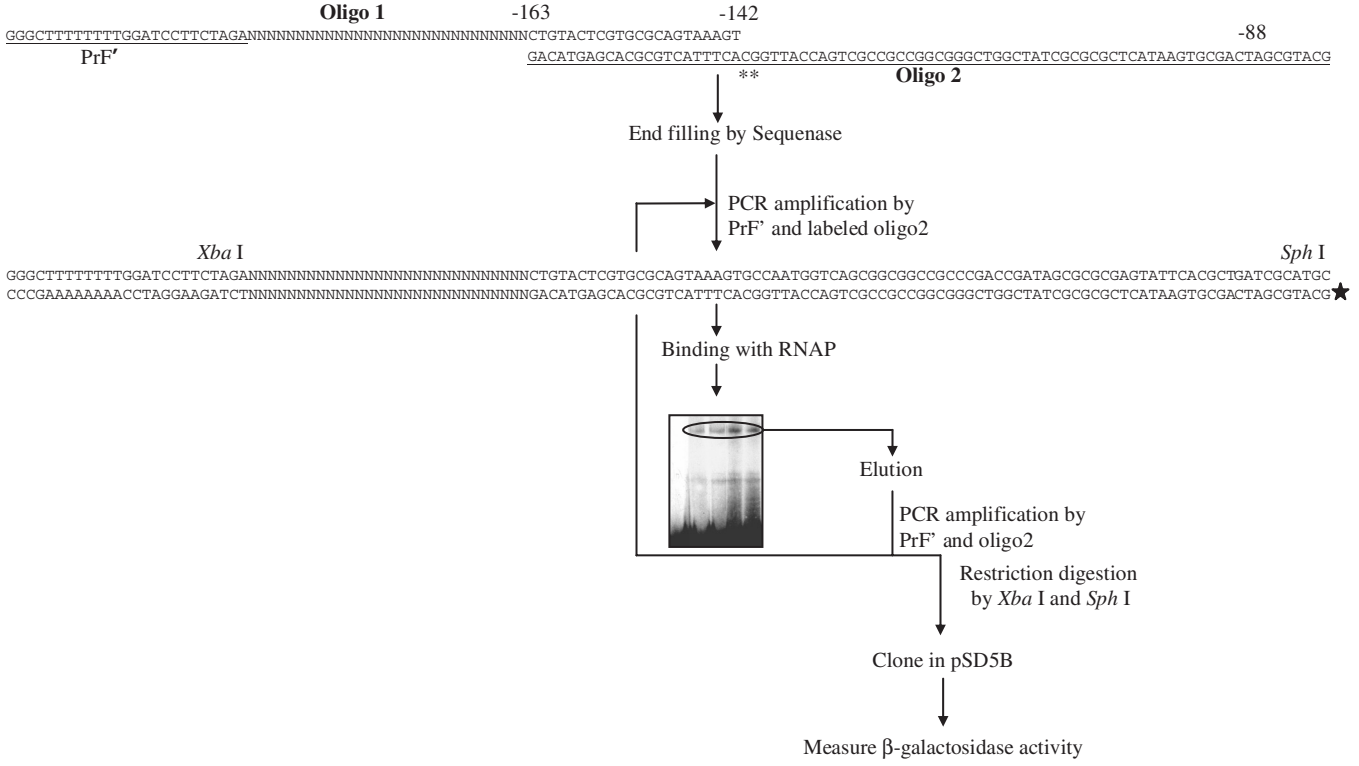

Figure 1. Generation of promoter fragments with random sequences by *in vitro* selection. Oligonucleotides were annealed and extended with Sequenase DNA polymerase to form a library of double-stranded DNA fragments carrying 29 bp long random sequence, flanked by constant sequences, at its 5′ and 3′ ends. The flanking sequence at the 5′ end comprised of a partial TER3 terminator sequence of pSD5B and the flanking sequence at the 3′ end carried the sequence of coding strand of *sigA* gene from position −142 to −163 (with respect to translational start codon). The oligo2 contained the mutant sequence (positions of mutations are marked by small asterisks) of template strand of *sigA* gene from position −88 to −163 (with respect to translational start codon). These oligonucleotides were designed in such a way that the resulting double-stranded DNA fragments contained an XbaI site upstream to random sequence and an SphI site downstream to it. The end-labeled DNA fragments were obtained by subjecting the double-stranded DNA fragment to PCR amplification by using primer PrF′ and labeled primer oligo2. Primers, PrF′ and oligo2, were designed in such a way that the amplicon carried sufficient bases (76 bases) downstream to the random sequence to ensure that the selected promoter sequences have sufficient bases downstream to the TSP. These end-labeled DNA fragments were subjected to EMSA by allowing them to form the complex with RNAP of *M.smegmatis*. The DNA fragments were subsequently eluted from the gel and re-amplified by PrF′ and oligo2. PCR amplicon was subsequently used for either its cloning into the pSD5B (after digestion by XbaI and SphI) or as a template for second round of PCR amplification using primer PrF′ and labeled primer oligo2. β-Galactosidase activity of the promoter clones was determined, as described in Materials and Methods. Bold star represents the labeled end of promoter DNA fragment.

activity of cloned DNA fragments in *M.smegmatis* (Figure 1). Ten rounds of *in vitro* selection were carried out, as after tenth round we could not monitor any complex on the gel (data not shown).

Construction of mutants

DNA fragments carrying the mutations and deletions/insertions in promoter sequences were constructed by two-step PCR amplification using overlapping modified primers, as described earlier (28). After cleavage with XbaI–SphI, these fragments were cloned in pSD5B and the recombinant vectors were used to transform *E.coli* XL1Blue and *M.smegmatis* MC^2155 strains, respectively, and the transformants were grown on Luria–Bertani plate containing kanamycin (25 μg/ml) and X-gal (60 μg/ml). Nucleotide sequence of DNA fragments was determined by the Sequenase version 2.0 according to the manufacturer's instructions.

Primer extension

The transcription start point (TSP) of A_{37} was determined by isolating RNA from *M.smegmatis*, harboring pSD5B-A_{37}. Primer extension was carried out by using the primer PrlacZ (Figure 4), as described earlier (11).

Electrophoretic mobility shift assay (EMSA)

EMSA was performed typically as described earlier (26), except whenever needed, heparin was added (10 μg/ml) after 10 min of binding of RNAP with promoter fragment. All the binding reactions were performed with PCR-amplified labeled promoter fragments as template. The DNA fragments were PCR amplified by using primers, PrF (labeled by using the [γ-^{32}P]ATP and T4 polynucleotide kinase, according to the manufacturer's specifications) and PrR (Figure 4A), resulting in a 217 bp amplicon, which contained 142 bp upstream of the TSP (+1) of A_{37} and 75 bp downstream of +1 (Figure 4C).

DNase I footprinting

DNase I footprinting of the RNAP–A_{37} promoter complexes was performed with 217 bp long PCR amplicon, labeled at single end as described above. Footprinting was carried out as described earlier (26).

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

Determination of promoter activity by β-galactosidase assay

Promoter activity was determined by conducting the β-galactosidase assay, as described by Miller (29).

RESULTS

Selection of strong transcriptional signals of mycobacteria

We made an attempt to generate the strong promoter sequences of mycobacteria by following a strategy, similar to saturation mutagenesis (30). However, owing to lack of enough information about the sequence requirements by mycobacterial transcriptional machinery, we restricted the length of random nucleotides to 29 bases (Figure 1), which represent the number of bases acquired between −10 and −35 hexameric sequences ([6 × 2] + 17) in other prokaryotic systems (31,32). However, the random DNA sequence library generated by us represents relatively small fraction of possible DNA sequences at the site of randomization, as compared to the one exhibiting 95% coverage (which would need enormous starting DNA material (>5 mM), and thus may not be experimentally feasible). Despite this caveat, as described below, we were able to select several strong transcription signals of mycobacteria.

The initial selection of strong promoter-like sequences was based on the ability of these random DNA fragments to bind with RNAP and form open complexes under varying stringencies in the presence of heparin, as described in Materials and Methods. Subsequently, by using β-galactosidase-based reporter plasmid the transcriptional activities of ~475 selected DNA sequence clones were determined in *M.smegmatis*, which varied from ~900 to ~39 000 nmol/min/mg (β-galactosidase activities; data not shown). For further analysis, 40 clones were selected for nucleotide sequencing which included all the clones exhibiting greater than ~2000 nmol/min/mg activity [which represents the activity of the constitutive *sigA* promoter (Figure 7A)], and some clones with the promoter activity of lower than 2000 nmol/min/mg. As shown in Figure 2, 26 out of 40 clones exhibited unique sequences.

Promoter prediction by NNPP program and sequence analysis

To define the promoter elements, the sequence of the above 26 clones was analyzed by NNPP program (http://www.fruitfly.org/seq_tools/nnppAbst.html). Each of the DNA sequences containing 29 bp long unique sequence, flanked by 15 bp long constant sequences, at its 5′ (5′-TTGGATCC-TTCTAGA-3′) and 3′ (5′-CTGTACTCGTGCGCA-3′) ends was subjected to promoter prediction analysis. The NNPP

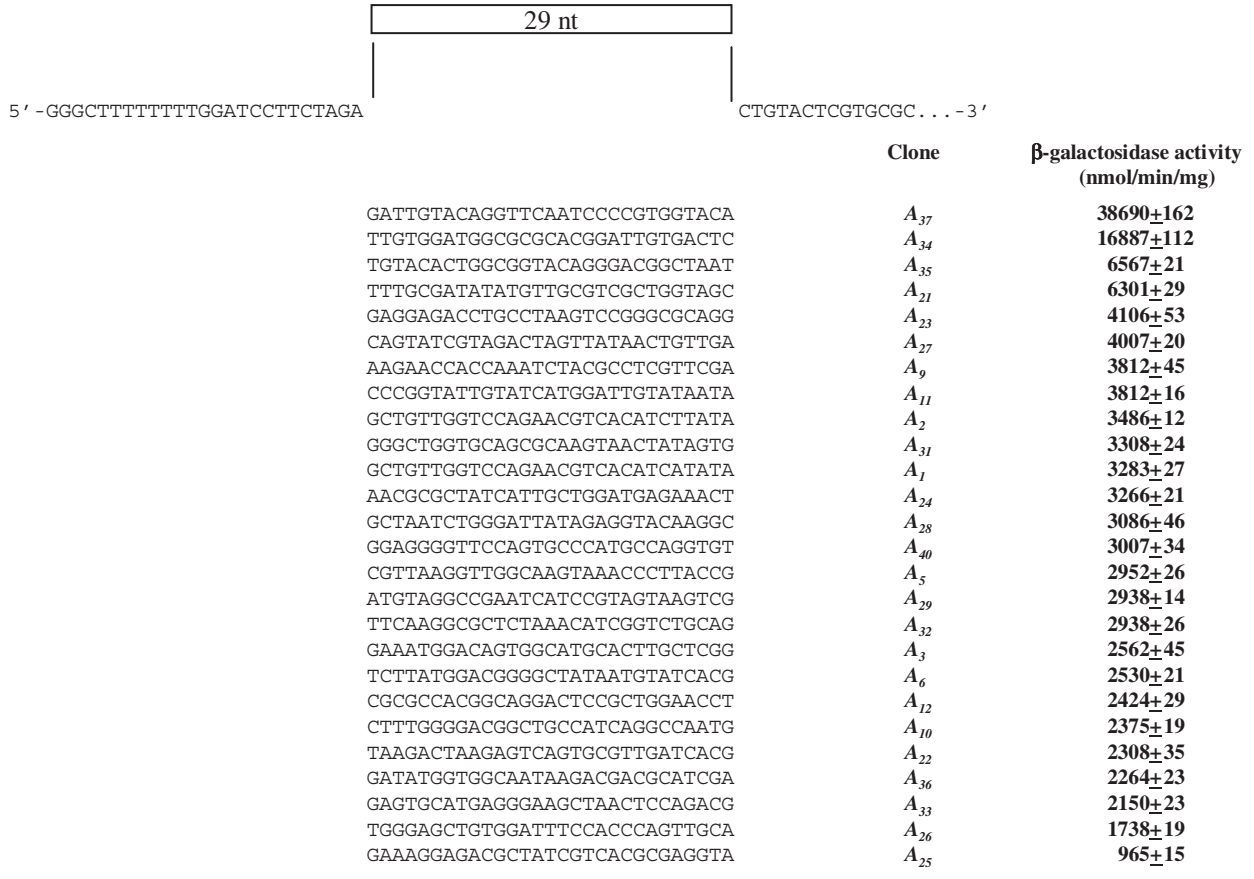

Figure 2. Analysis of unique promoter clones. The open box represents the position of 29 bp long unique sequence in each promoter clone, flanked by constant sequences, at its 5′ and 3′ ends, as described in Figure 1. Promoter activity was measured as β-galactosidase specific activity. The values represent mean of three separate assays.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

-35 -10 +1	Clone	Score
5′-TAGAT**TTGCGA**TATATGTTGCGTCGC**TGGTAGCCT**GTACT**C**GTGCGCA-3′	A_{21}	0.95
5′-TAGAGA**TTGTAC**AGGTTCAATCCCCG**TGGTACACT**GTACT**C**GTGCGCA-3′	A_{37}	0.93
5′-GATCC**TTCTAG**ACAGTATCGTAGACTAGT**TATAAC**TGTTG**A**CTGTACTCG-3′	A_{27}	0.71
5′-TTCTAG**ACCCGG**TATTGTATCATGGATTG**TATAAT**ACTGT**A**CTCGTGCGC-3′	A_{11}	0.68
5′-TCTAGA**AACGCG**CTATCATTGCTGGA**TGAGAAACT**CTGTA**C**TCGTGCGCA-3′	A_{24}	0.61
5′-TTCTAG**AGGGCT**GGTGCAGCGCAAGTAAC**TATAGT**GCTGT**A**CTCGTGCGC-3′	A_{31}	0.57
5′-TAG**ATGTAC**ACTGGCGGTACAGGGACGGC**TAATCT**GTACT**C**GTGCGCA-3′	A_{35}	0.50
5′-TAGAGCTG**TTGGTC**CAGAACGTCACATCA**TATACT**GTACT**C**GTGCGCA-3′	A_{01}	0.48
5′-TAGAGCTG**TTGGTC**CAGAACGTCACATCT**TATACT**GTACT**C**GTGCGCA-3′	A_{02}	0.44
5′-TTCTAG**ATCTTA**TGGACGGGGCTATAATG**TATCAC**GCTGT**A**CTCGTGCGC-3′	A_{06}	0.26

Figure 3. Promoter prediction by the NNPP program. Each of the DNA sequences containing 29 bp long unique sequence, flanked by 15 bp long constant sequences, at its 5′ and 3′ ends was used for the promoter prediction by NNPP program by using minimum promoter score >0.25. +1 indicates the putative TSP, identified by the NNPP program. The values represent the score assigned to each DNA sequence.

program recognized 10 out of 26 sequences as having a definitive TSP with a score of >0.25 (Figure 3). Moreover, these promoter sequences appeared to contain the conserved bases, 5′-WTS-3′ at −35 and 5′-TATACT-3′ at −10 positions, with respect to their putative TSPs and exhibited strong activities ranging from ∼3000 to 39 000 nmol/min/mg (Figures 2 and 3). Promoters, A_{21}, A_{24} and A_{37} also contained an extended −10 motif. These results indicate that despite using the incomplete DNA sequence library, we were able to select a few strong promoter sequences preferred by mycobacterial RNAP. However, as shown in Figures 2 and 3, activities of these promoter sequences were not corresponding to the scores assigned by the program. Therefore, we selected one of the promoter sequences, A_{37} for further characterization on the basis of its highest activity and the score >0.9, assigned by NNPP program.

To experimentally define the promoter elements of A_{37}, it was subjected to primer extension analysis, which revealed that transcription from this promoter starts from a purine residue G near the position predicted by program (Figure 4B). Analysis of the promoter sequence revealed the presence of an extended −10 sequence 5′-TGGTACACT-3′, 6 bp upstream to the +1 site (Figure 4C).

DNase I footprinting analysis of A_{37}–RNAP complex

In order to characterize the base positions in A_{37} promoter required for the binding of mycobacterial RNAP during the binary complex formation, the RNAP–A_{37} promoter complexes were subjected to DNase I treatment and the gel band intensities were analyzed by densitometric scanning by using NIH*Image* program. As shown in Figure 4D, the holoenzyme partially protected the sequences around −10 position and exhibited complete protection of the bases 5′-ACAGG-3′, near −35 position (from −31 to −27 positions). The binding of RNAP also revealed a few hyperactive sites at positions −24, −25, −34, −38, −41 and −43 (Figure 4D). Protection of bases near −10 and −35 positions of A_{37} implicated that these promoter regions could play an important role in the recognition of RNAP.

Base substitutions at −35 region and its influence on promoter functions

Based on the DNase I footprinting results, we evaluated the effect of base substitutions from positions −35 to −27 on the activity of A_{37} promoter. Since the presence of TGN motif in case of mycobacteria may mask the contribution of −35 region in the promoter activity (26), we first mutated the TGN motif of A_{37} and the subsequent mutations in the −35 region were carried out with $A_{37TG\text{-}}$. In the preliminary studies it was observed that substituting T at position −35 by G [$A_{37TG\text{-}}$($_{-35}$T-G)], TT at positions −35 and −34 by GG [$A_{37TG\text{-}}$($_{-35\text{-}34}$TT-GG)], and GT at positions −33 and −32 by CC [$A_{37TG\text{-}}$($_{-33\text{-}32}$GT-CC)] resulted in a substantial reduction in the $A_{37TG\text{-}}$ promoter activity. However, substitutions of bases at positions −31, −30, −29, −28 and −27 (A31C30 by GG [$A_{37TG\text{-}}$($_{-31\text{-}30}$AC-GG)], A29 by C [$A_{37TG\text{-}}$($_{-29}$A-C)] and G28G27 by TT[$A_{37TG\text{-}}$($_{-28\text{-}27}$GG-TT)]) exhibited only moderate effect on the promoter activity in *M.smegmatis*. Quantitative estimations indicated that the activities of ($A_{37TG\text{-}}$($_{-35}$T-G), $A_{37TG\text{-}}$($_{-35\text{-}34}$TT-GG) and $A_{37TG\text{-}}$($_{-33\text{-}32}$GT-CC) were reduced by 3-, 4.5- and 3-fold, respectively, in comparison with the activity of wild-type promoter ($A_{37TG\text{-}}$) while the activities of $A_{37TG\text{-}}$($_{-31\text{-}30}$AC-GG), $A_{37TG\text{-}}$($_{-29}$A-C) and $A_{37TG\text{-}}$($_{-28\text{-}27}$GG-TT) were similar to the activity of wild-type promoter (Figure 5A).

In order to analyze the effect of these substitutions at molecular level, both the wild-type promoter $A_{37TG\text{-}}$ and its derivatives were subjected to binding with RNAP. As shown in Figure 5B, mutations of bases from position −35 to −32 resulted in a significant reduction in binding (Lanes 2, 4 and 5). It was also observed that mutation of A residue at position −29 exhibited no effect on the binding of RNAP (Lane 3). Hence, these results were in agreement with the *in vivo* activities exhibited by the mutant promoters and suggested that −35 region of mycobacterial promoter is an important determinant of its activity.

Determination of the optimal base requirement at the −35 region

To determine if there is a sequence preference in the −35 promoter element, we examined each position in the −35 promoter region from positions −35 to −30, comparing the effect of each of the four possible nucleotides on promoter activity. The activity of these 24 promoter derivatives (6 positions × 4 bases) was measured in *M.smegmatis*. As shown in Figure 6, mutations at positions −35 to −30 exhibited substantial effect on promoter activity and revealed that the sequence $^{-35}$TTGCGA^{-30} represented the optimal base requirement at the −35 region of the $A_{37TG\text{-}}$ promoter.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

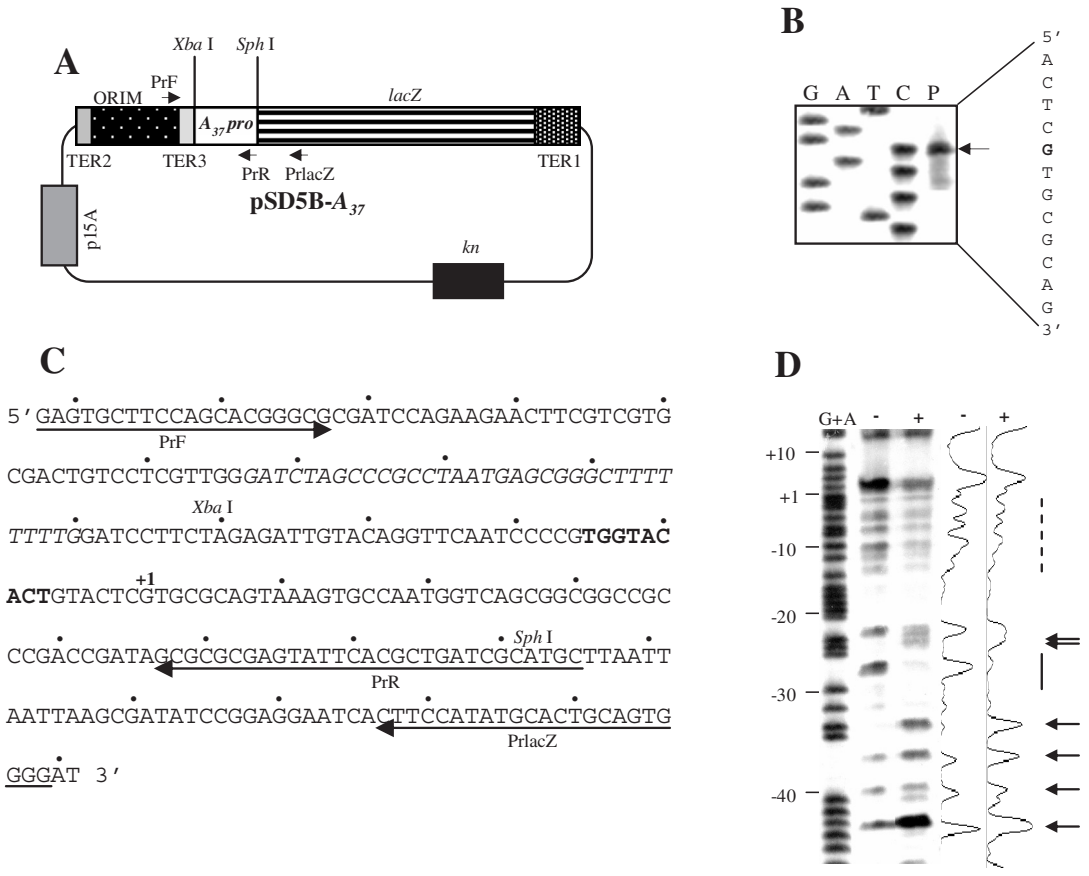

Figure 4. Analysis of A_{37} promoter. (**A**) Schematic illustration of pSD5B-A_{37} promoter clone. Relevant restriction sites, positions of various primers (PrF, PrR and PrlacZ) and antibiotic resistance marker (*kn*) are shown. TER1, TER2 and TER3 represent the transcriptional terminators, as described previously (24,25). (**B**) Determination of TSP of A_{37}. Primer extension reaction was carried out with end-labeled primer PrlacZ and RNA isolated from the *M.smegmatis* cells harboring pSD5B-A_{37} vector. The reactions were analyzed on urea–6% polyacrylamide gels. Sequencing reactions were also performed with PrlacZ and run alongside the corresponding primer extension reactions (lane P). Arrow represents position of TSP, which is shown as a bold letter in the sequence. (**C**) Sequence analysis of A_{37} promoter containing region of pSD5B-A_{37}. +1 represents TSP of A_{37}. Sequences written in bold indicate position of −10 sequence of A_{37}. Horizontal arrows represent the sequences of PrF, PrR and PrlacZ primers, respectively. Sequence written in italics represents the transcriptional terminator, TER3. Positions of restriction sites flanking the promoter region are shown. (**D**) DNase I protection assay using RNAP bound to the A_{37} promoter. The gel electrophoretic pattern and the corresponding densitometric scan of the coding strand are shown. Arrows indicate the positions of DNase I hyperactive sites due to the binding of RNAP. The numbers indicate the positions of respective bases. The solid vertical bar indicates the positions on the promoter region that are completely protected; the interrupted bar represents the sequence that shows partial protection. G+A represents sequencing ladder by the Maxam–Gilbert method; '+' and '−' represent DNase I reactions with and without RNAP, respectively. Densitometric scanning was performed by using NIH *Image* program.

Generality of the −35 sequence for promoter functions

Since, the −35 sequence, 5′-TTGCGA-3′ was derived, based on mutations at individual positions, we next wanted to evaluate the authenticity of this sequence and its contribution in the enhancement of promoter activity. Hence, we replaced the −35 sequence of A_{37TG-} with a sequence 5′-TTGCGA-3′. As shown in Figure 7A, the resulting promoter A_{37TG-}*con* exhibited an ∼2.5-fold enhancement in its activity in comparison with the activity of A_{37TG-} (Figure 7A). Similarly, we also substituted the −35 region sequence of another moderately active mycobacterial promoter *sigA* by the sequence, 5′-TTGCGA-3′. The resulting promoter *sigAprocon* exhibited a ∼6-fold higher activity in comparison with the activity of *sigA* (Figure 7A). These results suggested that 5′-TTGCGA-3′ could serve as an optimal −35 sequence for mycobacterial promoters.

To analyze the effect of the sequence 5′-TTGCGA-3′ on promoter recognition by RNAP, we compared the binding affinity of RNAP to A_{37TG-} and *sigA* promoters with their derivatives, A_{37TG-}*con* and *sigAprocon*, respectively, by EMSA. Different concentrations of RNAP were mixed with ^{32}P-labeled promoter fragments at 37°C to allow complex formation. The results showed concentration-dependent binding of RNAP with all the four promoter fragments (Figure 7B). However, promoter derivatives with the modified sequence, 5′-TTGCGA-3′ at −35 positions exhibited higher affinity than the corresponding wild-type promoters (Figure 7B). These results indicated the preference of mycobacterial RNA polymerase for the sequence 5′-TTGCGA-3′ at −35 region.

In order to determine the global distribution of this optimal −35 sequence in the 5′-untranslated regions (5′-UTRs) of *M.tuberculosis* genome, we analyzed putative promoter sequences in 5′-UTRs, up to 300 bp from the translational start site. Pattern search analysis of the 5′-UTRs of *M.tuberculosis* genome using the TubercuList

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

A

Sequence (-30 -20 -10 +1)	Clone	β-galactosidase activity (nmol/min/mg)
GATTGTACAGGTTCAATCCCCGTGGTACACTGTACTCG	A_{37}	38690±162
----------AC----------	A_{37TG-}	9909±110
--G----------AC----------	A_{37TG-}($_{-35}$T-G)	3625±45
--GG----------AC----------	A_{37TG-}($_{-35-34}$TT-GG)	2275±21
----CC----------AC----------	A_{37TG-}($_{-33-32}$GT-CC)	3229±34
------GG----------AC----------	A_{37TG-}($_{-31-30}$AC-GG)	7888±89
--------C----------AC----------	A_{37TG-}($_{-29}$A-C)	9001±97
---------TT----------AC----------	A_{37TG-}($_{-28-27}$GG-TT)	8984±129

B

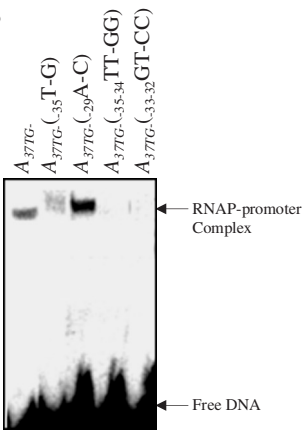

Figure 5. Base substitutions at −35 region and its influence on promoter functions. (**A**) Effects of base substitutions on the activity of A_{37TG-} promoter. Promoter activity was measured as β-galactosidase specific activity. The cell lysate of pSD5B-transformed *M.smegmatis* (exhibiting an activity of 12 ± 5 nmol/min/mg; data not shown) was used as a negative control. The values represent mean of three separate assays. '-' represents similar bases at the corresponding position. The numbers above promoter sequence indicate the positions of respective bases. (**B**) EMSA with wild-type and mutant A_{37TG-} promoter fragments. Binding reactions containing 1.5 nM of radiolabeled promoter DNA fragment and 0.1 μM RNA polymerase were carried out at 37°C for 10 min. The reactions were terminated by addition of 10 μg/ml heparin for 2 min. After incubation, samples were loaded on 4% polyacrylamide (30:1 acrylamide/Bis) gel.

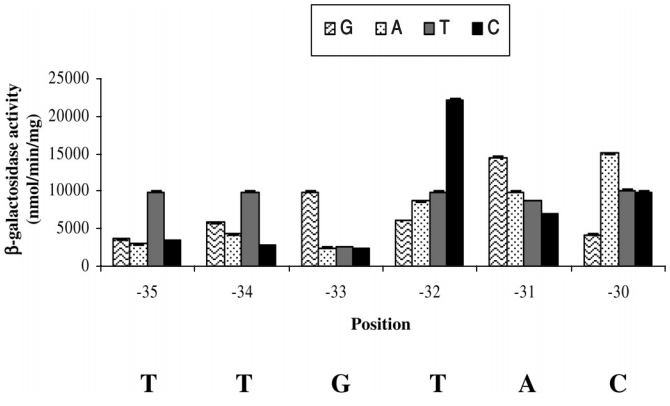

Figure 6. Effects of single-base substitutions on the activity of A_{37TG-} promoter in *M.smegmatis*. All three possible substitutions were carried out at each position from −35 to −30. The wild-type sequence at each position is shown below the figure. Promoter activity was measured as β-galactosidase specific activity. The cell lysate of pSD5B-transformed *M.smegmatis* (exhibiting an activity of 12 ± 5 nmol/min/mg; data not shown) was used as a negative control. The values represent mean of three separate assays.

(http://genolist.pasteur.fr/TubercuList/) website for the presence of a sequence 5′-TTGCGA-$N_{18\pm1}$-TANNNT-3′ revealed 3 perfect matching sequences (exhibited by 5′-UTRs of genes *pks11*, Rv3839 and *mmsA*) and 112 sequences with one mismatch in either of putative −35 and −10 hexamers (Table 1). In addition, majority of these sequences (47 out of 115) exhibited a distance of 18 bp between putative −35 and −10 hexamers (Table 1). To analyze the effects of base substitution at −35 position in the activity of these promoter-like sequences, we randomly selected the 5′-UTRs of two genes, *mmsA* and *gcvH* (Table 1). The DNA sequence of each of these 5′-UTRs encompassing the minimal promoter region from position −35 to +10 (based on the respective putative TSPs) was cloned in pSD5B at the promoter cloning sites, XbaI and SphI and the promoter activity was determined in *M.smegmatis*, as described above. As shown in Figure 7A, both these promoters exhibited weak activities in *M.smegmatis* (224 nmol/min/mg by *mmsA* and 169 nmol/min/mg by *gcvH*). By replacing the −35 sequence by a non-identical sequence, there was a significant reduction

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

A

Sequence (-30 -20 -10 +1)	Clone	β-galactosidase activity (nmol/min/mg)
GATTGTACAGGTTCAATCCCCGACGTACACTGTACTC**G**	A_{37TG-}	9909±110
GATTGCGAAGGTTCAATCCCCGACGTACACTGTACTC**G**	A_{37TG-}*con*	25393±214
GGGCTGTACTCGTGCGCAGTAAAGTTACAATGGTCAG**C**	*sigApro*	1997±15
GGTTGCGACTCGTGCGCAGTAAAGTTACAATGGTCAG**C**	*sigAprocon*	11143±119
GATTGCGAGCCATCCGCCACAGTTACTAGGATATCCAA	*mmsA*	224±45
GAAACCGAGCCATCCGCCACAGTTACTAGGATATCCAA	$mmsA_{\Delta-35}$	105±37
GATTGCGGCAACCAAAACAGCACCAATACGCTCAGCCC	*gcvH*	169±31
GAAACCGGCAACCAAAACAGCACCAATACGCTCAGCCC	$gcvH_{\Delta-35}$	79±23

B

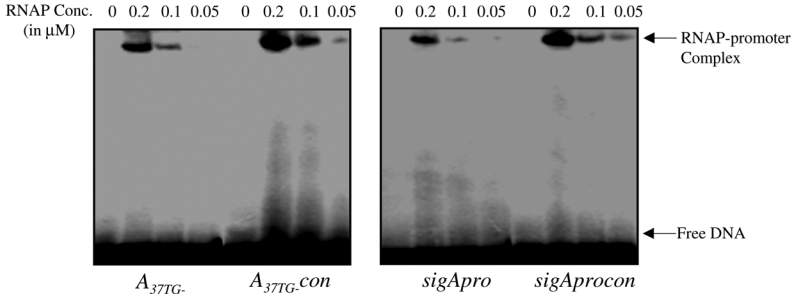

Figure 7. Functional analysis of promoter derivatives with altered sequence in the −35 region. (**A**) Comparative analysis of the activities of wild-type and mutant derivatives of mycobacterial promoters, respectively. *sigA* promoter fragment contains the promoter sequence from position −143 to +42 with respect to its TSP and both the *mmsA* and *gcvH* promoter fragments contain the promoter sequence from position −35 to +10 with respect to their TSPs. Promoter activity was measured as β-galactosidase specific activity. The cell lysate of pSD5B-transformed *M.smegmatis* (exhibiting an activity of 12 ± 5 nmol/min/mg; data not shown) was used as a negative control. The values represent mean of three separate assays. Underlined sequences represent the bases at −35 and −10 positions. The numbers above promoter sequence indicate the positions of respective bases. Boldface letters at +1 indicate the experimentally defined TSPs. (**B**) EMSA with wild-type and mutant derivatives of A_{37TG-} and *sigA* promoters, respectively. Binding reactions containing 1.5 nM of radiolabeled promoter DNA fragment and different concentrations of RNA polymerase were carried out at 37°C for 10 min. After incubation, samples were loaded on 4% polyacrylamide (30:1 acrylamide/Bis) gel.

in the activities of the respective promoters (105 nmol/min/mg by *mmsA* and 79 nmol/min/mg by *gcvH*, Figure 7A). These results further confirm the premise that the −35 sequence can influence the activity of mycobacterial promoters by modulating the promoter recognition by RNA polymerase and subsequent formation of open complex.

Determination of optimum distance between −35 and −10 sequences of mycobacterial promoters

Although the sequences between −35 and −10 regions may not directly contribute towards the specificity of recognition of promoter elements by RNAP, the placement of two sites at an appropriate distance is critical for the geometry of RNAP and subsequent recognition by RNAP (33). In order to determine the optimal intervening distance, spacer length was modulated from 15 to 19 bp (The distance between –35 and –10 sequences in A_{37TG-}*con* was 17 bp). As shown in Figure 8, the distance was altered by deleting or inserting bases at the center of spacer. Moreover, we inserted random bases in the spacer (Figure 8) and observed that the effect was not sequence specific (data not shown). Promoter derivatives with 15 or 19 bp long spacer sequences exhibited a drastic reduction in their activities (1912 and 5757 nmol/min/mg, respectively) in comparison with activity of the wild-type promoter (25 393 nmol/min/mg). Promoter derivative with 16 bp long spacer sequence exhibited significant promoter activity (11 278 nmol/min/mg); however, the highest activity was exhibited by promoter derivative with 18 bp spacer (36 309 nmol/min/mg).

Analysis of the functional differences between mycobacterial and *E.coli* promoters

Since, the overall architecture of mycobacterial and *E.coli* promoters appeared similar, we compared the activity of one of the derivatives of A_{37TG-}, A_{37TG-}(*conE.coli*) (which has *E.coli* consensus sequence 5′-TTGACA-3′ at −35 position with a spacer length of 16 bp) in *M.smegmatis* and *E.coli*. A_{37TG-}(*conE.coli*) exhibited a specific activity of

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

Table 1. Analysis of putative promoter sequences in 5′-UTRs of mycobacterial genome

Position from ATG	−35	−10	Gene
Promoter sequences with 17 bp spacer			
−70	5′-**TTGCGA**GATCGCGTAATGTGGCT	**TAAACT**-3′	**pks11**
−57	5′-**TTGCGA**CCAGACCGTGCATTAGT	**TAACTT**-3′	**Rv3839**
−234	5′-**TTGCGA**TGCGGTTCCAGTACGAA	**TCCCGT**-3′	**pntAa**
−32	5′-**TTGCGA**ACCAAGTCGCGAGTCGAG	**AGGGT**-3′	**Rv0157A**
−76	5′-**TTGCGA**GCAGACCCCGCCAGCTCG	**ACCGT**-3′	**galE2**
−235	5′-**TTGCGA**ATTATTGGTCAGATCGG	**TACCAC**-3′	**Rv0508**
−214	5′-**TTGCGA**GCGCAGCTGCGGCACGG	**TGGCAT**-3′	**Rv0897c**
−257	5′-**TTGCGA**CACGTGGCCGATCTTGT	**TATGAC**-3′	**Rv1208**
−189	5′-**TTGCGA**CGGTTGAGCATGCTGGT	**TGCGGT**-3′	**Rv1290A**
−97	5′-**TTGCGA**ACGGGTGGCATCGAAATC	**AACTT**-3′	**PE19**
−89	5′-**TTGCGA**AAGTATTACCGTCGGGT	**TTTGGT**-3′	**Rv1815**
−294	5′-**TTGCGA**TAACCCAACTCTCATGG	**TGCTTT**-3′	**Rv2298**
−190	5′-**TTGCGA**TGAGCAGCGTGGCCGAC	**TATCTC**-3′	**PPE40**
−231	5′-**TTGCGA**CGCCACGGTGCCCGTCGG	**ATGGT**-3′	**ggtB**
−168	5′-**TTGCGA**AGTCCAATCGCAAATAT	**TACGAA**-3′	**Rv2723**
−151	5′-**TTGCGA**ATCCTGTTGGGCCGCCCC	**AGTTT**-3′	**Rv2808**
−141	5′-**TTGCGA**CCAGTTCAACCTGTGCG	**TAATGG**-3′	**Rv3088**
−57	5′-**TTGCGA**ACCCAACTGGTCGCGGCT	**CCACT**-3′	**rmlB**
−141	5′-**TTGCGA**ATGTGACGGTAGATAGAG	**AAGCT**-3′	**mmpL8**
−120	5′-**TTGCGA**CGCGAGCGGTCATGGAGC	**AGGGT**-3′	**sodA**
−69	5′-**TTGCG**CAAGACTGCGGCGGAATC	**TAATAT**-3′	**Rv0331**
−97	5′-**TTGCG**CGGGTCGCCACACGCCCG	**TATCTT**-3′	**Rv0496**
−37	5′-**TTGCG**GAAGAATATCCGTAAGTC	**TAAACT**-3′	**Rv2034**
−37	5′-**TTGCG**GCGGATTCTCGCGCTGCA	**TACCGT**-3′	**Rv3447c**
−87	5′-**TTGCCA**TTAAACTGCACCCACGC	**TACCGT**-3′	**rpmB1**
−69	5′-**TTGCAA**TAGCAAGCACTTGGACG	**TACTCT**-3′	**Rv0474**
−268	5′-**TTGCCA**TGTTCTCACCAAGGGCG	**TACCGT**-3′	**Rv0654**
−119	5′-**TTGTGA**TTGGCGTGGGGGGGCGC	**TAACGT**-3′	**Rv0302**
−35	5′-**TTGTGA**TCCGTATTTCGGTTCAC	**TATATT**-3′	**Rv1773c**
−64	5′-**TAGCGA**AAACTGGGAATTTCCCC	**TACGTT**-3′	**Rv2331**
−243	5′-**TCGCGA**CGATATTTACGGTGCGC	**TACGGT**-3′	**Rv2475c**
−41	5′-**TGGCGA**GCTTCGTGCGTGTTCGG	**TAGCCT**-3′	**metU**
−294	5′-**GTGCGA**AAGAGCCGTCATTCGCC	**TAGAAT**-3′	**PE8**
Promoter sequences with 18 bp spacer			
−78	5′-**TTGCGA**GCCATCCGCCACAGTTAC	**TAGGAT**-3′	**mmsA**
−70	5′-**TTGCGA**GATCGCGTAATGTGGCTT	**AAACTT**-3′	**pks11**
−125	5′-**TTGCGA**GGGTCGGTTCGCGGCCCCA	**ACGCT**-3′	**pknF**
−230	5′-**TTGCGA**TGATCGCGCTGACGGGCTC	**AGCAT**-3′	**PPE28**
−274	5′-**TTGCGA**GTTCCGTTGCAATTTCGG	**TGTATT**-3′	**hisE**
−57	5′-**TTGCGA**TGCCCGCCGTTACCCTGTG	**ACTGT**-3′	**cyp128**
−46	5′-**TTGCGA**TCCCATAGCTTTTGGGGT	**TGAGGT**-3′	**Rv2285**
−127	5′-**TTGCGA**GCTATATCGCGGCGCTGT	**TGTGGT**-3′	**Rv2823c**
−144	5′-**TTGCGA**CGGGCGTGGGCGGCACCG	**TGGTAT**-3′	**Rv2912c**
−69	5′-**TTGCGA**GTTCGCTGAAGCAAGCAC	**TGCAGT**-3′	**Rv3114**
−181	5′-**TTGCGA**CCGAGAAGCTGAGCGACG	**TATTCG**-3′	**Rv3173c**
−156	5′-**TTGCGA**CCGAAAAGGTCACCCGCG	**TCCTGT**-3′	**Rv3269**
−33	5′-**TTGCGA**TGATTTTCCAACCGGGAGG	**AGAGT**-3′	**Rv3389c**
−76	5′-**TTGCGA**CGCATGTTCTTCACCGTC	**TATCCA**-3′	**cut3**
−43	5′-**TTGCGA**GACGGGCATATCGCCCAC	**TGTAGT**-3′	**cyp142**
−247	5′-**TTGCGA**TCGAGCTGGCCTGCCCCC	**TGGTTT**-3′	**fadE33**
−79	5′-**TTGCGA**TCGCCACGAGCGGTAACC	**TGGACT**-3′	**ftsH**
−141	5′-**TTGCGA**ATGTGACGGTAGATAGAG	**AAGCTT**-3′	**mmpL8**
−260	5′-**TTGCGA**CGCAGCATCGTGGTGCTC	**AACGAT**-3′	**Rv3861**
−80	5′-**TTGCGA**GATTGGCTGCCCGCATCC	**AATGAT**-3′	**Rv0090**
−150	5′-**TTGCGA**TGACCCGAAAAAGCCAGG	**TAGACG**-3′	**Rv0293c**
−49	5′-**TTGCGA**ACACATCGTTGATGGCCG	**TACTCG**-3′	**YrbEe2A**
−113	5′-**TTGCGA**TCCGCGGCGGGTCTATGG	**TGCCAT**-3′	**Rv0894**
−298	5′-**TTGCG**GACAACGTGCAACGGTTCG	**TACAGT**-3′	**Rv0615**
−279	5′-**TTGCG**TGTGCATCATCTGGCAGGC	**TATGTT**-3′	**Rv1510**
−82	5′-**TTGCG**GCAACCAAAACAGCACCAA	**TACGCT**-3′	**gcvH**
−43	5′-**TTGCG**CTTGAGCCAGCTAGGCGGG	**TAAATT**-3′	**Rv2657c**
−300	5′-**TTGCG**CTTGAGCCAGCTAGGCGGG	**TAAATT**-3′	**Rv2656c**
−70	5′-**TTGCG**CGGCAGGACTTTTGGGCCC	**TAGCTT**-3′	**pknI**
−213	5′-**TTGCG**GTACCGGCGGTACCGGCTT	**TAGTGT**-3′	**Rv3338**
−193	5′-**TTGCG**CAGAGCACGGCGATGCGGG	**TAATGT**-3′	**infA**
−64	5′-**TTGCG**CTCACCGGAGAGAAAACTG	**TAACGT**-3′	**Rv3526**
−111	5′-**TTGCCA**CTACCATCCGGTTGCGGA	**TAACGT**-3′	**Rv0950c**

Table 1. *Continued*

Position from ATG	−35	−10	Gene
−66	5′-**TTGCCA**CCGCCATTCTGCGTCGGT	**TAGTCT**-3′	**Rv1490**
−153	5′-**TTGCTA**AATTAGGGCAGCCTTGCC	**TATCAT**-3′	**Rv1847**
−54	5′-**TTGAGA**GTCGGGCGCCGGTTGCGG	**TAATCT**-3′	**ileT**
−279	5′-**TTGAGA**GTCGGGCGCCGGTTGCGG	**TAATCT**-3′	**alaT**
−141	5′-**TTGTGA**TCGAACGGGTGGGGCCTC	**TAGGTT**-3′	**rpfB**
−226	5′-**TTACGA**TCACCAGCTGGTGGGGCT	**TATGGT**-3′	**trmD**
−203	5′-**TTACGA**CTTCGGCGGCCGGTCCAA	**TATCGT**-3′	**tcrY**
−89	5′-**TGGCGA**TAACGCCGCCTCCGGAGG	**TACGTT**-3′	**Rv0193c**
−36	5′-**TCGCGA**CTTGTGTGACCGACAGGC	**TACGTT**-3′	**Rv0938**
−81	5′-**TAGCGA**TGTTTTTGCGAGATCGCG	**TAATGT**-3′	**pks11**
−149	5′-**TCGCGA**TGTTCATGTTAGTCCAGC	**TATATT**-3′	**Rv1730c**
−49	5′-**CTGCGA**CTACTAGGCCCGCGACCG	**TAAGCT**-3′	**purB**
−91	5′-**ATGCGA**CACGTAAACTTATCGGTG	**TACTTT**-3′	**mesT**
−49	5′-**CTGCGA**CGGCACCTATCGGGATCG	**TAATTT**-3′	**Rv3766**
Promoter sequences with 19 bp spacer			
−270	5′-**TTGCGA**TAAGTACCTCCGGCAATTC	**TATGAG**-3′	**gmhB**
−253	5′-**TTGCGA**TAAATGATGGTAGATCACG	**TCATTT**-3′	**Rv0246**
−77	5′-**TTGCGA**ACAATGACAGCCATGAGGC	**CAGTAT**-3′	**Rv0308**
−229	5′-**TTGCGA**CCGCGCTGGCGACCGGCAG	**TTCCGT**-3′	**desA2**
−70	5′-**TTGCGA**GATCGCGTAATGTGGCTTAA	**ACTTT**-3′	**pks11**
−107	5′-**TTGCGA**GCCGGGTGCATAGGCCCGG	**TACGGC**-3′	**Rv1846c**
−255	5′-**TTGCGA**TTGCCACCCTCAAAAAGGGC	**ACGTT**-3′	**Rv2433c**
−234	5′-**TTGCGA**TGCGTTCGCGATCGGTCCGG	**ACTCT**-3′	**pdhA**
−140	5′-**TTGCGA**TTGGACTTCGCAATCGTCG	**TTCGCT**-3′	**Rv2721c**
−282	5′-**TTGCGA**GGCCTGGTGGCGCTGGTGG	**TGCCAT**-3′	**PEPGRS47**
−127	5′-**TTGCGA**ACTTAGAGAAAGACTAGAAC	**ACGTT**-3′	**hmp**
−266	5′-**TTGCG**GCGCTGGGCGACAATCCCCG	**TAGAAT**-3′	**ileS**
−99	5′-**TTGCG**CCACGCCGTTGCGGTGTGAT	**TAATCT**-3′	**TB15.3**
−38	5′-**TTGCG**GCCCGGACTAGCGCGAAAGA	**TACTGT**-3′	**Rv1922**
−38	5′-**TTGCG**TCTGCTCGCGCACCCCGGAG	**TAGCGT**-3′	**Rv2449c**
−44	5′-**TTGCG**GTGCCGGTGTCGCCGGGCCG	**TAGGAT**-3′	**proU**
−157	5′-**TTGCG**GTGCCGGTGTCGCCGGGCCG	**TAGGAT**-3′	**tig**
−261	5′-**TTGCG**TTATACGACAGTCTGAGCAG	**TATATT**-3′	**PPE61**
−272	5′-**TTGCG**GAGGAGAACACCACATTCGT	**TATTCT**-3′	**Rv3737**
−234	5′-**TTGCG**CTTTCCCCAGGCCGGCGAAA	**TAACAT**-3′	**mmpL8**
−38	5′-**TTGCCA**GCGCTCGCGAACGCGCGGT	**TAACCT**-3′	**pepE**
−176	5′-**TTGCAA**ACGATACGCATGCTGTGTG	**TAGATT**-3′	**fadE34**
−72	5′-**TTTCGA**CAACCGTTGCACGGCACGG	**TAGCGT**-3′	**Rv0106**
−72	5′-**TTTCGA**CAACAGCTAGGTGGCACTG	**TACCGT**-3′	**Rv2059**
−218	5′-**TTCCGA**GTCACGATGTCGGCCTGGC	**TATCCT**-3′	**Rv2717c**
−110	5′-**TGGCGA**GTAGATACAACCATACGAG	**TAGGCT**-3′	**tuf**
−120	5′-**TCGCGA**CTGTGGAGGCTTGGCCGGT	**TAGCGT**-3′	**PE8**
−102	5′-**TCGCGA**CCCGACATTGCGAAAGTAT	**TACCGT**-3′	**Rv1815**
−86	5′-**TCGCGA**TGGACCGCTGTCGTCACTA	**TAACGT**-3′	**plcA**
−37	5′-**GTGCGA**TTCTGTGTCTGCTTGCCGG	**TAGACT**-3′	**Rv1096**
−123	5′-**ATGCGA**TGCTTTCGTCTTGAGCCAC	**TACCAT**-3′	**pks3**
−39	5′-**CTGCGA**CCCATTAAGTAGGTTAGGC	**TACATT**-3′	**MBT1**
−38	5′-**CTGCGA**CAACTCCAAGCGTGGCCGG	**TAACGT**-3′	**fadD28**
−276	5′-**ATGCGA**CGGCCATCACAGACCAAGG	**TATCGT**-3′	**Rv3098c**
−269	5′-**GTGCGA**CATTGACGCCATCGGCGAC	**TAGCTT**-3′	**Rv3277**

Promoter sequences in 5′-UTRs of *M.tuberculosis* genome, up to 300 bp from the translational start site were subjected to pattern search analysis by using the TubercuList website for the presence of a sequence 5′-TTGCGA-$N_{18\pm1}$-TANNNT-3′ with one mismatch in either of the putative −35 and −10 sequences. Putative −35 and −10 sequences are underlined. The conserved bases are given in boldface. Numbers on the right refer to the position of the first base of the putative −35 sequence with respect to the start codon of the putative ORF.

11 623 nmol/min/mg in *M.smegmatis*, which was comparable with the activity of A_{37TG-} (9909 nmol/min/mg). However, contrary to our expectations, A_{37TG-}(*conE.coli*) exhibited only negligible activity in *E.coli* (85 nmol/min/mg). These observations suggested that in *E.coli*, in addition to −35 and −10 sequences, there might be other determinants required for promoter function, which may not be essential

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

Sequence	Clone	β-galactosidase activity (nmol/min/mg)
GATTGCGAAGGTTCAATCCGACGTACACTGTACTCG	$A_{37\,TG\text{-}}con_{(N15)}$	1912±15
GATTGCGAAGGTTCAATCCCGACGTACACTGTACTCG	$A_{37\,TG\text{-}}con_{(N16)}$	11278±76
GATTGCGAAGGTTCAATCCCCGACGTACACTGTACTCG	$A_{37\,TG\text{-}}con_{(N17)}$	25393±214
GATTGCGAAGGTTCAATCCNCCGACGTACACTGTACTCG	$A_{37\,TG\text{-}}con_{(N18)}$	36309±290
GATTGCGAAGGTTCAATCCNNCCGACGTACACTGTACTCG	$A_{37\,TG\text{-}}con_{(N19)}$	5757±34

Figure 8. Effects of spacer length on the activity of $A_{37TG\text{-}}$ promoter. Spacer length of $A_{37TG\text{-}}$ promoter was altered by deleting or inserting a base in the spacer region between −10 and −35 sequences. Random bases were used for insertion (depicted by 'N'). Promoter activity was measured as β-galactosidase specific activity. The cell lysate of pSD5B-transformed *M.smegmatis* (exhibiting an activity of 12 ± 5 nmol/min/mg; data not shown) was used as a negative control. The values represent mean of three separate assays. Boldface letters indicate TSP.

Sequence	Clone	β-galactosidase activity (nmol/min/mg)	
		M.smegmatis	*E. coli*
-30 -20 -10 +1 GATTTGACAGGTTCAATCCCCGACGTACACTGTACTCG	$A_{37TG\text{-}}$(*conE. coli*)	11623±191	82±5
GATTTGACAGGTTCAATTTTATGTTTACACTGTACTCG	$A_{37TG\text{-}}$(*conE. coli*)*AT*+	12860±121	1277±15

Figure 9. Analysis of the functional differences between mycobacterial and *E.coli* promoters. Promoter activities were measured as β-galactosidase specific activity. The cell lysate of pSD5B-transformed *M.smegmatis* (exhibiting an activity of 12 ± 5 nmol/min/mg; data not shown) was used as a negative control. The values represent mean of three separate assays. Underlined sequences represent the conserved bases at −35 and −10 positions. The numbers above promoter sequence indicate the positions of respective bases. Boldface letter represents +1 position. In case of both the organisms, the samples for measuring promoter activities were obtained from cultures in their mid-log phase.

in mycobacteria. Recently, in a study by Liu *et al*. (34), it was reported that the −15 sequence plays a crucial role in determining the strength of *E.coli* promoters. *E.coli* promoters with a consensus −15 sequence 5′-TTTATGTT-3′, exhibited ∼2- to 8-fold higher activities compared to their counterparts devoid of this sequence. In order to find out whether the inability of $A_{37TG\text{-}}$ *(conE.coli)* promoter derivative to support transcription in *E.coli* was related to the presence of GC-rich sequence in the spacer region, we replaced the sequence $^{-20}$CCCCGACG^{-13} by an AT-rich sequence (Figure 9). The resulting promoter derivative, $A_{37TG\text{-}}$*(conE.coli)AT*+ showed ∼15-fold increased activity in *E.coli*, while in *M.smegmatis*, its activity remained unchanged (Figure 9). These results pointed out that the sequences in the spacer region and their AT richness could be one of the determinants in differentiating between mycobacterial and *E.coli* promoters.

DISCUSSION

Selection of strong transcriptional signals of mycobacteria and analysis of A_{37}

Although, the −10 region of mycobacterial promoters has been clearly identified, the additional features required for transcription from these promoters have not been clearly defined. In this study, we have carried out a detailed analysis to identify what other sequences/features apart from −10 region contribute to the activity of mycobacterial promoters. Since majority of the known housekeeping promoters of mycobacteria are weak (24) and are unlikely to carry consensus/nearly consensus recognition sequences, we were required to generate strong promoters which bind efficiently with the RNAP of mycobacteria which was obtained by following a strategy, similar to the saturation mutagenesis (30). However, owing to lack of sufficient knowledge about the mycobacterial promoter elements (except the Pribnow Box), we started with background information about the promoters from other prokaryotic systems. We generated DNA sequence library harboring ∼100 bp long DNA fragments containing random sequences in a stretch of 29 bases, which represent the number of bases acquired between −35 and −10 positions in a typical prokaryotic promoter (number of bases in two hexamers separated by a distance of 17 bp = [2 × 6] + 17) (31,32). Despite using the incomplete library of DNA sequences, we were able to select a few strong promoter sequences (Figure 3). We chose to characterize the promoter A_{37} from this library based on its extremely high activity and near-perfect score.

Thorough analysis of A_{37} revealed that its extremely high activity could be subscribed to cumulative effect of several features such as a purine at +1, a conserved −10 sequence along with an extended −10 motif. We have observed that replacing the base at +1 by any of the purine residues resulted in ∼2-fold increase in the promoter's activity in mycobacteria (N. Agarwal and A. K. Tyagi, unpublished data). In addition, we had earlier described the role of −10 and extended −10 sequences in mycobacterial promoters (11,26,35). In the DNase I footprinting experiments, hyperactivities of DNase I at −24/−25 positions of A_{37} indicated overexposure of

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

Table 2. Analysis of strong promoter sequences of mycobacteria

Gene	−35	−10	+1	Sp.	Ref.
BCG_*hsp60* tsA	TTGCAC	AAGAAT	G	18/6	(13)
Mfo_*blaF*	TTCAAA	TACGCT	A	19/7	(20)
L5_Pleft	TTGACA	CATTCT	A	18/6	(49)
L5_P2	TTGCTA	TACATT	G	18/7	(49)
Mtb_*rpsl*	TTGACC	TATTGT	G	18/8	(N. Agarwal and A. K. Tyagi, unpublished data)
Msm_*rpsl*	TTGTTT	TATTGT	G	18/8	(50)
Msm_*rrnB* P1	TTGACT	TAACTT	G	17/7	(51)
Msm_*rrnA* P2	TTGCAA	TATCTT	G	18/6	(51)
Msm_*rrnA* PCL1	TTGACG	TAGACT	G	18/6	(51)
Mtb_*rrnA* PCL1	TTGACT	TAGACT	G	18/6	(51)
Mfo_*rrnA* P2	TTGACC	TAATCT	A	18/5	(52)
Mfo_*rrnA* P3	TTGACA	TAAGCT	G	18/6	(52)

Promoter sequences of mycobacteria with strong activities were analyzed to find out the presence of conserved sequences at −35 and −10 positions and the length of the spacer region. +1 represents the TSP. Sp. represents the spacer length between −35 and −10 sequences and between −10 region and TSP, respectively. These sequences were obtained from the upstream region of various genes from different mycobacterial species: BCG, *M.bovis* BCG; Mfo, *M.fortuitum*; L5, Mycobacteriophage L5; Mtb, *M.tuberculosis* and Msm, *M.smegmatis*. The promoter sequences were obtained from different sources as mentioned in corresponding references.

the bases to DNAse I due to the presence of RNAP. This suggests that interaction of RNAP with A_{37} may result in the generation of a favorable conformation of the promoter possibly due to bending at −24/−25 positions for a better binding of holoenzyme to both the −35 and the −10 sequences.

Role of the −35 sequence for promoter functions

The reduced RNAP binding and consequently reduced promoter activity due to substitution of bases at −35 region suggested the requirement of an optimal −35 sequence in the activity of mycobacterial promoters. In this study we show that for the optimal activity and recognition of RNAP, a sequence at −35 region, 5′-TTGCGA-3′ is preferred by mycobacterial transcriptional machinery. However, as shown in Figure 3, most of the promoters in our library lack this consensus sequence at −35 positions, which is probably due to incomplete DNA sequence library used for the selection of promoter-like sequences by mycobacterial RNAP. Significant changes in the activities of the promoters, A_{37TG-}, *sigA, mmsA* and *gcvH* on the substitution of their respective −35 regions substantiated the importance of −35 region in the activity of a mycobacterial promoter. Further evidence for the role of −35 sequence in promoter function was provided by enhanced binding of the mycobacterial RNAP with A_{37TG-}*con* and *sigAprocon* promoter derivatives containing 5′-TTGCGA-3′ sequence at −35 region. However, as shown in Figure 6, substitutions of various individual bases at −35 site still resulted in substantial promoter activities, indicating that mycobacterial transcriptional machinery can tolerate variety of sequences at −35 position, as reported in our previous studies (11).

Determination of optimum distance between −35 and −10 sequences of mycobacterial promoters

The distance between −35 and −10 sequences is one of the important determinants of promoter activity (36–47). Alterations in the distance between −35 and −10 sequences revealed that unlike *E.coli* RNAP (where the optimum distance between −35 and −10 sequences is 17 bp), mycobacterial enzyme requires an 18 bp long spacer sequence for optimal promoter activity. These results along with the observations that >40% of the putative promoter sequences in 5′-UTRs, obtained by pattern search analysis, exhibited a distance of 18 bp between putative −35 and −10 sequences (Table 1), suggest that a distance of 18 bp between −35 and −10 sequences represents an optimal spacer length for mycobacterial promoters. Further support for this comes from our analysis of several known mycobacterial promoters, which revealed the presence of a spacer of 18 bp in most of the strong promoters (Table 2).

Analysis of the functional differences between mycobacterial and *E.coli* promoters

Despite the similarities with *E.coli* promoters, the mycobacterial promoters do not function efficiently in *E.coli* (24). In the present study also, we observed that A_{37TG-}*conE.coli*, a derivative of A_{37TG-} promoter, despite having an *E.coli* consensus sequence, 5′-TTGACA-3′ at −35 position did not exhibit significant activity in *E.coli*. Recently, in a study by Liu *et al.* (34), it was shown that the presence of GC-rich sequences in the spacer region drastically influences the strength of promoters in *E.coli*. This observation was further substantiated by the fact that majority of strong *E.coli* promoters have an AT content of >75–80%. Analysis of A_{37TG-}*con E.coli* promoter derivative indicated the presence of high GC content (∼60%) in the spacer region. Hence, we replaced the GC-rich spacer sequence of this promoter (from position −13 to −20) by a sequence resulting in 75% AT richness in the spacer region. This enhanced AT richness resulted in 15-fold higher activity of this promoter in *E.coli*. However, despite this significant increase, the activity of A_{37}*TG-(conE.coli)AT+* was still lower in *E.coli* than in mycobacterium. This could probably be due to a suboptimal distance between −35 and −15 sequences in this promoter (8 bp) compared to the distance reported by Liu *et al.* (34), where it was observed that the distance between −35 and

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

−15 sequences is an important determinant of the promoter activity in *E.coli*. It has been observed by Callaci *et al.* (48) that the inter-domain distance between regions 2.4 and 4.2 of *E.coli* σ^{70} is much shorter than the distance between −10 and −35 promoter elements. Hence, the AT-rich spacer sequence may be better suitable for appropriate binding of this region required by RNAP to establish optimal contacts with −10 and −35 hexameric sequences. Although, σ^{A} from mycobacteria has not been crystallized as yet, possibly, it may have a more appropriate distance between 2.4 and 4.2 regions, thus, making it less dependent on the maneuvering of promoter region affected by bending of the spacer sequence. This may provide an explanation as to why mycobacterial promoters may function with highly GC-rich spacer regions but exhibit significantly reduced activity in *E.coli*.

ACKNOWLEDGEMENTS

We thank Bhavna Varshney for assistance in the mutation studies, Radhika Gupta and Ruchi Jain for critical reading of the manuscript and Rajiv Chawla for efficient secretarial assistance in the preparation of the manuscript. N.A. is grateful to the Council of Scientific and Industrial Research, India for a fellowship. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Corbett,E.L., Watt,C.J., Walker,N., Maher,D., Williams,B.G., Raviglione,M.C. and Dye,C. (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch. Intern. Med.*, **163**, 1009–1021.
2. Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry,C.E.,III *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537–544.
3. Sun,R., Converse,P.J., Ko,C., Tyagi,S., Morrison,N.E. and Bishai,W.R. (2004) *Mycobacterium tuberculosis* ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol. Microbiol.*, **52**, 25–38.
4. Manganelli,R., Provvedi,R., Rodrigue,S., Beaucher,J., Gaudreau,L. and Smith,I. (2004) Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. *J. Bacteriol.*, **186**, 895–902.
5. Manganelli,R., Fattorini,L., Tan,D., Iona,E., Orefici,G., Altavilla,G., Cusatelli,P. and Smith,I. (2004) The extra cytoplasmic function sigma factor sigma(E) is essential for *Mycobacterium tuberculosis* virulence in mice. *Infect. Immun.*, **72**, 3038–3041.
6. Kaushal,D., Schroeder,B.G., Tyagi,S., Yoshimatsu,T., Scott,C., Ko,C., Carpenter,L., Mehrotra,J., Manabe,Y.C., Fleischmann,R.D. *et al.* (2002) Reduced immunopathology and mortality despite tissue persistence in a *Mycobacterium tuberculosis* mutant lacking alternative sigma factor, SigH. *Proc. Natl Acad. Sci. USA*, **99**, 8330–8335.
7. Geiman,D.E., Kaushal,D., Ko,C., Tyagi,S., Manabe,Y.C., Schroeder,B.G., Fleischmann,R.D., Morrison,N.E., Converse,P.J., Chen,P. *et al.* (2004) Attenuation of late-stage disease in mice infected by the *Mycobacterium tuberculosis* mutant lacking the SigF alternate sigma factor and identification of SigF-dependent genes by microarray analysis. *Infect. Immun.*, **72**, 1733–1745.
8. Calamita,H., Ko,C., Tyagi,S., Yoshimatsu,T., Morrison,N.E. and Bishai,W.R. (2005) The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. *Cell Microbiol.*, **7**, 233–244.
9. Ando,M., Yoshimatsu,T., Ko,C., Converse,P.J. and Bishai,W.R. (2003) Deletion of *Mycobacterium tuberculosis* sigma factor E results in delayed time to death with bacterial persistence in the lungs of aerosol-infected mice. *Infect. Immun.*, **71**, 7170–7172.
10. Hahn,M.Y., Raman,S., Anaya,M. and Husson,R.N. (2005) The *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. *J. Bacteriol.*, **187**, 7062–7071.
11. Bashyam,M.D., Kaushal,D., Dasgupta,S.K. and Tyagi,A.K. (1996) A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J. Bacteriol.*, **178**, 4847–4853.
12. Gomez,M. and Smith,I. (2000) Determinants of mycobacterial gene expression. In Hatful,G.F. and Jacobs,W.R.,Jr (eds), *Molecular Genetics of Mycobacteria*. ASM Press, Washington DC, pp. 111–129.
13. Stover,C.K., de la Cruz,V.F., Fuerst,T.R., Burlein,J.E., Benson,L.A., Bennett,L.T., Bansal,G.P., Young,J.F., Lee,M.H., Hatfull,G.F. *et al.* (1991) New use of BCG for recombinant vaccines. *Nature*, **351**, 456–460.
14. Suzuki,Y., Nagata,A. and Yamada,T. (1991) Analysis of the promoter region in the rRNA operon from *Mycobacterium bovis* BCG. *Antonie Van Leeuwenhoek*, **60**, 7–11.
15. Mehra,V., Sweetser,D. and Young,R.A. (1986) Efficient mapping of protein antigenic determinants. *Proc. Natl Acad. Sci. USA*, **83**, 7013–7017.
16. Shinnick,T.M. (1987) The 65-kilodalton antigen of *Mycobacterium tuberculosis*. *J. Bacteriol.*, **169**, 1080–1088.
17. Thole,J.E., Keulen,W.J., De Bruyn,J., Kolk,A.H., Groothuis,D.G., Berwald,L.G., Tiesjema,R.H. and van Embden,J.D. (1987) Characterization, sequence determination, and immunogenicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli* K-12. *Infect. Immun.*, **55**, 1466–1475.
18. Collins,M.E., Moss,M.T., Wall,S. and Dale,J.W. (1987) Expression of biotinylated proteins in mycobacterial gene libraries. *FEMS Microbiol. Lett.*, **43**, 53–56.
19. Murray,A., Winter,N., Lagranderie,M., Hill,D.F., Rauzier,J., Timm,J., Leclerc,C., Moriarty,K.M., Gheorghiu,M. and Gicquel,B. (1992) Expression of *Escherichia coli* beta-galactosidase in *Mycobacterium bovis* BCG using an expression system isolated from *Mycobacterium paratuberculosis* which induced humoral and cellular immune responses. *Mol. Microbiol.*, **6**, 3331–3342.
20. Timm,J., Perilli,M.G., Duez,C., Trias,J., Orefici,G., Fattorini,L., Amicosante,G., Oratore,A., Joris,B., Frere,J.M. *et al.* (1994) Transcription and expression analysis, using lacZ and phoA gene fusions of *Mycobacterium fortuitum* beta-lactamase genes cloned from a natural isolate and a high level beta-lactamase producer. *Mol. Microbiol.*, **12**, 491–504.
21. Satchidanandam,V., Amara,R.R., Uchil,P.D. and Singh,V. (2003) The regulatory elements of the *Mycobacterium tuberculosis* gene Rv3881c function efficiently in *Escherichia coli*. *FEMS Microbiol. Lett.*, **218**, 365–370.
22. Mulder,M.A., Zappe,H. and Steyn,L.M. (1999) The *Mycobacterium tuberculosis* katG promoter region contains a novel upstream activator. *Microbiology*, **145**, 2507–2518.
23. Dellagostin,O.A., Esposito,G., Eales,L.J., Dale,J.W. and McFadden,J. (1995) Activity of mycobacterial promoters during intracellular and extracellular growth. *Microbiology*, **141**, 1785–1792.
24. Das Gupta,S.K., Bashyam,M.D. and Tyagi,A.K. (1993) Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. *J. Bacteriol.*, **175**, 5186–5192.
25. Jain,S., Kaushal,D., DasGupta,S.K. and Tyagi,A.K. (1997) Construction of shuttle vectors for genetic manipulation and molecular analysis of mycobacteria. *Gene*, **190**, 37–44.
26. Agarwal,N. and Tyagi,A.K. (2003) Role of 5′-TGN-3′ motif in the interaction of mycobacterial RNA polymerase with a promoter of 'extended −10' class. *FEMS Microbiol. Lett.*, **225**, 75–83.
27. Pollock,R. and Treisman,R. (1990) A sensitive method for the determination of protein–DNA binding specificities. *Nucleic Acids Res.*, **18**, 6197–6204.
28. Good,L. and Nazar,R.N. (1992) An improved thermal cycle for two-step PCR-based targeted mutagenesis. *Nucleic Acids Res.*, **20**, 4934.
29. Miller,J.F. (1972) *Experiments In Molecular Genetics*. Cold Spring Harbor, NY.
30. Baliga,N.S. (2001) Promoter analysis by saturation mutagenesis. *Biol. Proced. Online*, **3**, 64–69.
31. Harley,C.B. and Reynolds,R.P. (1987) Analysis of *E.coli* promoter sequences. *Nucleic Acids Res.*, **15**, 2343–2361.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012

32. Helmann,J.D. (1995) Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.*, **23**, 2351–2360.
33. Dombroski,A.J., Johnson,B.D., Lonetto,M. and Gross,C.A. (1996) The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proc. Natl Acad. Sci. USA*, **93**, 8858–8862.
34. Liu,M., Tolstorukov,M., Zhurkin,V., Garges,S. and Adhya,S. (2004) A mutant spacer sequence between −35 and −10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. *Proc. Natl Acad. Sci. USA*, **101**, 6911–6916.
35. Bashyam,M.D. and Tyagi,A.K. (1998) Identification and analysis of 'extended −10' promoters from mycobacteria. *J. Bacteriol.*, **180**, 2568–2573.
36. Rossi,J.J., Soberon,X., Marumoto,Y., McMahon,J. and Itakura,K. (1983) Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives. *Proc. Natl Acad. Sci. USA*, **80**, 3203–3207.
37. Aoyama,T., Takanami,M., Ohtsuka,E., Taniyama,Y., Marumoto,R., Sato,H. and Ikehara,M. (1983) Essential structure of *E.coli* promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res.*, **11**, 5855–5864.
38. Mulligan,M.E., Brosius,J. and McClure,W.R. (1985) Characterization *in vitro* of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the TAC promoter. *J. Biol. Chem.*, **260**, 3529–3538.
39. Jacquet,M.A. and Ehrlich,R. (1985) *In vivo* and *in vitro* effect of mutations in *tetA* promoter from pSC101: insertion of poly(dA.dT) stretch in the spacer region does not inactivate the promoter. *Biochimie*, **67**, 987–997.
40. Auble,D.T., Allen,T.L. and deHaseth,P.L. (1986) Promoter recognition by *Escherichia coli* RNA polymerase. Effects of substitutions in the spacer DNA separating the −10 and −35 regions. *J. Biol. Chem.*, **261**, 11202–11206.
41. Borowiec,J.A. and Gralla,J.D. (1987) All three elements of the lac ps promoter mediate its transcriptional response to DNA supercoiling. *J. Mol. Biol.*, **195**, 89–97.
42. Aoyama,T. and Takanami,M. (1988) Supercoiling response of *E.coli* promoters with different spacer lengths. *Biochim. Biophys. Acta*, **949**, 311–317.
43. Auble,D.T. and deHaseth,P.L. (1988) Promoter recognition by *Escherichia coli* RNA polymerase. Influence of DNA structure in the spacer separating the −10 and −35 regions. *J. Mol. Biol.*, **202**, 471–482.
44. Jacquet,M.A., Ehrlich,R. and Reiss,C. (1989) *In vivo* gene expression directed by synthetic promoter constructions restricted to the −10 and −35 consensus hexamers of *E.coli*. *Nucleic Acids Res.*, **17**, 2933–2945.
45. Lozinski,T., Markiewicz,W.T., Wyrzykiewicz,T.K. and Wierzchowski,K.L. (1989) Effect of the sequence-dependent structure of the 17 bp AT spacer on the strength of consensuslike *E.coli* promoters *in vivo*. *Nucleic Acids Res.*, **17**, 3855–3863.
46. Ayers,D.G., Auble,D.T. and deHaseth,P.L. (1989) Promoter recognition by *Escherichia coli* RNA polymerase. Role of the spacer DNA in functional complex formation. *J. Mol. Biol.*, **207**, 749–756.
47. Typas,A. and Hengge,R. (2006) Role of the spacer between the −35 and −10 regions in sigmas promoter selectivity in *Escherichia coli*. *Mol. Microbiol.*, **59**, 1037–1051.
48. Callaci,S., Heyduk,E. and Heyduk,T. (1999) Core RNA polymerase from *E.coli* induces a major change in the domain arrangement of the sigma 70 subunit. *Mol. Cell*, **3**, 229–238.
49. Nesbit,C.E., Levin,M.E., Donnelly-Wu,M.K. and Hatfull,G.F. (1995) Transcriptional regulation of repressor synthesis in Mycobacteriophage L5. *Mol. Microbiol.*, **17**, 1045–1056.
50. Kenney,T.J. and Churchward,G. (1996) Genetic analysis of the *Mycobacterium smegmatis rps*L promoter. *J. Bacteriol.*, **178**, 3564–3571.
51. Gonzalez-y-Merchand,J.A., Colston,M.J. and Cox,R.A. (1996) The rRNA operons of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: comparison of promoter elements and of neighbouring upstream genes. *Microbiology*, **142**, 667–674.
52. Gonzalez-y-Merchand,J.A., Garcia,M.J., Gonzalez-Rico,S., Colston,M.J. and Cox,R.A. (1997) Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA. *J. Bacteriol.*, **179**, 6949–6958.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 10, 2012