JOURNAL OF BACTERIOLOGY, Aug. 1996, p. 4847–4853
0021-9193/96/$04.00+0

Vol. 178, No. 16

A Study of the Mycobacterial Transcriptional Apparatus: Identification of Novel Features in Promoter Elements

MURALI D. BASHYAM, DEEPAK KAUSHAL, SUJOY K. DASGUPTA,† AND ANIL K. TYAGI*

Department of Biochemistry, University of Delhi South Campus, New Delhi-110021, India

Received 27 February 1996/Accepted 8 June 1996

Our earlier studies on transcriptional signals of mycobacteria had revealed that (i) strong promoters occur less frequently in the slowly growing pathogen *Mycobacterium tuberculosis* H37Rv than in the fast-growing saprophyte *M. smegmatis* and (ii) mycobacterial promoters function poorly in *Escherichia coli*. We now present evidence that RNA polymerases of *M. smegmatis*, *M. tuberculosis*, and *M. bovis* BCG recognize promoter elements with comparable efficiencies. Analysis of these randomly isolated mycobacterial promoters by DNA sequencing, primer extension, and deletion experiments revealed that their −10 regions are highly similar to those of *E. coli* promoters, in contrast to their −35 regions, which can tolerate a greater variety of sequences, owing presumably to the presence of multiple sigma factors with different or overlapping specificities for −35 regions, as reported earlier for the *Streptomyces* promoters. A comparison of the −10 and −35 binding domains of MysA, HrdB, and RpoD (the principal sigma factors of *M. smegmatis*, *Streptomyces aureofaciens*, and *E. coli*, respectively) showed that all three sigma factors have nearly identical −10 binding domains. However, the −35 binding domains of the principal mycobacterial and streptomycete sigma factors, although nearly identical to each other, are vastly different from the corresponding region of the sigma factor of *E. coli*. Thus, the transcriptional signals of mycobacteria have features in common with *Streptomyces* promoters but differ from those of *E. coli* because of major differences in the −35 regions of the promoters and the corresponding binding domain in the sigma factor.

The genus *Mycobacterium* includes pathogens that cause the highest number of bacterial infections and human deaths every year (2). Because of the emergence of multidrug-resistant strains of *Mycobacterium tuberculosis* (7) and the increasing incidence of mycobacterial infections in human immunodeficiency virus-positive cases (4), tuberculosis has emerged as a major global threat to health. Despite their highly pathogenic nature, progress towards an understanding of gene structure and gene expression in mycobacteria has been slow. The lack of information on mycobacterial transcriptional signals has particularly impeded our understanding of the regulation of gene expression in these organisms.

Only a few mycobacterial promoters have been studied. These include promoters for the 16S rRNA genes of *M. smegmatis* (13), *M. tuberculosis* (12, 30), and *M. leprae* (26); the *bla* gene of *M. fortuitum* (29); the *ask*β gene of *M. smegmatis* (5); the *hsp-60* gene of *M. bovis* BCG (18); the *cpn-60* gene of *M. tuberculosis* (15); the 85A antigen gene of *M. tuberculosis* (17); the *P*AN promoter from *M. paratuberculosis* transposon IS*900* (21); and the three promoters responsible for transcribing the repressor-like gp71 protein of mycobacteriophage L5 (22). It is, however, difficult to assess the nature of transcriptional signals in mycobacteria by comparison of these promoter sequences, as many of these belong to specifically regulated genes and thus may not contribute towards the elucidation of constitutive gene expression in mycobacteria. The purpose of an unbiased study of mycobacterial transcriptional signals will be best achieved by analysis of randomly isolated promoters by using a promoter probe vector. We had earlier reported the isolation of a battery of promoters from the fast-growing saprophyte *M. smegmatis* and the slowly growing pathogen *M. tuberculosis* H37Rv by using chloramphenicol acetyltransferase (CAT)-based vector pSD7 (8). It was observed that most mycobacterial promoters function poorly in *Escherichia coli* (8). We report here our findings on the nature of the mycobacterial transcriptional apparatus and the specificity of sequence requirements for initiation of transcription in mycobacteria.

* Corresponding author. Mailing address: Department of Biochemistry, University of Delhi South Campus, Benito Juarez Road, New Delhi-110021, India. Fax: 91-11-6885270 or 91-11-6886427.

† Present address: Department of Microbiology, Bose Institute, Calcutta 700054, India.

MATERIALS AND METHODS

Bacterial strains and plasmids. *M. smegmatis* LR222 (provided by J. T. Crawford, Centers for Disease Control and Prevention, Atlanta, Ga.), *M. tuberculosis* H37Ra, and *M. bovis* BCG (provided by J. S. Tyagi, Department of Biotechnology, All India Institute of Medical Sciences, New Delhi) were used as mycobacterial host strains. All transformations in *E. coli* were performed with strain DH5α. The promoter-containing derivatives of pSD7 were obtained as previously described (8). Plasmid pGEM-5Zf(+) was purchased from Promega Corporation, Madison, Wis.

Bacterial culture methods, transformation, and CAT assay. *M. smegmatis* and *E. coli* were grown and transformed as described earlier (8). *M. tuberculosis* and *M. bovis* BCG were grown in Middlebrook 7H9 broth (Difco) supplemented with albumin-glucose complex (Difco) or on Middlebrook 7H10 agar (Difco) supplemented with oleic acid-albumin-glucose complex (Difco) in the presence of 25 μg of cycloheximide per ml. Cells for electroporation were prepared as described previously (11). *M. tuberculosis* and *M. bovis* BCG were transformed by using a Cell Porator (Bethesda Research Laboratories, Gaithersburg, Md.) at a field strength of 15 kV/cm. Transformants carrying the promoter clones were selected on Luria-Bertani (Difco) agar for *E. coli* and *M. smegmatis* and 7H10 agar for *M. bovis* BCG and *M. tuberculosis*, containing kanamycin (25 μg/ml) and kanamycin with chloramphenicol (20 μg/ml). Colonies appearing on the plates containing kanamycin and chloramphenicol were used to determine CAT activity. CAT assays were performed as described earlier (8, 9).

Enzymes and chemicals. Restriction enzymes, *E. coli* Klenow fragment, T4 DNA ligase, T4 DNA polymerase, and T4 polynucleotide kinase were purchased from New England Biolabs Inc., Beverly, Mass. [^{14}C]chloramphenicol (specific activity, 50 mCi/mmol) was purchased from Amersham International plc., Amersham, United Kingdom. The Sequenase Version 2.0 DNA sequencing kit, Moloney murine leukemia virus reverse transcriptase, and pancreatic RNase inhibitor were purchased from United States Biochemical Corporation, Cleveland, Ohio. Calf intestinal alkaline phosphatase was purchased from Promega Corporation. [γ-^{32}P]ATP and [α-^{35}S]dATP were obtained from the Board of Radiation and Isotope Technology, Bombay, India.

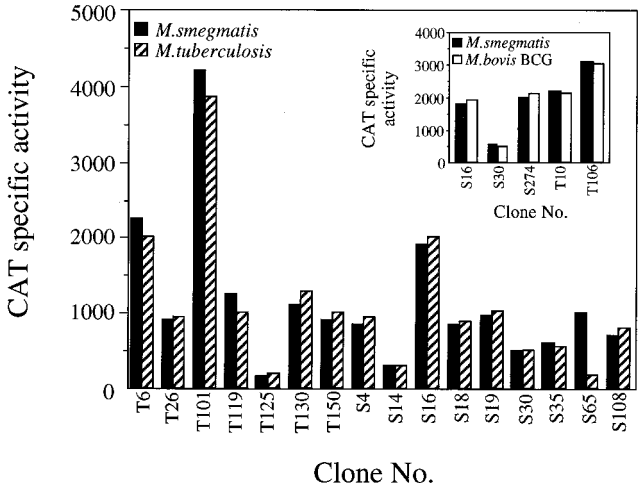


FIG. 1. Comparative assessment of mycobacterial promoters in *M. smegmatis*, *M. tuberculosis* H37Ra, and *M. bovis* BCG. The CAT activity supported by *M. tuberculosis* H37Rv promoters (T6 to T150) and *M. smegmatis* promoters (S4 to S108) was determined in *M. smegmatis* LR222 (solid bars) and *M. tuberculosis* H37Ra (hatched bars). CAT specific activity is expressed as nanomoles per minute per milligram of protein. The inset shows the activities of three *M. smegmatis* promoters (S16, S30, and S274) and two *M. tuberculosis* promoters (T10 and T106) in *M. smegmatis* and *M. bovis* BCG.

DNA sequencing and determination of the transcriptional start points (TSP) of mycobacterial promoters. Two primers were synthesized for sequencing of both strands of the promoter fragments. The 20-mer forward primer (complementary to the mycobacterial origin of replication [ORIM]), starting 92 bp upstream of the unique *Bam*HI site in pSD7 (5′-GAGTGCTTCCAGCACGG GCG-3′) and the 27-mer reverse primer (complementary to the *cat* gene), starting 65 bp downstream of the unique *Bam*HI site (5′-CCTGAAAATCTCG TCGAAGCTCGGCGG-3′), were designated Pr.P1 and R.T.Pr., respectively (see Fig. 2). Sequencing reactions were carried out by using Sequenase Version 2.0 in accordance with the manufacturer's instructions. Isolation of RNAs from the *M. smegmatis* transformants and determination of the TSP of the mycobacterial promoters were carried out as described earlier (1).

Deletion mapping of promoters S35 and T3. Various modifications were made in plasmid pSD7 containing *M. smegmatis* promoter S35 (designated pS35) by utilizing the unique *Xba*I restriction site present at position −20 relative to the transcriptional start point of the promoter. We separately deleted the DNA sequences downstream and upstream of the *Xba*I site. The resulting constructs, pS35.1 and pS35.2, contained only the −35 region and only the −10 region of promoter S35, respectively (see Fig. 4). Subsequently, three DNA fragments of different sizes, derived from the multiple cloning region of pGEM-5Z f(+), a 43-mer *Sph*I-*Pst*I fragment, a 48-mer *Nco*I-*Nde*I fragment, and a 94-mer *Apa*I-*Nsi*I fragment, were separately cloned into the *Xba*I site of pS35 (see Fig. 4). For each of the three clonings, recombinants containing the particular fragment in both orientations were selected, resulting in a total of six modified constructs (pS35.3a to pS35.5b; see Fig. 4). Similar modifications were carried out in plasmid pSD7 containing *M. tuberculosis* promoter T3 (designated pT3) by utilizing the unique *Kpn*I site present at position −22 relative to the transcriptional start point (see Fig. 4). All of the modified constructs of S35 and T3 were confirmed by DNA sequencing.

Cloning of *E. coli* promoters in pSD7. Genomic DNA from *E. coli* DH5α was digested with *Sau*3AI, and fragments ranging from 0.1 to 1 kb were ligated to pSD7 cleaved with *Bam*HI. The ligation mixture was used to transform *E. coli*, and promoter clones were selected on LB agar containing kanamycin (25 μg/ml) and chloramphenicol (20 μg/ml).

RESULTS

Comparative assessment of mycobacterial promoters in *M. smegmatis* and *M. tuberculosis* H37Ra. The *M. tuberculosis* and *M. smegmatis* promoters that we analyzed in this study had been isolated earlier in *M. smegmatis* by using pSD7 (8). Our

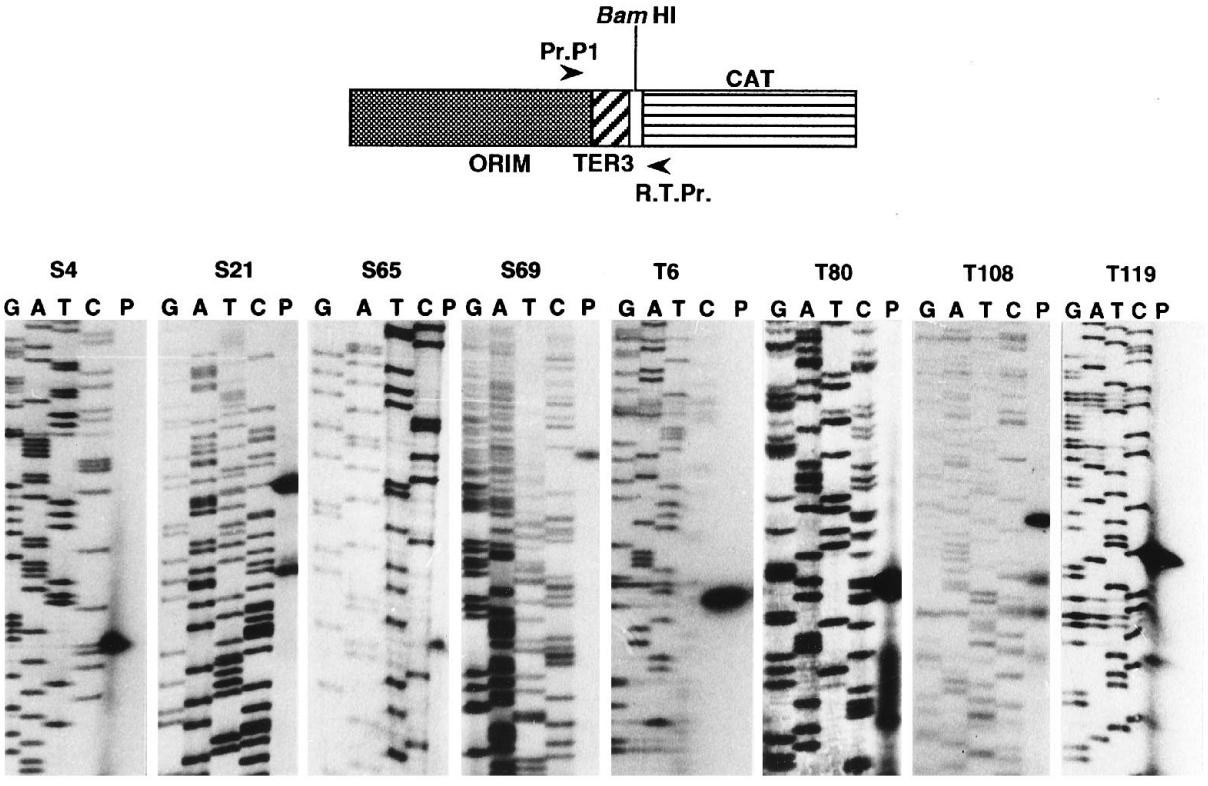


FIG. 2. Determination of TSP of mycobacterial promoters. Primer extension reactions were carried out with end-labelled primer R.T.Pr. and RNAs isolated from the mycobacterial promoter clones. The reactions were analyzed on urea–6% polyacrylamide gels. Sequencing reactions were also performed with R.T.Pr. and run alongside the corresponding primer extension reactions. The autoradiogram shows the TSP of four *M. smegmatis* promoter clones, S4, S21, S65, and S69, and four *M. tuberculosis* promoter clones, T6, T80, T108, and T119 (lane P in each of the eight reactions). The positions of the two primers (Pr.P1 and R.T.Pr.) are shown.

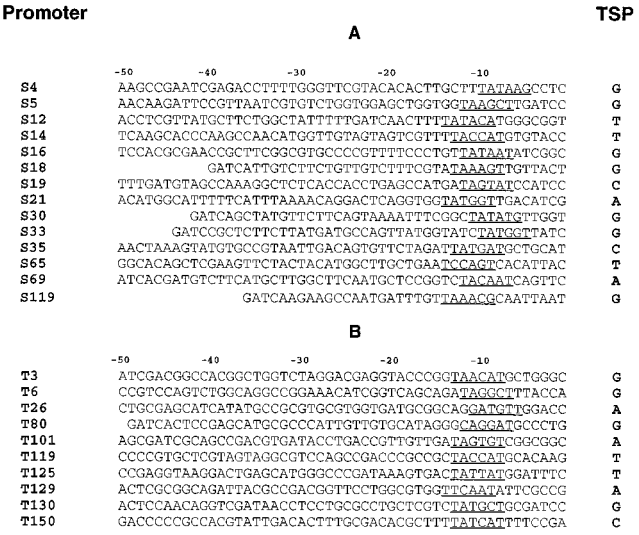


FIG. 3. Alignment of mycobacterial promoters. Fourteen *M. smegmatis* promoters (A) and 10 *M. tuberculosis* promoters (B) were aligned on the basis of their TSP. The TSP of each promoter is shown at the extreme right. The nucleotide positions of the promoters with respect to the TSP are shown at the top. The −10 region hexameric sequence for each promoter is underlined.

experiments had demonstrated that strong promoters occur less frequently in *M. tuberculosis* than in *M. smegmatis*. Although these observations are consistent with a lower rate of transcription, a lower RNA content per unit of DNA, and a slower growth rate of *M. tuberculosis* than of *M. smegmatis* (10), it was conceivable that the promoters of *M. tuberculosis* had diverged from those of *M. smegmatis* to such an extent that they are not optimally recognized by the RNA polymerase of the latter. Hence, we initially compared the recognition of promoters in *M. smegmatis* and *M. tuberculosis*. Sixteen randomly selected promoter plasmids (9 from M. smegmatis and 7 from *M. tuberculosis*) were electroporated into *M. tuberculosis* H37Ra. All 16 promoter plasmids conferred chloramphenicol resistance on *M. tuberculosis* H37Ra. As shown in Fig. 1, all of the promoter plasmids exhibited comparable CAT activities in *M. tuberculosis* and *M. smegmatis*, with the exception of S65. The CAT activities supported by some of the mycobacterial promoter plasmids were also compared in *M. bovis* BCG and *M. smegmatis* (Fig. 1, inset), and it was observed that these promoters supported equivalent CAT activities in both species. Moreover, evaluation of the TSP of two of the mycobacterial promoter constructs (T150 and S30) in *M. tuberculosis* and *M. smegmatis* showed that transcription initiates from the same position in both species (data not shown). Thus, the efficiency, as well as the specificity, of transcriptional recognition in these species of mycobacteria appears to be conserved.

Alignment of mycobacterial promoter sequences on the basis of their TSP. Nucleotide sequences of 35 mycobacterial promoter constructs, including 20 from *M. smegmatis* and 15 from *M. tuberculosis*, were determined. The TSP of several promoters were determined by the primer extension technique (Fig. 2). Since the RNase-processed ends may also contribute to these results, as a precautionary measure the primer extension experiment for each promoter construct was repeated a minimum of five times to define a 5′ RNA end as a TSP. No additional distinct transcription products were observed upstream of the designated TSP when autoradiograms were overexposed. In addition, we were able to detect a specific conserved sequence present at nearly the same position upstream (−10 region) of the TSP for each of the promoter clones. These conserved sequences corresponded to the Pribnow box present in nearly all classical prokaryotic promoters. An alignment of 24 mycobacterial promoters (14 promoters from *M. smegmatis* and 10 promoters from *M. tuberculosis*) is shown in Fig. 3. Transcription initiated in a majority of the cases (9 of 14 promoters of *M. smegmatis* and 7 of 10 promoters of *M. tuberculosis*) at a purine nucleotide. Analysis of the promoter sequences from positions −1 to −50 showed that the promoters from *M. tuberculosis* have a higher GC content (57%) than the promoters from *M. smegmatis* (43%).

All of the mycobacterial promoters analyzed in this study contain a highly conserved Pribnow box-like sequence around the −10 position (Table 1). The conserved hexameric sequences were 100% T, 93% A, 50% T, 57% A, 43% a, and 71% T for *M. smegmatis* promoters and 80% T, 90% A, 60% Y, 40% g, 60% A, and 100% T for *M. tuberculosis* promoters (Table 1 footnote). The first, second, and sixth bases in the hexameric sequence (T, A, and T, respectively), which represent the functionally more important positions in *E. coli* promoters (24), are most strongly conserved in both mycobacterial species. However, upon the basis of the limited number of promoter sequences analyzed in this study, it appears that in *M. tuberculosis*, less important positions (the third and fourth) have become relatively more susceptible to GC intrusions to accommodate the higher GC content of its promoters.

The most prominent features of mycobacterial promoters are that the sequences in the −35 regions do not bear any homology with the TTGACA motif present in promoters of *E. coli* and several other bacteria and that apparently no single sequence was found to be conserved in the −35 regions of mycobacterial promoters.

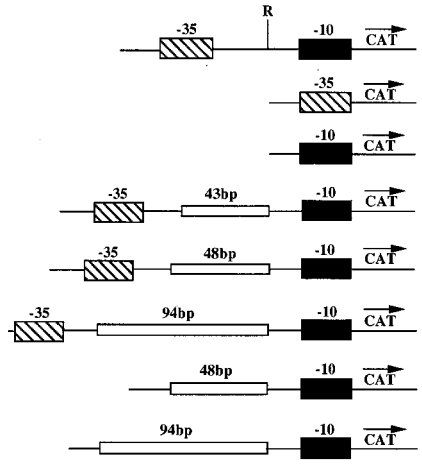
Functional dissection of mycobacterial promoter regions. To determine the functional organization of mycobacterial promoters, we performed several modifications of *M. smegmatis* promoter construct pS35 and *M. tuberculosis* promoter construct pT3 (Fig. 4).

Promoter constructs pS35.1 and pS35.2, harboring only the

TABLE 1. Conservation of sequences in the −10 regions of mycobacterial promoters[a]

M. smegmatis		*M. tuberculosis*	
Promoter	−10 region	Promoter	−10 region
S4	TATAAG	T3	TAACAT
S5	TAAGCT	T6	TAGGCT
S12	TATACA	T26	GATGTT
S14	TACCAT	T80	CAGGAT
S16	TATAAT	T101	TAGTGT
S18	TAAAGT	T119	TACCAT
S19	TAGTAT	T125	TATTAT
S21	TATGGT	T129	TTCAAT
S30	TATATG	T130	TATGCT
S33	TATGGT	T150	TATCAT
S35	TATGAT		
S65	TCCAGT		
S69	TACAAT		
S119	TAAACG		

[a] The consensus for the *M. smegmatis* and *M. tuberculosis* promoters was determined by aligning the mycobacterial promoter sequences on the basis of their TSP. The frequencies of occurrence of bases are as follows (capital letters indicate bases exhibiting conservation of 50% or more, and the letter Y indicates a pyrimidine base). *M. smegmatis* −10 region, T, 100%; A, 93%; T, 50%; A, 57%; a, 43%; T, 71%. *M. tuberculosis* −10 region: T, 80%; A, 90%; Y, 60%; g, 40%; A, 60%; T, 100%. *E. coli* −10 consensus: T, 77%; A, 76%; T, 60%; A, 61%; A, 56%; T, 82%.



Promoter construct	CAT specific activity	Promoter construct	CAT specific activity
pS35	640	pT3	1125
pS35.1	2	pT3.1	2
pS35.2	50	pT3.2	176
pS35.3	a) 314 b) 100	pT3.3	a) 771 b) 2000
pS35.4	a) 51 b) 235	pT3.4	a) 68 b) 700
pS35.5	a) 402 b) 75	pT3.5	a) 470 b) 1350
pS35.4′	a) 48 b) 231	pT3.4′	a) 70 b) 650
pS35.5′	a) 414 b) 67	pT3.5′	a) 500 b) 1450

FIG. 4. Functional dissection of the −10 and −35 regions of *M. smegmatis* promoter S35 and *M. tuberculosis* promoter T3. Promoter construct pS35 represents pSD7 containing promoter S35 of *M. smegmatis*, and pT3 represents pSD7 containing promoter T3 of *M. tuberculosis*. The letter R represents the *Xba*I restriction site, situated at the −20 position with respect to the TSP of promoter S35, and the *Kpn*I restriction site, situated at the −22 position with respect to the TSP of promoter T3. The definitions of the modified promoter constructs are as follows. The −10 and −35 regions were deleted from pS35 to yield pS35.1 and pS35.2, respectively. pS35.3a and pS35.3b contain the 43-bp *Sph*I-*Pst*I fragment, derived from the multiple cloning region of pGEM-5Zf(+), cloned into the *Xba*I site in either orientation. pS35.4a and pS35.4b contain the 48-bp *Nco*I-*Nde*I fragment cloned in either orientation. pS35.5a and pS35.5b contain the 94-bp *Apa*I-*Nsi*I fragment cloned in either orientation. Constructs pS35.4′a to pS35.5′b were obtained by deleting the original −35 region from the respective parent promoter constructs, namely, pS35.4a to pS35.5b. Identical modifications were performed with the *Kpn*I site in the pT3 constructs as described in the text. The CAT specific activity supported by each modified construct is expressed as nanomoles per minute per milligram of protein.

−35 region and only the −10 region, respectively, were tested for the ability to confer chloramphenicol resistance on *M. smegmatis*. Construct pS35.1 failed to transform *M. smegmatis* to chloramphenicol resistance and did not support a significant level of CAT activity, indicating that the −35 region alone is insufficient to support transcription in mycobacteria. However, construct pS35.2 retained enough CAT activity (Fig. 4) to confer chloramphenicol resistance on *M. smegmatis*. To interpret these observations, we carried out further modifications of promoter construct pS35. Three different DNA fragments, derived from the multiple cloning region of plasmid pGEM-5Zf(+), were separately cloned into the unique *Xba*I site of pS35 in both orientations (Fig. 4). Insertion of different sequences between the −10 and −35 regions in pS35 resulted in the generation of mosaic promoters (pS35.3a to pS35.5b in Fig. 4) with diverse −35 regions that supported different CAT activities ranging from 8 to 65% of the activity of the parent construct (Fig. 4).

The functional contributions of different regions of the promoter construct pT3 were also evaluated by carrying out modifications similar to those described for promoter construct pS35. Modified promoter constructs pT3.1 and pT3.2, containing only the −35 region and only the −10 region of promoter T3, respectively, were tested for chloramphenicol resistance and CAT activity in *M. smegmatis*. Construct pT3.1 failed to confer chloramphenicol resistance on *M. smegmatis* and did not support any significant CAT activity, indicating an absolute requirement of the −10 region for initiation of transcription. However, pT3.2 retained sufficient CAT activity, despite deletion of the −35 region (Fig. 4), to confer chloramphenicol resistance on *M. smegmatis*. Subsequently, we cloned three different DNA fragments from pGEM-5Zf(+) (as described earlier for the S35 promoter) into the *Kpn*I site in plasmid pT3 in both orientations. The resulting constructs (pT3.3a to pT3.5b) also transformed *M. smegmatis* to chloramphenicol resistance. The modified constructs supported CAT activities ranging from 6 to 200% of the activity supported by pT3 (Fig. 4).

The fact that constructs pS35.3a to pS35.5b, as well as constructs pT3.3a to pT3.5b, retained transcriptional activity despite the modifications in the promoter elements raised the possibilities that either the sequence in the inserted DNA fragment contributed an alternative −35 sequence or the original −35 sequence was still functional as a promoter module despite the distance created by the insertions. However, it is noteworthy that insertion of each fragment in both orientations (which had no effect on the position of the original −35 region) resulted in drastically different CAT activities (Fig. 4). Insertion of each fragment in different orientations created different sequences at the −35 region, which were responsible for such variable CAT activities. Moreover, when we deleted the original −35 sequence from eight of these modified constructs (pS35.4a to pS35.5b and pT3.4a to pT3.5b), the resulting new constructs, viz., pS35.4′a to pS35.5′b and pT3.4′a to pT3.5′b, exhibited CAT activities comparable to those of their respective parent constructs (Fig. 4). These results show that alternative sequences generated at the −35 position, rather than the original −35 sequence acting from a position further upstream, were responsible for the observed transcriptional activities of all of the modified promoter constructs. It is worth mentioning here that the sequences in the −35 regions of these modified promoter constructs (pS35.3a to pS35.5b and pT3.3a to pT3.5b) exhibited no significant homology within themselves.

Assessment of *E. coli* promoters in mycobacteria. Since the −10 regions in the promoters of both *E. coli* and mycobacteria have common sequence requirements and the −35 region of mycobacterial promoters appears to tolerate a greater variety of sequences than does that of *E. coli* promoters, our prediction was that a larger percentage of *E. coli* promoters should be recognized by the mycobacterial RNA polymerase than vice versa. To perform a systematic analysis, we first determined the

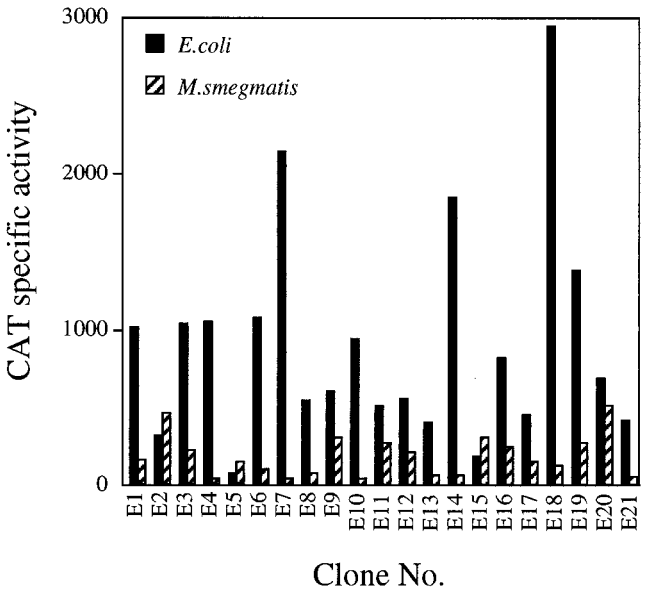


FIG. 5. Comparative assessment of *E. coli* promoters in *E. coli* and *M. smegmatis*. The CAT activity supported by each *E. coli* promoter plasmid was determined from the respective *E. coli* and *M. smegmatis* transformants. CAT specific activity is expressed as nanomoles per minute per milligram of protein.

TSP of two *M. smegmatis* promoters, namely, S16 and S30 (which also supported transcription in *E. coli*) (8) from the respective *M. smegmatis* and *E. coli* transformants. In both cases, transcription initiated at the same position (data not shown). These observations suggested that the mycobacterial and *E. coli* RNA polymerases recognize similar regions in the promoter fragments. We then constructed a random library of *E. coli* promoters in promoter probe vector pSD7. Plasmids isolated from 100 individual *E. coli* promoter clones were individually electroporated into *M. smegmatis*. Seventy-five *E. coli* promoter plasmids conferred chloramphenicol resistance on *M. smegmatis*, in contrast to our earlier observations that only 12 of 100 *M. smegmatis* promoters conferred chloramphenicol resistance on *E. coli* (8). The CAT activities supported by 21 randomly selected *E. coli* promoters were compared in *E. coli* and *M. smegmatis* (Fig. 5). Generally, the *E. coli* promoters supported very suboptimal CAT activity in mycobacteria, but this was particularly true of strong *E. coli* promoters. A noteworthy observation, however, was that about two-thirds of the *E. coli* promoters supported CAT activities in *M. smegmatis* that were significantly higher than 10% of the corresponding CAT activities in *E. coli*, including four promoters (E2, E5, E15, and E20) which supported nearly equal or higher CAT activities in *M. smegmatis* compared with the corresponding activities in *E. coli*. On the contrary, we had observed earlier that most of the promoters from *M. smegmatis* and *M. tuberculosis* supported CAT activities in *E. coli* that were less than 10% of the corresponding activities in *M. smegmatis* (8). These results corroborate our observations that the −35 regions of mycobacterial promoters can tolerate a larger variety of sequences than can those of *E. coli.*

DISCUSSION

In this communication, we describe the characterization of promoters that we had earlier isolated from *M. smegmatis* and *M. tuberculosis* by employing shuttle vector pSD7 (8). We show that fast-growing *M. smegmatis* and slowly growing *M. tuberculosis* and *M. bovis* BCG recognize mycobacterial promoters with similar efficiencies. This conclusion is based on the comparable CAT activities supported by several mycobacterial promoters in these species (Fig. 1) and is further substantiated by the observation that initiation of transcription as determined for promoters T150 and S30 in *M. smegmatis* and *M. tuberculosis* occurs from identical positions. These results suggest that the basic transcriptional machineries of these mycobacterial species have determinants of transcriptional specificity in common. We therefore believe that *M. smegmatis* can be safely used as a surrogate host for expression of at least the constitutively expressed genes from slowly growing pathogenic mycobacteria. The situation may be different for promoters of specifically regulated genes, as observed by Timm and coworkers (28) for the *bla* gene promoter from *M. fortuitum* and the *p*AN and *p*sul3 promoters from mobile genetic elements, and possibly for promoter S65 in this study (Fig. 1).

The alignment of mycobacterial promoters upstream of the TSP revealed a highly conserved, Pribnow box-like hexamer located around the −10 position in both *M. tuberculosis* and *M. smegmatis*. However, our results suggest that GC pressure has affected the promoter regions of *M. smegmatis* and *M. tuberculosis* to different extents, with accumulation of relatively higher GC content in *M. tuberculosis* promoters. Sequence comparison within *E. coli* promoters reveals considerable diversity in the conserved hexamers, although these promoters must have structural features necessary for interaction with RNA polymerase in common (19). A similar situation may exist in *M. smegmatis* and *M. tuberculosis* promoters with respect to the −10 region, despite the differences in GC content between the consensus sequences. This is substantiated by our observation that a given promoter is recognized equally well by RNA polymerases from *M. smegmatis* and *M. tuberculosis* H37Ra. The genomic GC contents of *M. smegmatis* and *M. tuberculosis* species are comparable (6); hence, a high GC content in the promoters of *M. tuberculosis* may have been selected during the course of evolution. It is not clear what the driving force is for selection of this higher GC content in the promoter regions of *M. tuberculosis* compared with *M. smegmatis. M. tuberculosis* promoters are usually weaker than *M. smegmatis* promoters (8). Possibly, this higher GC content of *M. tuberculosis* promoters has a bearing on the lower strength of its promoters. Being an intracellular parasite of humans, *M. tuberculosis* evades the immune response of the host by staying dormant in macrophages; hence, weaker promoters may be advantageous to its pathogenic nature.

Among the sequences present in the −35 regions of mycobacterial promoters, we were unable to find a single strongly conserved sequence. Although some promoters in mycobacteria have been reported to carry in the −35 region a sequence that closely resembles the typical prokaryotic −35 region consensus TTGACA (17) (T150 in Fig. 3), the absence of this *E. coli*-like sequence appears to be a distinctive feature of mycobacterial promoters. Our experiments on the functional dissection of mycobacterial promoter regions (Fig. 4) substantiate our findings that the −35 regions of mycobacterial promoters can accommodate a larger variety of sequences than the −10 regions, albeit with variable transcriptional activities. These observations are consistent with those reported by Sarkis and colleagues (25) for the firefly luciferase gene fused to the mycobacteriophage L5 genome. Those workers had reported that the firefly luciferase gene could apparently be expressed in mycobacteria when sequences upstream of the −10 region were replaced with two different sequences from mycobacteriophage L5 DNA (25). Kremer and coworkers observed that deletion of 4 bp or insertion of 64 bp between the −10 and −35

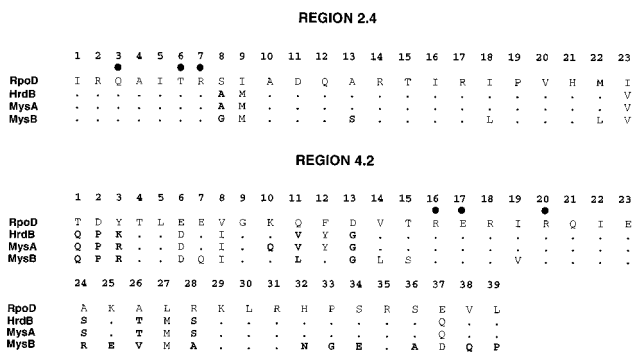


FIG. 6. Comparison of the amino acid sequences in the promoter binding domains of the sigma factors of *M. smegmatis* (MysA and MysB) and the principal sigma factors of *S. aureofaciens* (HrdB) and *E. coli* (RpoD). Regions 2.4 (responsible for binding to the −10 region of the promoter) and 4.2 (responsible for binding to the −35 region of the promoter) of the four sigma factors are aligned. The amino acids already demonstrated to participate directly in making specific contacts with promoter sequences are highlighted by large dots. Small dots represent identical amino acids, conserved substitutions are represented by plain letters, and nonconserved substitutions are shown by boldface letters. Conserved substitutions are defined as the following groups: I, L, M, and V; A and G; S and T; K, H, and R; D, E, N, and Q; F, Y, and W; C; and P. The amino acid sequences are from reference 20 for RpoD of *E. coli*, from reference 16 for HrdB of *S. aureofaciens*, and from reference 23 for MysA and MysB of *M. smegmatis*. The numbers above the amino acids do not represent actual amino acid positions in the respective sigma factors.

regions of the *M. tuberculosis* 85A antigen promoter abolished only 50% of the promoter activity (17) and suggested that although the sequence at the −35 position is essential for transcription, its location may not be critical. However, those investigators did not provide any experimental evidence that the original −35 sequence, and not an alternative −35 sequence generated as a result of the insertion, was responsible for the observed activity. By removing the original −35 sequence from constructs S35.4a to S35.5b and T3.4a to T3.5b (Fig. 4), we were able to show that the alternative sequence generated at the −35 site, rather than the original −35 sequence acting from a different location, was responsible for the observed transcriptional activity. These findings are further supported by our observations that a much larger proportion of *E. coli* promoters function in mycobacteria than vice versa.

The tolerance of large variety of sequences in the −35 region of mycobacterial promoters distinguishes them from most other prokaryotic promoters. To our knowledge, such features in the transcriptional signals have been reported only for *Streptomyces* spp., which, like mycobacteria, belong to the actinomycete family. Strohl has studied 139 streptomycete promoters and shown that a majority (80%) of streptomycete promoters do not function in *E. coli*, show no homology to *E. coli* promoters in the −35 region, and contain diverse sequences in the −35 region (27). The presence of multiple sigma factors (3, 31) and the potential for a high degree of transcriptional flexibility resulting from the capacity to synthesize different sigma factors is now firmly established in *Streptomyces* spp. (31). Mycobacteria may also contain multiple sigma factors, which might explain the heterogeneity in the −35 regions of promoters. Two constitutively expressed sigma factors have already been identified from *M. smegmatis*, *M. tuberculosis*, and *M. leprae*, and the corresponding genes *mysA* and *mysB* have been cloned and sequenced from *M. smegmatis* (23). Predich and colleagues have proposed MysA as the principal sigma factor of *M. smegmatis* (23). Given the observations that (i) mycobacterial promoters may have features in common with streptomycete promoters and (ii) mycobacterial genes are usually better expressed in *Streptomyces* spp. than in *E. coli* (14), we compared the sequences of the 2.4 (responsible for binding to the −10 region) and 4.2 (responsible for binding to the −35 region) regions of MysA and MysB with the corresponding regions of the principal sigma factors of *Streptomyces aureofaciens* (HrdB) and *E. coli* (RpoD) (Fig. 6) to elucidate the nature of similarities and/or differences between the transcriptional apparatus of these genera. The amino acid sequences of the 2.4 regions of MysA and HrdB are identical and differ from the corresponding sequence in RpoD in only 3 of a total of 23 amino acids. Two of these are conserved changes, implying that the −10 binding regions in the sigma factors of these three bacteria are nearly identical (Fig. 6). However, we observed a higher degree of variability in the 4.2 regions of these sigma factors. In a stretch of 39 amino acids, the sequences of MysA and HrdB are nearly identical, except at positions 3 and 10, where lysine residues in HrdB have been replaced by arginine and glutamine residues, respectively, in MysA. However, MysA differs from RpoD at 14 positions, of which 9 represent nonconserved changes (Fig. 6). Our earlier observations that a majority of mycobacterial promoters do not function in *E. coli* (8) could be ascribed to poor interaction of the 4.2 region of the *E. coli* sigma factor with the −35 regions of mycobacterial promoters.

The close identity of MysA and HrdB in the 2.4 and 4.2 regions emphasizes the similarities in the transcriptional apparatuses of mycobacteria and *Streptomyces* spp. and provides a plausible explanation for the fact that *Streptomyces* is a better host than *E. coli* for expression of mycobacterial genes (14). It is significant that the amino acid sequences of MysA and MysB in the 4.2 region differ at 18 of 39 positions and 12 of these changes are nonconserved (Fig. 6). It is conceivable that the differences in the −35 binding domains of MysA and MysB reflect on the diverse sequences present in the −35 regions of mycobacterial promoters. The total cellular transcription in mycobacteria may represent a division of labor involving RNA polymerase with associated sigma factors, each recognizing different or overlapping consensus sequences in the −35 regions. Although only one sigma factor with a single −35 consensus is responsible for transcription from all vegetatively expressed genes in *E. coli*, a large number of *E. coli* promoters exhibit sequences in their −35 regions which vary to different extents from the consensus element TTGACA. All of these sequences are, however, recognized by the *E. coli* RNA polymerase, albeit with differing efficiencies. It is therefore conceivable that if a cell expresses multiple vegetative sigma factors with different consensus sequence requirements for the −35 regions of promoters, the total number of sequences which can function at the −35 region of its promoters would be much larger than in *E. coli* and consequently the sequence requirement in the −35 regions of promoters may seem degenerate, as reported earlier for *Streptomyces* spp. and seen in this study for mycobacteria, if these promoters are analyzed on the basis of alignment of sequences with respect to the TSP. In vitro transcription experiments with a number of mycobacterial promoters using reconstituted RNA polymerase containing either MysA or MysB should provide some definite answers. It is not clear how the presence of more than one sigma factor may be useful for these organisms. A high degree of transcriptional flexibility may be helpful in responding to varying environmental conditions, as well as to support the complex series of interactions with the host.

We believe that a rational approach to understanding the transcriptional signals of mycobacteria will enhance our knowledge about regulation of gene expression in mycobacteria and will also have a significant impact in elucidating the mecha-

nism(s) of pathogenesis of these organisms at the molecular level. These studies will also help in the development of new tools for the study of mycobacterial molecular genetics, as well as for the expression of native and foreign genes of immunological interest in BCG for development of more effective recombinant vaccines.

ACKNOWLEDGMENTS

We thank Manisha Jain and Birinder Kaur for excellent technical assistance, Shruti Jain and Jaya S. Tyagi for critical reading of the manuscript, and Rajiv Chawla and Yatinder Mohan Dua for efficient preparation of the manuscript.

M.D.B., D.K., and S.K.D. are grateful to CSIR for financial assistance. This work was supported by a grant from the Department of Biotechnology, New Delhi, India.

REFERENCES

1. **Bashyam, M. D., and A. K. Tyagi.** 1994. An efficient and high-yielding method for isolation of RNA from mycobacteria. BioTechniques **17:**834–836.
2. **Bloom, B. R., and C. J. L. Murray.** 1992. Tuberculosis: commentary on a reemergent killer. Science **257:**1055–1064.
3. **Buttner, M. J.** 1989. RNA polymerase heterogeneity in *Streptomyces coelicolor* A3(2). Mol. Microbiol. **3:**1653–1659.
4. **Chaisson, R. E., and G. Slutkin.** 1989. Tuberculosis and human immunodeficiency virus infection. J. Infect. Dis. **159:**96–100.
5. **Cirillo, J. D., T. R. Weisbrod, L. Pascopella, B. R. Bloom, and W. R. Jacobs, Jr.** 1994. Isolation and characterization of the aspartokinase and aspartate semialdehyde dehydrogenase operon from mycobacteria. Mol. Microbiol. **11:**629–639.
6. **Clark-Curtiss, J. E.** 1990. Genome structure of mycobacteria, p. 77–96. *In* J. McFadden (ed.), Molecular biology of mycobacteria. Surrey University Press, London.
7. **Culliton, B. J.** 1992. Drug-resistant TB may bring epidemic. Nature (London) **356:**473.
8. **DasGupta, S. K., M. D. Bashyam, and A. K. Tyagi.** 1993. Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. J. Bacteriol. **175:**5186–5192.
9. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2:**1044–1051.
10. **Harshey, R. M., and T. Ramakrishnan.** 1977. Rate of ribonucleic acid chain growth in *Mycobacterium tuberculosis* $H_{37}R_v$. J. Bacteriol. **129:**616–622.
11. **Jacobs, W. R., Jr., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. A. Udani, W. Jones, R. G. Barletta, and B. R. Bloom.** 1991. Genetic systems for mycobacteria. Methods Enzymol. **204:**537–555.
12. **Ji, Y.-E., M. J. Colston, and R. A. Cox.** 1994. Nucleotide sequence and secondary structures of precursor 16S rRNA of slow growing mycobacteria. Microbiology **140:**123–132.
13. **Ji, Y.-E., M. J. Colston, and R. A. Cox.** 1994. The ribosomal RNA (*rrn*) operons of fast-growing mycobacteria: primary and secondary structures and their relation to *rrn* operons of pathogenic slow-growers. Microbiology **140:** 2829–2840.
14. **Kieser, T., M. T. Moss, J. W. Dale, and D. A. Hopwood.** 1986. Cloning and expression of *Mycobacterium bovis* BCG DNA in "*Streptomyces lividans*." J. Bacteriol. **168:**72–80.
15. **Kong, T. H., A. R. M. Coates, P. D. Butcher, C. J. Hickman, and T. M. Shinnick.** 1993. *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. Proc. Natl. Acad. Sci. USA **90:**2608–2612.
16. **Kormanec, J., M. Farkasovsky, and L. Potuckova.** 1992. Four genes in *Streptomyces aureofaciens* containing a domain characteristic of principal sigma factors. Gene **122:**63–70.
17. **Kremer, L., A. Baulard, J. Estaquier, J. Content, A. Capron, and C. Locht.** 1995. Analysis of the *Mycobacterium tuberculosis* 85A antigen promoter region. J. Bacteriol. **177:**642–653.
18. **Levin, M. E., and G. F. Hatfull.** 1993. *Mycobacterium smegmatis* RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. Mol. Microbiol. **8:**277–285.
19. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:**1507–1516.
20. **Lonetto, M., M. Gribskov, and C. A. Gross.** 1992. The σ^{70} family: sequence conservation and evolutionary relationships. J. Bacteriol. **174:**3843–3849.
21. **Murray, A., N. Winter, M. Lagranderie, D. F. Hill, J. Rauzier, J. Timm, C. Leclerc, K. M. Moriarty, M. Gheorghiu, and B. Gicquel.** 1992. Expression of *Escherichia coli* β-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.
22. **Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull.** 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol. Microbiol. **17:**1045–1056.
23. **Predich, M., L. Doukhan, G. Nair, and I. Smith.** 1995. Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. Mol. Microbiol. **15:**355–366.
24. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:** 319–353.
25. **Sarkis, G. J., W. R. Jacobs, Jr., and G. F. Hatfull.** 1995. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. Mol. Microbiol. **15:**1055–1067.
26. **Sela, S., and J. E. Clark-Curtiss.** 1991. Cloning and characterization of the *Mycobacterium leprae* putative ribosomal RNA promoter in *Escherichia coli*. Gene **98:**123–127.
27. **Strohl, W. R.** 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. **20:**961–974.
28. **Timm, J., E. M. Lim, and B. Gicquel.** 1994. *Escherichia coli*-mycobacteria shuttle vectors for operon and gene fusions to *lacZ*: the pJEM series. J. Bacteriol. **176:**6749–6753.
29. **Timm, J., M. G. Perilli, C. Duez, J. Trias, G. Orefici, L. Fattorini, G. Amicosante, A. Oratore, B. Joris, J. M. Frere, A. P. Pugsley, and B. Gicquel.** 1994. Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* β-lactamase genes cloned from a natural isolate and a high-level β-lactamase producer. Mol. Microbiol. **12:**491–504.
30. **Verma, A., A. K. Kinger, and J. S. Tyagi.** 1994. Functional analysis of transcription of the *Mycobacterium tuberculosis* 16S rDNA-encoding gene. Gene **148:**113–118.
31. **Westpheling, J., M. Ranes, and R. Losick.** 1985. RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature (London) **313:**22–27.