1995

# Compilation and analysis of Mycobacterium paratuberculosis promoters

John Patrick Bannantine
*Iowa State University*

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road. Ann Arbor. MI 48106-1346 USA
313/761-4700 800/521-0600

# Compilation and analysis of *Mycobacterium paratuberculosis* promoters

by

John Patrick Bannantine

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Microbiology, Immunology and Preventive Medicine
Major: Microbiology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1995

TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

## ABSTRACT

*Mycobacterium paratuberculosis* transcriptional and translational signals were studied to gain insight into gene expression in this organism. To study *M. paratuberculosis* promoter structures, a more versatile promoter selection vector was constructed from the pKO1 parent vector. This new plasmid, pJJ2, was used to identify 11 promoter fragments from an *M. paratuberculosis* DNA library. In addition, a previously characterized *M. paratuberculosis* promoter, contained in a 493-bp *Eco*RI fragment, was cloned into the new vector to test the efficacy in cloning novel *M. paratuberculosis* promoters. In a related study, an expression probe shuttle plasmid (pYUB76) was employed to clone *M. paratuberculosis* expression signals that could be studied directly in mycobacteria. Using this vector to identify *M. paratuberculosis* expression signals, we have determined the nucleotide sequence of ten promoter-containing fragments and have compared these sequences to those of several previously reported *Mycobacterium* promoters. Hexanucleotide sequences centered approximately 35 and 10 base pairs upstream from the experimentally determined transcription start sites revealed a consensus that is different from *E. coli.* Compilation of these promoter sequences identified the -35 (T, T/G, G, A/C, G, T) and the -10 (C, A, G, C, C, G) conserved hexanucleotides.

# GENERAL INTRODUCTION

## Objective

*Mycobacterium paratuberculosis* is the etiologic agent of paratuberculosis, a chronic intestinal granulomatous infection of domestic or captive ruminants characterized by multiplication of the organisms within macrophages in the intestinal mucosa (Chiodini *et al.*, 1984; Merkal 1984; Thoen and Baum, 1988). A detailed study of the molecular basis of pathogenesis in *M. paratuberculosis* is crucial to the understanding of the organism's ability to survive within host macrophages as well as to gain insight for vaccine development. The transcriptional machinery of mycobacteria in general, and *M. paratuberculosis* in particular, has not been well defined. *M. paratuberculosis* is seldom included in studies of this genus because of its extremely slow growth and culture difficulties. Moreover, the potential health hazard associated with its manipulation may have deterred many workers. Mycobacteria are among the most difficult prokaryotes to lyse due to the thick, lipid-rich composition of their cell walls. Chemical or enzymatic lysis is relatively tedious and requires long incubation periods. Moreover, mechanical disruption often results in incomplete lysis (Hurley *et al.*, 1987; Hines *et al.*, 1991). Thus, *M. paratuberculosis* remains an underdeveloped genetic system. To date, only four genes have been cloned and sequenced (Stevenson *et al.*, 1991; Gilot *et al.*, 1993; Green *et al.*,

1989; Hance *et al.*, 1989) and only two promoters have been characterized in this species (Murray *et al.*, 1992; Thomas *et al.*, 1992). A method of genetic exchange has only recently been established (Foley-Thomas *et al.*, 1995). By contrast, there are well over 150 characterized *Escherichia coli* promoters that have been assembled and analyzed in detail (O'Neill, 1989).

It was against this background that the project of cloning *M. paratuberculosis* promoters was undertaken. The focus was to extend the current knowledge of promoter architecture and gene expression in *M. paratuberculosis* by characterizing additional promoter elements with the long range goal of adapting this organism to the IVET (*in vivo* expression technology) method recently developed (Mekalanos, 1992; Mahan *et al.*, 1993) and refined (Mahan *et al.*, 1995) in the laboratory of John Mekalanos. Therefore, this project represents an initial step, that being the cloning and characterization of *M. paratuberculosis* promoters, with the ultimate goal of acquiring the ability to search for environmentally induced changes in mycobacterial gene expression which may provide clues to virulence mechanisms. Furthermore, a thorough analysis of *M. paratuberculosis* expression sequences was undertaken with the goal of cloning a strong promoter to optimize the expression of foreign antigens in *Mycobacterium bovis* BCG (Barletta *et al.*, 1990).

Initial studies were aimed at improving upon the promoter-probe vector pKO1 and using it to examine mycobacterial

transcription signals in *E. coli.* One novel promoter-containing *M. paratuberculosis* DNA fragment was cloned and sequenced with the modified pKO1 vector (pJJ2). In addition, a previously characterized *M. paratuberculosis* promoter (Thomas *et al.*, 1992) was cloned into pJJ2 to show that modifications to the vector did not affect the *galK* reporter gene functions. This resulting construct, designated pTB10, was further characterized by subcloning and transcription start site mapping. These studies were abandoned when another promoter-probe vector, pYUB76 (Barletta *et al.*, 1992), became available which could replicate in both *E. coli* and *Mycobacterium smegmatis* hosts. This allowed transcription initiation studies to be performed directly in mycobacteria. To perform these studies, a *M. paratuberculosis* DNA library was cloned into the promoter-probe shuttle vector pYUB76. This vector contains a promoterless *lacZ* gene that serves as the reporter gene in mycobacteria and *E. coli.* Over 100 promoter-containing clones were identified after electroporation of the library into *E. coli* DH10B. Identification was based on *lacZ* (which encodes β-galactosidase) expression on Xgal containing media. In one experiment, 12 *lacZ* positive colonies were selected, minipreped and used to electrotransform *M. smegmatis* $mc^2155$. Then a comparitive analysis of expression levels between the two surrogate hosts was quantitatively assessed. It was observed that when the reporter gene was expressed at high levels in one host, expression in the other host was only transient. Thus it was concluded that there is a

divergence in the transcriptional signals between *E. coli* and mycobacteria. In another experiment, that same library in *E. coli* DH10B was pooled and electroporated into *M. smegmatis*. Only 24 *lacZ* positive colonies were obtained from this experiment of which the 10 strongest were sequenced. Computer analysis of the resulting sequences showed no strong homologies in the databases using default search parameters. Putative open reading frames were identified. Total cell RNA was extracted *from M. smegmatis* harboring the selected *lacZ* positive constructs and transcriptional activity was quantitated using slot blot analysis techniques. In addition the 5' mRNA start sites were identified by primer extension analysis and nucleotide sequences were aligned relative to the start sites. Based on these data, rudimentary consensus promoter sequences were identified.

## Dissertation organization

The research data in this dissertation are presented in the form in which they are to be submitted for publication. Chapter 2 will be submitted to the journal *Gene* and Chapter 3 will be submitted to the *Journal of Bacteriology*. The literature review section, Chapter 1, is intended to provide a general understanding of the progress made in gene expression studies for mycobacteria relative to *E. coli*, a well developed system. Citations from Chapter 1 are listed at the end of the dissertation, whereas the citations from Chapter 2 and 3 are

listed at the end of the respective chapters. Following the last paper (Chapter 3) is a general conclusions section that summarizes the research results and provides a perpectus of future research in the field of mycobacterial gene expression.

# CHAPTER 1. LITERATURE REVIEW

## Introduction

The mycobacteria fall into two general catagories based on their growth rates. The fast growing mycobacteria, consisting of such strains as *M. fortuitum*, *M. phlei*, and *M. smegmatis*, have doubling times of 2-3 h and will produce a colony from a single cell in 3-4 days (Wayne and Kubica, 1986). In contrast, the slow-growing mycobacteria such as *M. avium*, *M. bovis*, *M. tuberculosis*, and *M. paratuberculosis* species double every 18-26 h and thus yield colonies from single cells in 14-28 days. The slow growing mycobacteria are typically pathogenic whereas the fast growing strains are considered saprophytes. Interestingly, the slow growing mycobacteria possess a rRNA operon in single copy and the faster growing species have two copies (Bercovier *et al.*, 1986). Bercovier *et al.* (1986) used a copy of the *E. coli* rRNA operon as a probe in hybridization experiments with chromosomal DNA from two fast-growing strains and four slow growing mycobacteria. *E. coli*, one of the faster growing prokaryotes has seven rRNA operon copies.

*M. paratuberculosis* is a slow-growing, acid-fast bacterium that often survives in the immediate environment, especially in the feces of ruminants (Thoen and Baum, 1988). *M. paratuberculosis* requires 1 to 2 months to isolate on primary culture, and requires an iron chelating compound called mycobactin (Merkal and Curran, 1974;

Jørgensen 1982). The mycobactin requirement exists in other slow-growing strains of *Mycobacterium* (i. e. *M. avium* and *M. intracellulare*) and was first thought to be a characteristic of pathogenic mycobacteria. It was later discovered that some strains simply do not produce enough mycobactin and therefore must aquire it exogenously (Merkal and McCullough, 1982).

Johne's disease is the cause of reduced milk production and survival in cattle, sheep, goats and other ruminant animals (Thoen and Baum, 1988). The disease is prevalent in parts of the U. S. (Merkal *et al.*, 1987; Thoen and Baum, 1988) and has been reported in countries throughout the world (Chiodini *et al.*, 1984; NcNab *et al.*, 1991). Economic losses for the U. S. dairy industry exceed 1.5 billion annually (Cocito *et al.*, 1994). *Mycobacterium paratuberculosis* has been identified as the etiologic agent (Jöhne and Frothingham, 1895). Isolation of *M. paratuberculosis* from intestinal tissue of Crohn's patients, has lead to the suspicion of a potential pathogenic role for humans (Cocito *et al.*, 1994; McFadden *et al.*, 1987; McFadden *et al.*, 1992).

*M. paratuberculosis* enters young animals from the mother's milk, crosses the intestinal mucosa and is phagocytized by macrophages in the lamina propria (Momotani *et al.*, 1988). The acid-fast organism is resistant to the antimicrobial activity of the macrophages and, in fact, may utilize its intracellular environment for nutrients (Merkal, 1984). Paratuberculosis involves an impairment of the function of the intestinal passages, which produce

the main symptoms of diarrhea and weight loss. The disease progresses through three clinical stages (Cocito *et al.*, 1994). In stages one and two, the animal has no signs of disease and is catagorized as subclinical. The detectable excretion of organisms defines the second stage from the first. In the third stage, the animal is both clinical and excretory. The dominant late signs are emaciation, decreased milk production, loss of appetite, diffuse edema and anemia. The economic impact is felt when infected cattle do not reach market weight and when milk production is low, both of which are directly attributable to disease symtoms. The organism has been found to multiply within the intestinal tract, and produces an aggrevated inflammatory reaction (massive migration of neutrophils, macrophages, giant cells, etc.) that has been implicated as the cause of the lack of nutrient absorption. The fact that the organism displays a restricted affinity for the intestinal tract is in contrast to other pathogenic mycobacteria that have affinity for lung tissue (i. e. *M. tuberculosis*) or nerve tissue (i. e. *M. leprae*). In general, mycobacterial pathogens induce chronic disease that evolves over a period of years, with pathological damage resulting from detrimental activation of the host immune system rather than from release of potent toxins by the bacteria. Therefore mycobacterial virulence is related more to their ability to persist in tissues and to activate the immune system rather than the direct damage of host cells. Resistance to killing by host immune cells is, in part, a reflection of the structure of the mycobacterial cell wall.

Mycobacteria synthesize a lipid-rich cell wall which undoubtedly confers some degree of protection against the damaging effects of complement components and phagocyte-derived free radicals (Chan *et al.*, 1989).

Most of the current literature on *M. paratuberculosis* is directed at diagnostic methods that will identify the organism faster than culture techniques (Abbas *et al.*, 1983; Ambrosio *et al.*, 1991; Billman *et al.*, 1992; Hurley *et al.*, 1989; Molina *et al.*, 1991; Moss *et al.*, 1991; Poupart *et al.*, 1993; Thoresen and Saxegaard, 1991). As mentioned previously, *M. paratuberculosis* is an extremely slow growing bacterium, and, as such, identification by culture methods is inefficient for the purpose of diagnosis. The prolonged incubation time and the presence of subclinical cases, permits infected animals to shed large amounts of bacilli in their feces before detection. Moreover, current diagnostic methodology is unable to detect infection in the early stages of the disease (Chiodini *et al.*, 1984; Cocito *et al.*, 1994). Early detection is also hindered by the lack of specific DNA probes that do not cross react with related species. IS*900* has been examined as a possible diagnostic probe (Moss *et al.*, 1991), but it has been found to cross react with *M. avium*, a closely related strain (Kunze *et al.*, 1992). Idexx coroporation has a commercially available IS*900*-based diagnostic kit that the company claims will identify *M. paratuberculosis* from infected feces. The complement fixation test and enzyme-linked immunosorbent assay tests lack both sensitivity and specificity. The fecal culture remains

the most reliable diagnostic test, but results may take up to 2 months due to the slow growth of the organism. In addition, animals that shed a low number of bacilli may give false-negative results. Recently, polymerase chain reaction tests have been developed to diagnose *M. paratuberculosis*. Vary *et al.*, (1990) used a DNA sequence from IS*900*, which is present in multiple copies in the genome of *M. paratuberculosis*, to develop a *M. paratuberculosis* specific PCR reaction for the diagnosis of Johne's disease. With this assay, detection of 100 bacilli/ml of feces was demonstrated.

Few *M. paratuberculosis* virulence determinants have been studied. Lipoarabinoman (Sugden *et al.*, 1991) and glycopeptidolipid (Camphausen *et al.*, 1985) are major immunogens, but their function in pathogenesis remains unclear.

## Procaryotic promoter biology

Gene expression in bacteria is regulated primarily at the level of transcription. Transcription and its regulation have been studied extensively in several gram-negative bacteria, especially *E. coli*, but mostly neglected in gram-positive (with the exception of *Bacillus subtilis*) and acid-fast bacteria. The first step in transcription is the binding of RNA polymerase to a DNA sequence. This sequence is termed a promoter region. The events which must occur at a promoter site to initiate transcription include the attachment of RNA polymerase in the correct orientation and locally separating the DNA

strands to expose the bases to be transcribed. The sigma (σ) subunit of RNA polymerase is involved in the recognition of a promoter region, which has been elegantly illustrated in *Bacillus subtilis* (Losick and Pero, 1981) by substitution of four different σ factors during sporulation. Additionally, some lytic phages encode new σ subunits for the host RNA polymerase, which act to program transcription of phage-specific genes (Talkington and Pero, 1979). Two regions in most σ factors in bacteria probably determine the specificity of promoter utilization by making sequence specific contacts at two regions of promoters located approximately 10 and 35 bp upstream from the start point of transcription (Rezinkoff *et al.*, 1985). The promoter that is recognized by RNA polymerase when the enzyme is combined with one specific σ factor (σ 70), is considered the typical consensus prokaryotic promoter (Reznikoff *et al.*, 1985). Environmental changes or other circumstances may lead to a replacement of σ 70 by other σ factors. When this occurs, the promoter specificity of the RNA polymerase is altered so that a different group of genes is expressed. For example, under conditions of heat stress, the alternative σ factor is σ 32 in *E. coli* (Reznikoff *et al.*, 1985). In *Pseudomonas aeruginosa*, σ 54-dependent promoters contain a conserved GG-$N_{10}$-GC motif (Savioz *et al.*, 1993) which is far removed from the typical hexameric sequences recognized by σ 70.

Strong or weak promoters are classified as such based on the frequency of transcription initiation where the higher the frequency, the stronger the promoter. The frequency of transcription initiation

in *E. coli* can vary from genes which are transcribed less than once per generation to genes that are transcribed once every other second (McClure, 1985). Studies on *E. coli* promoters have suggested that the following factors can affect the strength of a promoter: (1) nucleotide sequence of the -35 region, (2) nucleotide sequence of the -10 region, (3) spacing between the -35 and -10 regions, (4) nucleotide sequence in the flanking region upstream from the -35 region (Rosenberg and Court, 1979). Factors (1), (2), and (3) will be analyzed in detail throughout the present study on *M. paratuberculosis* promoter elements. Concerning factor (4), some powerful *E. coli* promoters such as the λ phage $p_L$ are preceded by adenine and thymine (A+T) rich blocks (Horn and Wells, 1981a). Evidence has been presented that the upstream A+T-rich region is responsible for the strong promoter activity in the λ $p_L$ promoter (Horn and Wells, 1981b). The -35 and -10 regions are descriptively named with respect to their location from the transcription initiation site and have been compiled and characterized in detail in over 160 *E. coli* promoters (Hawley and McClure, 1983). Based on sequence homologies among 112 well-defined promoters, a very solid consensus promoter sequence could be established. This is only one of many studies (O'Neill, 1989; O'Neill and Chiafari, 1989) that qualify the *E. coli* promoter to be the most extensive and best studied single class of DNA regulatory sequence. Mutations that increase transcription initiation frequencies generally convert a nonconsensus base pair within a promoter to the consensus base pair for that

position. The reverse is true for those mutations where transcription initiation frequencies are decreased. The presence or absence of accessory factors such as cAMP-CAP or repressor molecules can also affect the efficiency of a promoter by several hundred fold (Reznikoff *et al.*, 1985). If promoter strength can be determined by how closely a sequence matches the consensus sequence, it should be possible to predict the strength of a promoter by its DNA sequence alone. Such an analysis was performed by Mulligan *et al.* (1984). They established homology scores which measured the match of promoters to the consensus sequence and obtained promoter strength values *in vitro*. A similar study was performed by Staden (1984) who took advantage of the established *E. coli* consensus promoter sequence to calculate weighted matrices that enabled identification of -35 and -10 regions without experimentally determining the transcription start site.

Finally, it is interesting to note that although over 160 *E. coli* promoters have been sequenced, no two promoter regions with the identical sequence have been observed, and most promoters do not contain the perfect consensus -10 and -35 sequences. In fact it is quite rare to find a perfect consensus sequence for any $\sigma$ subunit.

## Gene expression in mycobacteria

Most of the studies on cloning and expression of mycobacterial genes have been done in *E. coli.* Caution must be used when evaluating such studies however. This is because of the belief that organisms such as mycobacteria with a high guanine plus cytosine (G+C) content in their genomes must use a very different transcriptional control system from that in *E. coli.* This effect could no doubt be extended to translational signals as well, especially considering the potential for gene regulation by codon usage (Saier, 1995). This view is supported by the failure of the *M. tuberculosis* 85A antigen promoter to direct transcription of the alkaline phosphatase reporter gene in *E. coli* (Kremer, *et al.*, 1995). Other studies have yielded similar results (Das Gupta *et al.*, 1993). However, the inherent obstacles in expression of mycobacterial genes in *E. coli* is not absolute. Thole *et al.*, (1985) analyzed a gene library of *M. bovis* BCG in the lambda vector EMBL3 and found that expression of selected antigens was weak but detectable by Western blotting. Also, as discussed below, a mycobacterial heat shock promoter functions in *E. coli* (Thole *et al.*, 1987). This evidence suggests that at least some mycobacterial genes can be expressed in *E. coli,* but probably less efficiently than in their native host. This led to a search for an alternative cloning host that was more closely related to the mycobacteria. *Streptomyces* species emerged as a possible candidate since the DNA of these bacteria also has a high G+C

content. When a comparitive study between *E. coli* and *S. lividans* as cloning hosts was performed, it was concluded that *Streptomyces* is preferred over *E. coli* to clone mycobacterial expression signals (Kieser *et al.*, 1986). Only recently, has it become possible to study gene expression or promoter biology directly in mycobacteria. Host-vector systems have been developed that allow study of transcriptional activity in *M. smegmatis* and *M. bovis* BCG (Barletta *et al.*, 1992). The fast-growing *M. smegmatis* species has rapidly become the genetic "workhorse" for research in this genus with the development of efficiently transformable strains (Snapper *et al.*, 1990). As mentioned previously, *E. coli* promoters are probably the best characterized class of DNA regulatory sequence. Promoter biology in *Bacillus subtilis* has also been well characterized by the laboratories of Losick and Moran (Kenney and Moran, 1991; Kroos *et al.*, 1989; Rather *et al.*, 1986; Moran, 1989; Satola *et al.*, 1991;). Considerably less is known, however, about promoters of species in the genus *Mycobacterium.* Complex studies of gene expression such as global regulation, σ factor switching, and positive and negative regulation are all well established in *E. coli* and *B. subtilis* but are not yet possible in mycobacteria.

Despite their highly pathogenic nature and importance as heterologous expression systems (Barletta *et al.*, 1990; Dellagostin *et al.*, 1993; Stover *et al.*, 1991; Stover *et al.*, 1993), gene expression in mycobacteria is poorly understood because of the great difficulties encountered when working with these organisms. As recently as

1991, no intact mRNA had been successfully extracted from mycobacteria due to the short half-life of prokaryotic mRNA and the difficulty in achieving rapid lysis (Patel *et al.*, 1991; Bashyam and Tyagi, 1994). This inability to extract RNA from mycobacteria forced investigators to perform very limited characterization of transcriptionally active mycobacterial cloned fragments in a heterologous host. Those investigators that did perform more detailed studies such as transcription start site experiments, used *E. coli* RNA when analyzing mycobacterial promoters (Sela and Clark-Curtiss, 1991). Another contributing factor for the lack of knowledge regarding mycobacterial promoter biology is that the σ factor of RNA polymerase has not yet been identified or studied in this system. However some advancements in the field of mycobacterial genetics have been made, especially recently. Several enzymatic and mechanical methods for DNA extraction have been reported (Hill *et al.*, 1972; Patel *et al.*, 1986; Whipple *et al.*, 1987) and have led to the establishment of gene libraries for *Mycobacterium leprae* (Clark-Curtiss *et al.*, 1985) and *M. tuberculosis* (Young *et al.*, 1985; Cole and Smith, 1994). The subsequent development of suitable hosts and vector systems for working with mycobacteria (Jacobs *et al.*, 1991; Snapper *et al.*, 1990; Hinshelwood and Stoker, 1992; Qin *et al.*, 1994) meant that many of the difficulties encountered while working directly with pathogenic mycobacteria could be alleviated. Shuttle cosmids have been constructed that contain an origin of replication that functions in *E. coli*, an origin of replication that functions in

mycobacteria, a kanamycin-resistance gene that functions in both *E. coli* and mycobacteria, bacteriophage λ *cos* sequence that permits these molecules to be packaged into bacteriophage λ heads, and unique restriction sites for construction of genomic libraries (Jacobs *et al.*, 1991). A rudimentary transposon mutagenesis system has been developed (Martin *et al.*, 1990; McAdam *et al.*, 1994). Moreover, promoter-probe shuttle vectors have been engineered which can replicate between *E. coli* and *M. smegmatis* (Barletta *et al.*, 1992; Das Gupta *et al.*, 1993). Yet, despite these recent advancements, little is known about mycobacterial gene promoters and their activities. There has been some evidence suggesting that mycobacterial promoters are distantly related to promoters of other microorganisms (Sirakova *et al.*, 1989; Das Gupta *et al.*, 1993). A mycobacteriophage (L5) as been discovered and sequenced and may become an important tool in mycobacterial genetics (Lee and Hatfull, 1993; Hatfull, 1995). The most extensively studied mycobacterial promoter is that which controls expression of the heat shock protein genes (Thole *et al.*, 1987; Stover *et al.*, 1991). Those promoters controlling the expression of heat shock proteins are among the rare ones that have been shown to be active in *E. coli* (Thole *et al.*, 1987). Therefore, heat-shock promoters, which are functional in both mycobacteria and *E. coli*, should be useful in controlling expression of recombinant genes in *E. coli*-mycobacteria shuttle vectors. More recently, other mycobacterial promoters isolated from *M. paratuberculosis* (Thomas *et al.*, 1992; Murray *et al.*, 1992) and *M.*

*fortuitum* (Timm *et al.*, 1994) were found to contain well-conserved -10 and -35 regions. Caution must be used when evaluating these studies since these sequences were all active in *E. coli* and therefore may not be representative of true mycobacterial promoter structures.

## Recombinant *Mycobacterium bovis* BCG as a vaccine

An important application in the study of mycobacterial expression signals is to enable expression of foreign antigens in the attenuated *M. bovis* BCG (Bacille Calmette-Guerin) strain. Since mycobacteria are excellent adjuvents for humoral response as well as powerful mediators of cellular immunity (Aldovini and Young, 1991), the expression of foreign antigens in BCG would allow the construction of recombinant vaccine strains against human diseases of bacterial, viral, and parasitic etiology. *M. bovis* BCG was derived in 1906 from a virulent bovine strain by passage with ox bile for 231 passages (Calmette *et al.*, 1933). BCG is now the most widely used vaccine in the world against tuberculosis (Fine, 1988), although it is currently not used in the United States because of efficacy issues (Christensen, 1993; Bloom and Fine, 1994) and vaccinated persons would exhibit a positive reaction with the tuberculin skin test (Bloom and Fine, 1994). In other countries, the low frequency of serious complications and inexpensive production make BCG an ideal candidate for a recombinant vaccine vehicle. Since BCG, particularly

its cell wall, is a highly effective adjuvant, conjugation of an antigen to an adjuvant would be unnecessary when using recombinant BCG. In order to obtain a high level of foreign antigen expression in BCG, it is essential have strong mycobacterial expression signals. Currently there exists a multicopy plasmid and an integative expression vector, both of which carry the regulatory sequences for major BCG heat shock proteins to allow expression of foreign antigens (Stover *et al.*, 1991). However, even under the control of the same promoter, the amount of protein expressed in this system varies from 10% to less than 0.1% of the total protein extracted (Murray *et al.*, 1992). There is also variable expression of heterologous antigens among several *M. bovis* BCG strains (Burlein *et al.*, 1994). With the cloning of additional mycobacterial gene expression signals, it may be possible to find promoters that are stronger and drive expression at more consistent levels than the heat shock promoters currently in use. Three different human immunodeficiency virus type 1 (HIV-1) polypeptides have been expressed in BCG under the control of the mycobacterial hsp70 promoter (Aldovini and Young, 1991) and shown to elicit antibody and T cell responses. BCG also offers potential advantages as a vector for mucosal delivery of antigens. A single intranasal vaccination with recombinant BCG expressing the outer-surface protein A antigen from *Borrelia burgdorferi*, the etiological agent of lyme disease, results in a prolonged protective IgG response and a sustained secretory IgA response which is disseminated throughout the mucosal immune system (Stover *et al.*,

1993; Langermann *et al.*, 1994). The protective effects of recombinant BCG will undoubtedly be established for several pathogens in the future. It may even be further extended to allow simultaneous expression of multiple protective antigens of different pathogens, all in one vaccine.

# CHAPTER 2. IMPROVEMENT OF THE PROMOTER-PROBE VECTOR pKO1 FOR CLONING TRANSCRIPTION SIGNALS FROM *Mycobacterium paratuberculosis.*

A paper to be submitted to the Journal *Gene*

John P. Bannantine, C. O. Thoen, and R. E. Andrews, Jr.

## ABSTRACT

The promoter selection vector pKO1 has been modified to include the unique cloning sites *Bam*HI, *Sac*I, *Kpn*I, *Ava*I, *Xba*I, and *Pst*I. The new vectors (pJJ1 and pJJ2) allow the transcriptional fusion of genes to galactokinase (*galK*) in all three translational reading frames. Insertion of *M. paratuberculosis* DNA fragments upstream of *galK* into the unique *Bam*HI site of pJJ2 results in expression of *galK*. This newly constructed vector has been used to show that a previously sequenced *M. paratuberculosis* promoter-containing fragment (Thomas *et al.*, 1992) and 11 novel *M. paratuberculosis* promoter-containing fragments produced *galK* activity in *E. coli*. One of the 11 promoters (pTB28), which displayed high levels of *galK* activity on MacConkey's medium, was sequenced. Potential -35 and -10 promoter regions were identified from this sequence. S1 nuclease digestion of the 493 bp promoter-containing fragment from Thomas

*et al.* was used to determine the transcription start site in *E. coli.* The start point was 124 nucleotides upstream from the putative initiation codon and the -35 and -10 regions resembled the *E. coli* consensus promoter. The 493 bp fragment was also subcloned into the promoter-probe shuttle vector pYUB75 and designated pAJB60. When pAJB60 was transformed into both *E. coli* and *M. smegmatis* hosts, the *lacZ* reporter gene was expressed only in the mycobacterial host.

## INTRODUCTION

A number of plasmid vectors have been constructed for the purpose of cloning and the subsequent evaluation of DNA fragments with promoter activity (McKenney *et al.*, 1981; Bagdasarian *et al.*, 1983; Barletta *et al.*, 1992; Timm *et al.*, 1994). Such promoter-probe vectors rely on the in-phase fusion of the promoter-containing DNA to a structural gene or a selectable marker. The plasmid pKO1 is a commonly used promoter-probe vector for *E. coli* (McKenney *et al.*, 1981; Sirakova *et al.*, 1989). In this vector, the *E. coli* galactokinase gene (*galK*) provides the assayable marker to which transcriptional regulatory signals are fused. The promoterless *galK* reporter gene has been used successfully to detect mycobacterial promoter elements in a heterologous environment (Sirakova *et al.*, 1989; Thomas *et al.*, 1992). The limitations of pKO1 as a promoter selection vector include the lack of unique restriction sites upstream of *galK*

for the cloning of transcriptionally active DNA fragments and the distance of the cloning sites relative to the start of the *galK* reporter gene. The location of the restriction sites in pKO1 is not immediately upstream of *galK* which forces a cloned promoter to transcribe additional unnecessary plasmid sequences prior to reaching the reporter gene. For these reasons, pKO1 was also modified by removing the unnecessary vector sequences without disrupting or deleting the translational stops upstream of *galK*.

The genus *Mycobacterium* includes major pathogens such as *Mycobacterium tuberculosis* and *M. leprae*, the etiological agents of tuberculosis and leprosy. *M. avium* and *M. paratuberculosis* are closely related intracellular pathogens affecting humans and animals. The avirulent strain of *M. bovis* that has been extensively used as a tuberculosis vaccine, is also an attractive vector for the construction of live recombinant vaccines because of its strong immunogenicity. As a result, there is currently much interest in studying the molecular biology of mycobacteria.

Gene expression in mycobacteria, particularly *Mycobacterium paratuberculosis* is ill-defined. This can be attributed to a variety of factors including their slow growth rates, lack of genetic exchange methods such as transduction or conjugation, and inability to construct, detect and assay for mutations. The cloning and analysis of mycobacterial gene expression signals may provide clues to its pathogenic mechanisms as well as improve upon a system to drive expression of foreign antigens in *M. bovis* BCG.

In the present study, we describe the construction of pJJ1 and pJJ2 promoter selection vectors, which have the advantage of convenient multiple cloning sites as well as allowing transcriptional fusions to *galK*. The utility of the vectors for producing transcriptional fusions was shown by cloning and assessing gene regulatory sequences from both a previously characterized *M. paratuberculosis* promoter (Thomas *et al.*, 1992) and an *M. paratuberculosis* genomic DNA library.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** *M. paratuberculosis* ATCC 19698 was provided by D. Whipple NADC, Ames, IA and served as the source from which genomic DNA was isolated. *E. coli* HB101 (*galK, recA*), *E. coli* RR1 (*gal*K), and *E. coli* TG-1 ($F^+$) were used for cloning *M. paratuberculosis* promoters and production of single-stranded DNA, respectively. The promoter-probe plasmid pKO1 (Thomas *et al.*, 1992) was a laboratory stock. Helper phage R408 for production of single-stranded DNA templates was obtained from F. C. Minion, Department of Microbiology, Immunology, and Preventive Medicine (MIPM), Iowa State University, Ames, IA. Plasmid pKRB8 was provided by G. Phillips, Department of MIPM. The shuttle vectors pYUB75 and pYUB76 (Barletta *et al.*, 1992) and *M. smegmatis* $mc^2155$ were provided by R. Barletta, Department of Veterinary Science, University of Nebraska, Lincoln, NE. Plasmids pYUB75 and

pYUB76 have a kanamycin resistance marker for selection, contain a promoterless *lacZ* gene, and replicates in both *E. coli* and *M. smegmatis*.

**Growth conditions and extraction of *M. paratuberculosis* DNA.** For genomic DNA isolation, *M. paratuberculosis* was incubated without shaking in Middlebrook 7H9 broth (Difco, Detriot, MI) containing 0.5% Tween-80, OADC enrichment (Difco) and 2 mg/l mycobactin J (Allied Monitor, Fayette, MO). The cells were harvested after 5 weeks of growth in tissue culture flasks at 37°C. *M. paratuberculosis* pellets (1 to 3 g) were frozen at -70°C and genomic DNA was extracted by the method of Whipple *et al.* (1987).

**Construction of Plasmid pJJ1 and pJJ2.** To create pJJ1 and pJJ2, the polylinker along with the spectinomycin/streptomycin resistance cassette was excised from pKRB8 by digestion with *Eco*RI and cloned into the *Eco*RI site of pKO1 (Figure 1). The resulting construct, designated pKO8, was then digested with *Hin*dIII and allowed to self-ligate to remove half of the symmetrical polylinker, pKO1 sequences 5' to the *galK* gene, and the spectinomycin/ streptomycin cassette. This new construct was designated pJJ2. The pJJ1 construct was created by digestion of pKO8 with *Bam*HI followed by vector religation. Only the spectinomycin/ streptomycin cassette

and a portion of the multiple cloning site were removed by this treatment.

**Cloning of *M. paratuberculosis* promoters in pJJ2.** Genomic DNA from *M. paratuberculosis* was partially digested with *Sau*3AI and fractionated by agarose gel electrophoresis. Fragments of this digestion in the 100 to 1000 bp range were purified by electroelution from an agarose gel. A ligation reaction was prepared by mixing approximately 0.5 μg of purified genomic DNA fragments with 1 μg of dephosphorylated, *Bam*HI-cleaved pJJ2 in a final volume of 10 μl. Half of the ligation mixture was used to transform *E. coli* RR1 using a highly efficient $CaCl_2$ method (Miller, 1992). The transformed cells were plated on MacConkey's containing 1% galactose and Ampicillin ($Ap^r$,100 μg/ml).

The 493 bp *Eco*RI promoter-containing fragment from pAG5 (Thomas *et al.*, 1992) was gel purified and cloned into the *Eco*RI site of p34E (Tsang *et al.*, 1991). The fragment was removed from p34E by digestion with *Bam*HI and cloned into the *Bam*HI site of pYUB75 (Barletta *et al.*, 1992) where *lacZ* serves as the reporter gene. The ligation mixture was transformed by electroporation into *E. coli* DH10B and plated on Luria-Bertani (LB) medium containing 50 μg/ml kanamycin. The LB medium also contained 5-bromo-4-chloro-3-indolyl-β-D galactopyranoside (Xgal) added to each plate as described by Sambrook *et al.*, (1989). Kanamycin resistant colonies obtained from this experiment were screened by plasmid extraction

(Qiagen, Inc., Chatsworth, CA) and restriction analysis for confirmation of the construct which was given the designation pAJB60. This construct was then electroporated into *M. smegmatis* $mc^2155$ to assess β-galactosidase expression in that background. The selective media for *M. smegmatis* electroporations was Middlebrook 7H9 medium (Difco, Detriot, MI) with 50 μg/ml kanamycin and Xgal added as described above.

**Estimation of promoter strength by β-galactosidase assays.** The β-galactosidase assays for *E. coli* were performed as described by Miller (1972) with one modification. The minimal medium for growth recommended by Miller was not employed due to poor growth of *E. coli* DH10B. Instead, LB medium was substitued for minimal medium. β-galactosidase assays for *M. smegmatis* were performed as described by Barletta *et al.*, (1992). β-galactosidase activities were reported in Miller units (Miller, 1972).

**RNA extraction.** Total cell RNA was isolated from *E. coli* by routine methods (Ausubel *et al.*, 1989).

**DNA sequencing.** Sequencing reactions were performed manually with Sequenase® Version 2.0 sequencing kit (United States Biochemical Corporation, Cleveland, OH) for generating molecular weight markers with phage M13 as the DNA template to resolve the size of the S1 mapped product.

The *galK* positive clone with the smallest insert, pTB28, was subcloned into pGEM3Zf(+) (Promega, Madison, WI) for DNA sequencing. The promoter-containing insert was isolated from pTB28 by a *Hin*dIII-*Eco*RI double digestion. The vector pGEM3Zf(+) was digested in a similar manner and included in a ligation reaction that also contained the pTB28 fragment. The resulting construct was designated pSEQ4.

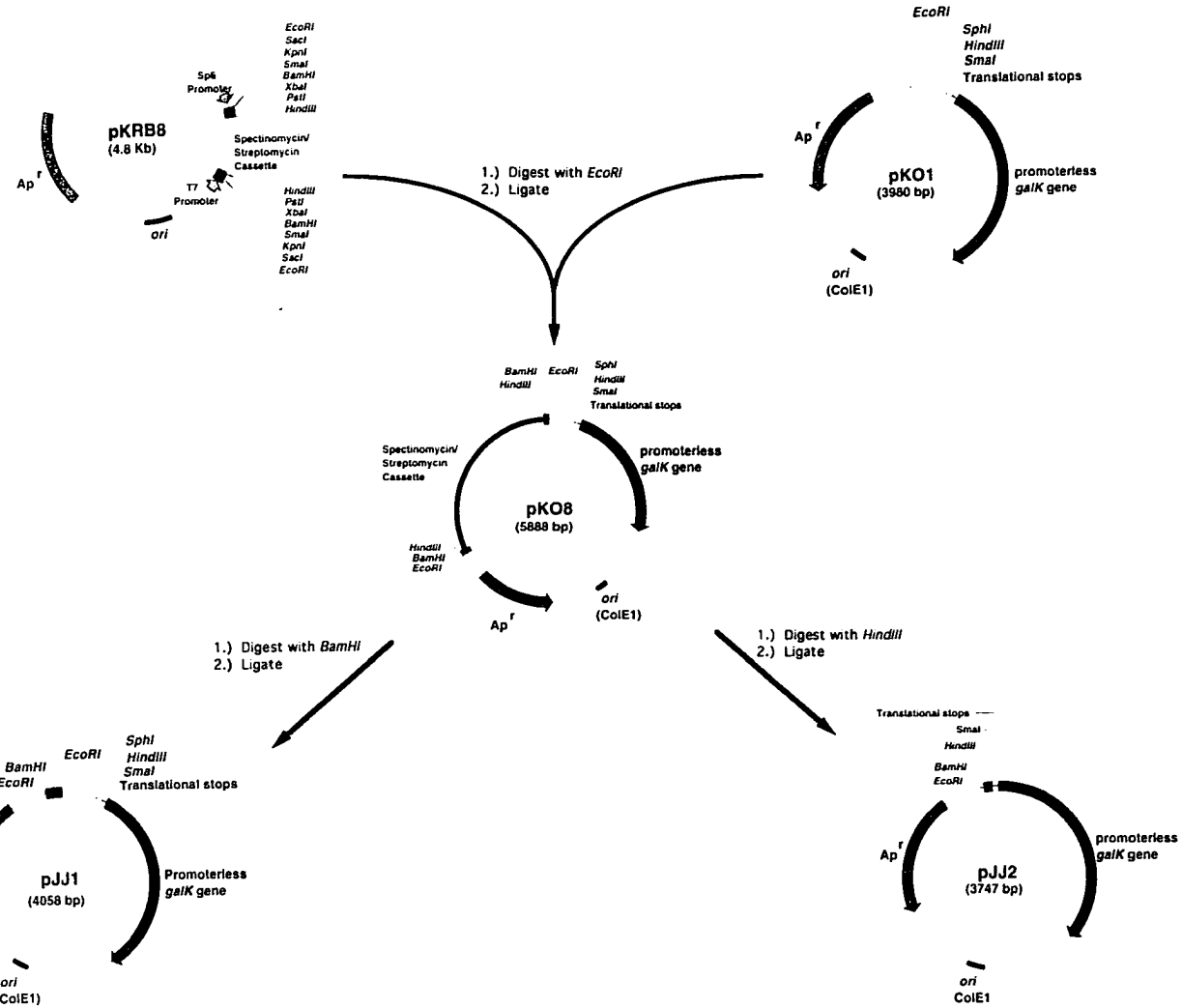
The pSEQ4 insert was sequenced by using the *Taq* DyeDeoxy® Terminator Cycle Sequencing Kit (Perkin Elmer/Applied Biosystems Division, Foster City, CA). The sequencing primers included the universal forward and reverse M13 primers that were complementary to regions flanking the multiple cloning sites on pGEM3Zf(+). Each sequencing reaction included 100 ng of template DNA and 10 pmol of sequencing primer in a total volume of 20 µl. Each reaction underwent 25 cycles (98°C for 15 s, 50°C for 10 s, 60°C for 4 min) in the GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer, Norwalk, Conn.). The sequencing products were purified through Centri-Sep columns (Princeton Separations, Adelphia, N.J.) and were loaded onto 6% denaturing polyacrylamide gels in an automated sequencer (373A DNA Sequencer; Perkin elmer/Applied Biosystems Division). Template DNA for automated sequencing was prepared by one of two methods; PEG preparation (Tartof and Hobbs, 1987) or Qiagen extraction. The Qiagen miniprep was performed according to the instructions of the manufacturer (Qiagen).

**S1-nuclease mapping.** The 493 bp *Eco*RI fragment from pAG5 was isolated from a 0.8% low melting agarose gel. The 5' hydroxyl end was radiolabeled with [$\gamma$-$^{32}$P]ATP (3000 mCi mmole$^{-1}$, Amersham) using polynucleotide kinase (10U). Unincorporated radioactivity was removed by two consecutive ethanol precipitations. *E. coli* RNA (30 µg) and radiolabeled DNA probe (0.5 µg) were mixed and precipitated with ethanol and 3M sodium acetate. The pellet was resuspended in 45 µl of hybridization buffer (40mM piperazine-N-N'-bis-ethanesulphonic acid (PIPES), 1mM EDTA (pH 8.0), 0.4M NaCl, 80% formamide) and incubated at 62°C for 12 h. DNA/RNA hybrids were subjected to S1-nuclease (300U) digestion at 37°C for 1-2h. The reactions were terminated by the addition of 80 µl of S1-nuclease stop buffer (4M ammonium acetate, 50mM EDTA (pH 8.0), 50 µg/ml tRNA). Reactions were heat denatured at 90°C for 2 min. then separated on a 6% denaturing polyacrylamide gel (Sambrook *et al.*, 1989).

## RESULTS

**Construction of pJJ1 and pJJ2.** pJJ1 and pJJ2 are transcriptional fusion vectors where promoter fragments are cloned upstream of the *galK* ribsomal binding site (RBS) and translational stop codons in all three reading frames. Figure 1 shows the general cloning strategy used to construct these vectors and Figure 2 shows the deduced nucleotide sequence comparison of pKO1, pJJ1, and pJJ2

Figure 1. Construction of plasmids pJJ1 and pJJ2. The 2.8-kb *Eco* RI fragment of pKRB8 was cloned into the *Eco*RI site of pKO1. The resulting plasmid, pKO8, was digested with either *Bam*HI (pJJ1) or *Hin*d III (pJJ2) and self-ligated to remove the spectinomycin/streptomycin resistance cassette.


pKRB8
(4.8 Kb)
Ap r
ori
T7 Promoter
Sp6 Promoter
Spectinomycin/ Streptomycin Cassette
EcoRI SacI KpnI SmaI BamHI XbaI PstI HindIII
HindIII PstI XbaI BamHI SmaI KpnI SacI EcoRI
1.) Digest with EcoRI
2.) Ligate
EcoRI
SphI
HindIII
SmaI
Translational stops
pKO1
(3980 bp)
promoterless galK gene
ori (ColE1)
BamHI EcoRI HindIII
SphI HindIII SmaI Translational stops
Spectinomycin/ Streptomycin Cassette
promoterless galK gene
pKO8
(5888 bp)
HindIII BamHI EcoRI
ori (ColE1)
1.) Digest with BamHI
2.) Ligate
1.) Digest with HindIII
2.) Ligate
BamHI EcoRI
EcoRI
SphI HindIII SmaI Translational stops
pJJ1
(4058 bp)
Promoterless galK gene
ori (ColE1)
Translational stops
SmaI
HindIII
BamHI EcoRI
pJJ2
(3747 bp)
promoterless galK gene
ori ColE1

Plasmid pKO1:

***Eco*RI**
GAATTCTGGC GAATCCTCTG ACCAGCCAGA AAACGACCTT TCTGTGGTGA AACCGGATGC TGCAATTCAG AGCGCCAGCA

AGTGGGGGAC AGCAGAAGGA CCTGACCGCC GCAGAGTGGA TGTTTGACAT GGTGAAGACT ATCGCACCAT CAGCCAGAAA

***Eco*RV**
ACCGAATTTT GCTGGGTGGG CTAACGATAT CCGCCTGATG CGTGAACGTG ACGGACGTAA CCACCGCGAC ATGTGTGTGC

***Sph*I** ***Hin*dIII** ***Sma*I**
TGTTCCGCTG GGCATGCCAG GACAACTTCT GGTCCGGTAA CGTGCTGAGC CCGGCCAAGC TTACTCCCCA TCCCCGGGCA

ATAAGGGCTG CACGCGCACT TTTATCCGCC TCTGCTGCGC TCCGCCACCG TACGTAAATT TATGGTTGGT TATGAAATGC

TGGCAGAGAC CCAGCGAGAC CTGACCGCAG AACAGGCAGC AGAGCGTTTG CGCGCAGTCA GCGATATCCA TTTTCGCGAA

TCCGGAGTGT AAGAAATG <-- *galK* start

Plasmid pJJ1:

***Eco*RI** ***Sac*I** ***Kpn*I** ***Bam*HI** ***Kpn*I** ***Sac*I** ***Eco*RI**
GAATTCGAGC TCGGTACCCG GGGATCCCCG GGTACCGAGC TCGAATTCTG GCGAATCCTC TGACCAGCCA GAAAACGACC

TTTCTGTGGT GAAACCGGAT GCTGCAATTC AGAGCGCCAG CAAGTGGGGG ACAGCAGAAG GACCTGACCG CCGCAGAGTG

***Eco*RV**
GATGTTTGAC ATGGTGAAGA CTATCGCACC ATCAGCCAGA AAACCGAATT TTGCTGGGTG GGCTAACGAT ATCCGCCTGA

***Sph*I**
TGCGTGAACG TGACGGACGT AACCACCGCG ACATGTGTGT GCTGTTCCGC TGGGCATGCC AGGACAACTT CTGGTCCGGT

***Hin*dIII** ***Sma*I**
AACGTGCTGA GCCCGGCCAAGC TTACTCCCCA TCCCCGGGCA ATAAGGGCTG CACGCGCACT TTTATCCGCC TCTGCTGCGC

TCCGCCACCG TACGTAAATT TATGGTTGGT TATGAAATGC TGGCAGAGAC CCAGCGAGAC CTGACCGCAG AACAGGCAGC

AGAGCGTTTG CGCGCAGTCA GCGATATCCA TTTTCGCGAA TCCGGAGTGT AAGAAATG <-- *galK* start

Plasmid pJJ2:

*Eco*RI *Sac*I *Kpn*I *Bam*HI *Xba*I *Pst*I *Sph*I *Hin*dIII *Sma*I
GAATTCGAGC TCGGTACCCG GGGATCCTCT AGAGTCGACC TGCAGGCATG CAAGCTTACT CCCCATCCCC GGGCAATAAG

GGCTGCACGC GCACTTTTAT CCGCCTCTGC TGCGCTCCGC CACCGTACGT AAATTTATGG TTGGTTATGA

AATGCTGGCA GAGACCCAGC GAGACCTGAC CGCAGAACAG GCAGCAGAGC GTTTGCGCGC AGTCAGCGAT

ATCCATTTTC GCGAATCCGG AGTGTAAGAA ATG <- *gal K* start

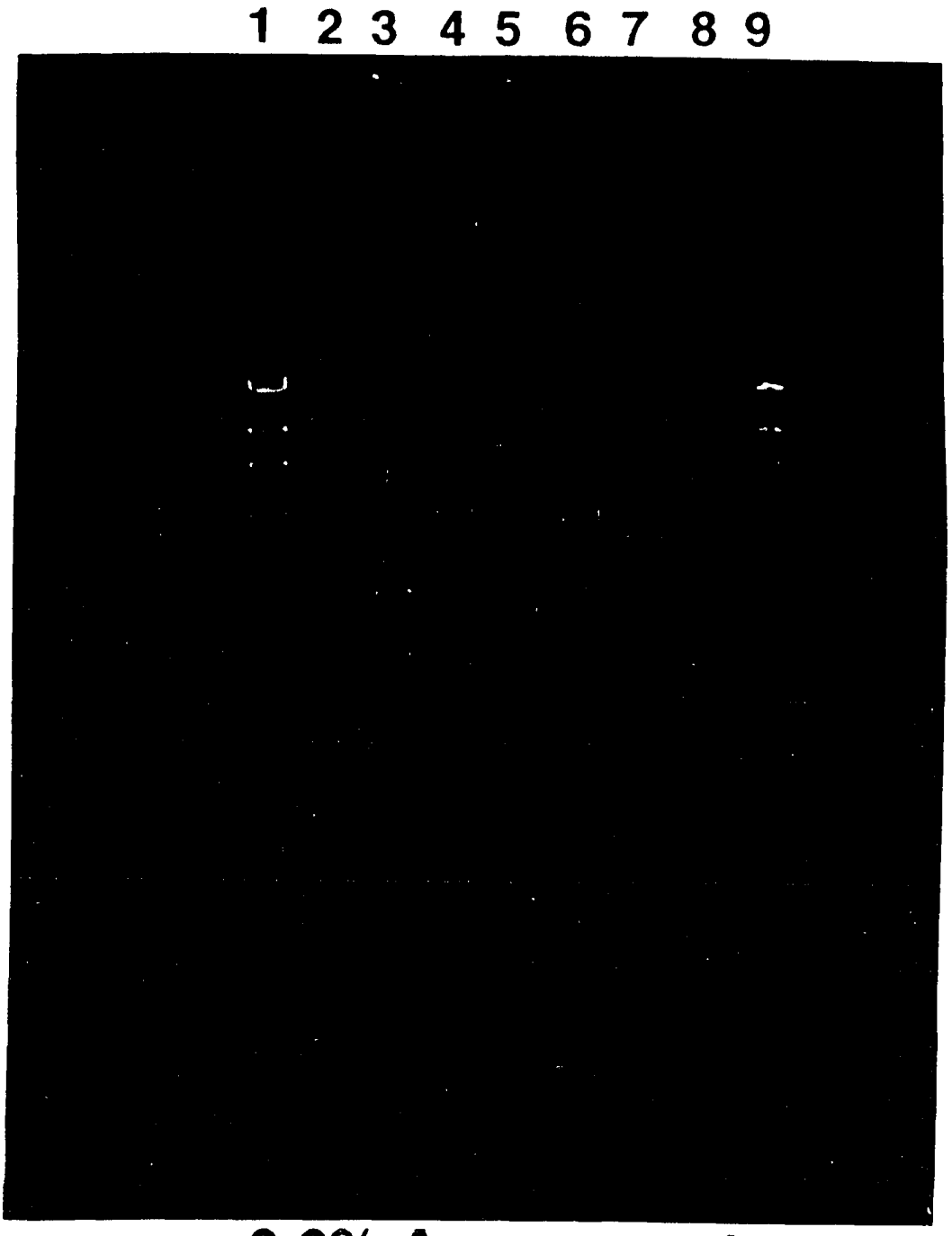
Figure 2. Deduced nucleotide sequence comparison of the regions upstream of *galK* in pKO1, pJJ1, and pJJ2. Restriction sites are indicated by an underlined recognition sequence with the restriction enzyme designation above it. The *galK* translational start is indicated and the upstream translational stops are shown in bold.

immediately 5' to the *galK* gene. The translational stops from the pKO1 parent were preserved in pJJ1 and pJJ2 (Figure 2). Figure 3 shows the restriction patterns of the described plasmids that verifies their construction. The plasmid-encoded antibiotic ampicillin serves as the selectable marker, whereas the red-white colony color on MacConkey's allows for quick identification of promoter-containing inserts based on acid byproducts from galactose utilization. *E. coli* cells harboring pJJ1 or pJJ2 showed no detectable *galK* activity on MacConkey medium.

**Cloning of the pAG5 insert and *M. paratuberculosis* *Sau*3AI fragments into pJJ2.** In order to clone mycobacterial promoters, ligations and transformations were carried out as mentioned in Materials and Methods. The ligation reaction yielded a 1,553 clone library of which 11 were *galK* positive. The frequency of promoter isolation was expressed as a percent with the number of red colonies (promoter-containing clones) divided by the number of white colonies. Thus 0.7% of this non-representative library contains putative promoter clones. Three *galK* positive clones (pTB12, pTB22 and pTB28) contained strong promoters as determined by intensity of red color and were found to contain insert sizes of 1400, 720, and 600 bp respectively. The smallest clone (pTB28) was subcloned into pGEM3Zf(+) and sequenced.

Previously, a *M. paratuberculosis* DNA library was cloned into pKO1 (Thomas *et al.*, 1992). One galactokinase positive clone from

Figure 3. Restriction analysis and agarose gel electrophoresis of pJJ1, pJJ2, and intermediate constructs pKO1, pKO8, and pKRB8. Shown is a 0.8% agarose gel containing the following lane assignments: lane 1, *Eco*RI-digested pKRB8; lane 2, *Hin*dIII-digested pKRB8; lane 3, *Bam*HI-digested pKO8; lane 4, *Bam*HI-digested pJJ1; lane 5, *Hin*dIII-digested pKO8; lane 9, *Bam*HI-digested pJJ2.




0.8% Agarose gel

this library (pAG5) contains a 493 bp *Eco*RI fragment which was subcloned into pJJ2 to give pTB10 and used to test the efficacy of the newly constructed vector. Both pTB10 and pAG5 display similar levels of *galK* activity as determined by intensity of red color on MacConkey's plates (data not shown). These data show that modifications of pKO1 to construct pJJ2 did not alter the *galK* reporter activities.

**Subcloning of the pAG5 insert into pYUB75 and expression of *lacZ* in *M. smegmatis*.** It has been shown previously that the 493 bp *M. paratuberculosis* insert of clone pAG5 acts as a promoter element in *E. coli* (Thomas *et al.*, 1992). To establish that the 493 bp DNA fragment contains legitimate mycobacterial promoter sequences, the fragment must drive expression of a reporter gene in a mycobacterial host. To accomplish this objective, the insert of pAG5 was subcloned into pYUB75, a promoter selection vector that replicates in *M. smegmatis*. This subclone, designated pAJB60, was transformed by electroporation into both the *E. coli* and *M. smegmatis* backgrounds where expression of the *lacZ* reporter gene was assessed quantitatively (Table 1). *E. coli* (pUC19) was used as a positive control with *M. smegmatis* and *E. coli* harboring pYUB75 serving as negative controls. Interestingly, a *lacZ* fusion was not produced in *E. coli*, but was in *M. smegmatis* since β-galactosidase was shown to be expressed in *M. smegmatis*, but was not detectable in *E. coli* (Table 1). Nucleotide sequencing of pAJB60

(Bannantine *et al.*, in preparation) has revealed that the 493 bp *M. paratuberculosis* fragment was cloned in the opposite orientation relative to pTB10 or pAG5 (Thomas *et al.*, 1992).

Table 1. Measurement of promoter activity from the 493 bp *M. paratuberculosis* DNA fragment.

| Strain | Colony color[a] | β-galactosidase activity[b] |
|---|---|---|
| *E. coli* DH10B (pUC19) | Blue | 91±8.4 |
| *E. coli* DH10B (pYUB75) | White | 0.35±0.071 |
| *M. smegmatis* (pYUB75) | White | 0.26±0.058 |
| *E. coli* DH10B (pAJB60) | White | 0.35±0.083 |
| *M. smegmatis* (pAJB60) | Blue | 1.9±0.29 |

[a] Colony color was scored on Xgal containing medium.
[b] Reported figures are in Miller Units (Miller, 1972) and are the average from three independent trials ± standard dev.

**Nucleotide Sequencing of the pSEQ4 insert.** The pSEQ4 insert is 518 bp long and has a G+C content of 64.1%, which is slightly lower than the published figure of 66-67% for *M. paratuberculosis* (McFadden *et al.*, 1987). To locate possible promoter sites, the sequence was examined for mycobacterial consensus regions that have been recently established (Bannantine *et al.*, in preparation). The 518 bp sequence is shown in Figure 4 along with the potential

GATCTCCAGCGGCGCGGTGACCACCAGACAGGCGGCGAGCACCACCAGGTATTGCCACCGAT

CCGCCAAGAGCGCAGTACCCCCACGCTCACGGGCGGCGTTGAAATAGGTAGTAATACGTCGG

GGTCTTGTCGTCGAGCGCCCGCTCCTGGCCAGTGGACAACAGCGTCCAACCCGCGGTGAACCG

CCGTTCGATTTCGGCGCGATCGGCATCGACGCCACACCCATCGAGCGGGCGTCCAACACGCTG

ATCATCATCTTCGGCTCGAAGCGGACCTTTACAGCCTCAGTCATATCTGTCACTCTCTTCTGGG

TCGCGGAGGGGCTACGGAGAGCCTTACCGGAAACTAGCCGATTGGCAAGCTGCGGCCGCCGC

CGGGCTGGTTA**TTGCGG**GTCGAACAGCACCGGGA**CCGACG**TCGGCGACCGGAACACCTGGC
-35 -10
***
CCCGGATGTGGGGATC**GAGGAAG**TCATCGCGGTGAACCCCAAGGTGTTCTTCCAGGTGTAC
RBS

TGGCTGGGTGGGCGCGACGC

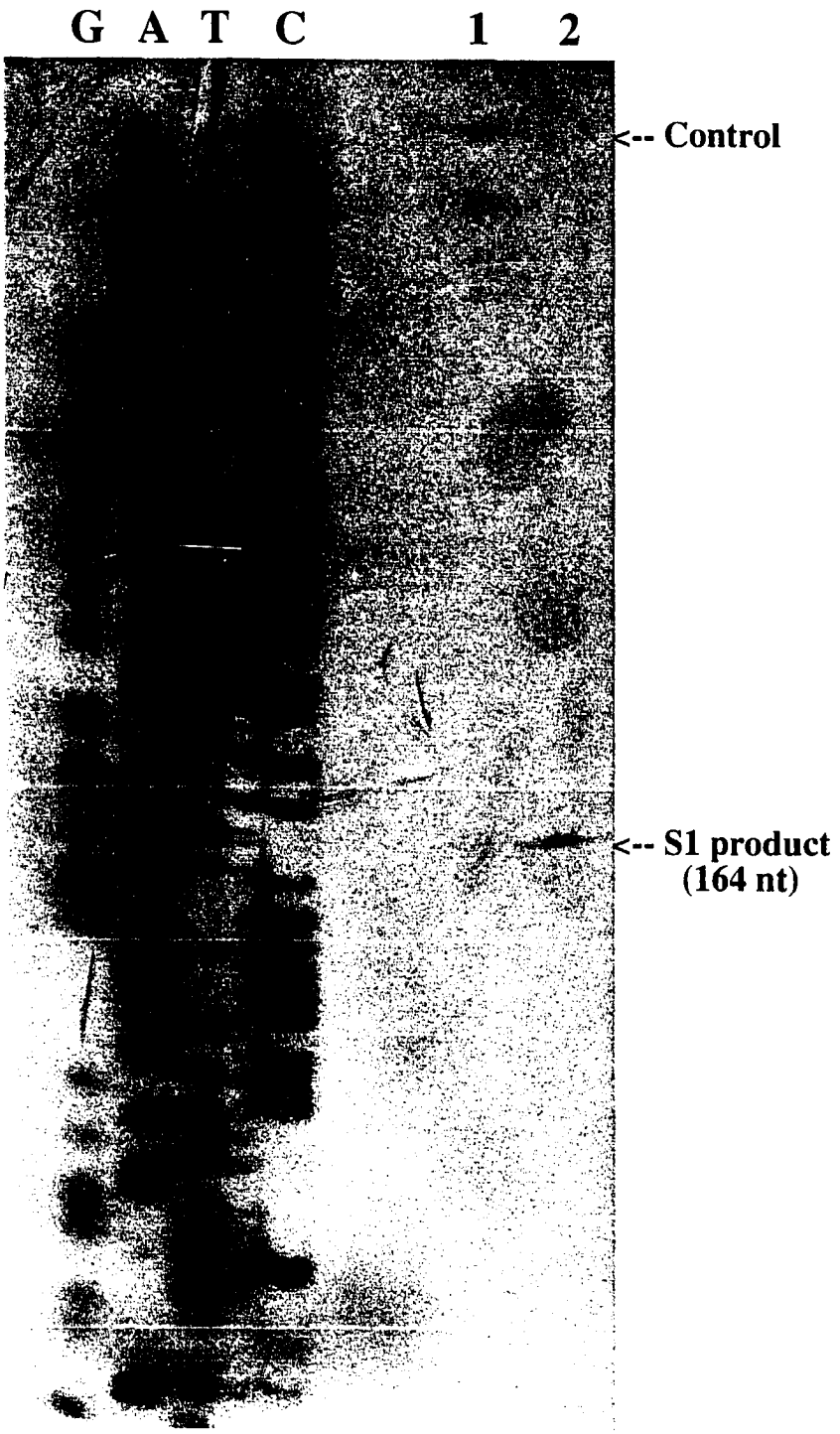
Figure 4. Nucleotide sequence of the 518 bp insert of pSEQ4. Only mycobacterial DNA is shown, no vector sequences. The sequence is written 5' to 3' and in pTB28, the *galK* gene would be on the 3' end. The potential transcriptional signals (-35 and -10 regions) as identified by matching the conserved promoter regions established by Bannantine *et al.* (in preparation) and the ribosome binding site (RBS) are highlighted in bold letters. The putative start codon is underlined. The three "*" above the TGA codon indicate the stop codon present in frame 3.

-35 and -10 regions that are highlighted in bold letters and closely resemble conserved promoter regions in mycobacteria. Nine of twelve base pairs in the -35 and -10 regions are consensus nucleotides. These identified conserved regions combined with a purine-rich ribosome binding site and well-positioned initiation

codon provide strong evidence for the proper location of the promoter region.

**S1 nuclease mapping.** To determine the 5' end of the transcript from pTB10 in *E. coli,* 20 μg of total cell RNA from *E. coli* was hybridized to 0.5 μg of the denatured, [$\gamma$-$^{32}$P]ATP labeled 493-bp fragment. Hybridization conditions were those favoring the formation of DNA-RNA hybrids (Sambrook *et al.*, 1989). The length of protected single-stranded DNA fragments was determined by comparison with M13 DNA sequencing ladders (Figure 5). The S1 mapped product of 164 bp nucleotides in length, indicating a 5' mRNA start site 15 nucleotides downstream from the -10 region reported by Thomas *et al.* 1992 (Figure 6).

Figure 5. S1 nuclease mapping of the 493 bp pAG5 fragment in *E. coli* HB101. The phage M13 nucleotide sequence is shown in the lanes labeled G, A, T, and C. Lane 1, the DNA:RNA hybrid which was not subjected to S1 nuclease digestion. Lane 2 contains the S1 nuclease digested product which has a length of 164 nucleotides (nt).


G A T C
1 2
<-- Control
<-- S1 product
(164 nt)

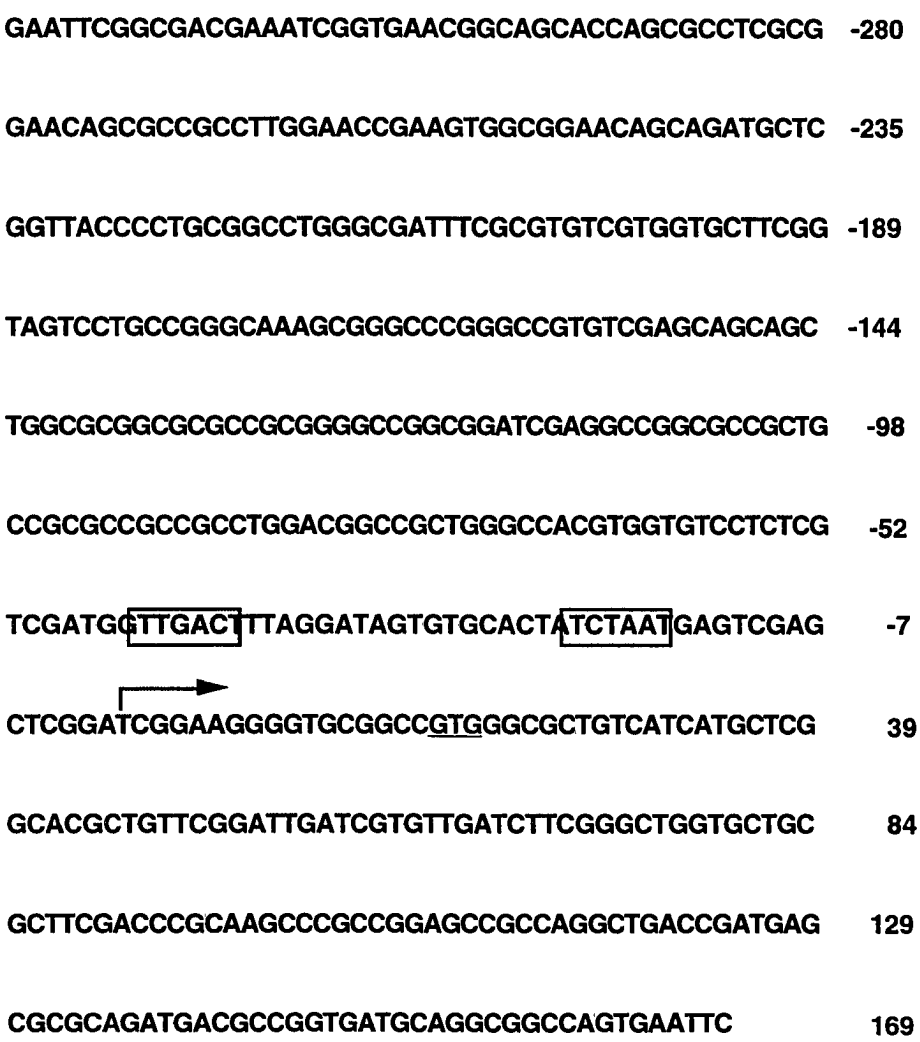



Figure 6. Nucleotide sequence of the 493 bp fragment with salient features. Shown is the nucleotide sequence of the pAG5 insert as determined by Thomas *et al.* (1992) and confirmed in this study. The putative -35 and -10 regions are shown in shaded boxes and the transcription start site, determined by S1 nuclease mapping is indicated by an arrow over the nucleotide. The potential translational start codon is underlined.

# DISCUSSION

The utility of pKO1 as a promoter-probe selection vector is limited by a lack of unique restriction sites 5' to the *galK* reporter gene. *M. paratuberculosis* genomic DNA is cleaved most efficiently with *Sau*3AI for generating restriction fragments in the 0.2 to 1 Kb range which are large enough to contain a single promoter region but small enough to be easily sequenced. However, *Sau*3AI generates cohesive ends which are not compatible for ligation to the restriction sites that exist in pKO1. To overcome this limitation, several unique restrictions sites were added to construct the pKO1 derivatives pJJ1 and pJJ2. The vector pJJ2 was tested for its ability to recognize transcriptionally active *M. paratuberculosis* fragments. Of 11 *galK*-positive clones, pTB28 was selected for subcloning and sequencing because of its small insert size (518 bp) and strong *galK* activity. The sequenced fragment was found to contain potential mycobacterial consensus promoter regions. Detailed analysis of mRNA will be necessary in order to determine whether the consensus sequences indicated here do in fact correspond to the transcriptional start site. With regard to cloning mycobacterial promoters, a limitation of pKO1 is that it cannot replicate in any mycobacterial host and ampicillin is not a selectable marker in mycobacteria (Kaneda and Yabu, 1983).

In this investigation, a sequenced *M. paratuberculosis* promoter element (Thomas, *et al.*, 1992) of 493 bp was subcloned into pJJ2 and the *Escherichia-Mycobacterium* shuttle vector (pYUB75) and

phenotypically shown to express the reporter gene in both vector systems. This same sequence drove expression of galaktokinase in *E. coli,* but not *lacZ* in *E. coli.* This is because the insert is in the opposite orientation in pAJB60 relative to pTB10. Therefore *E. coli* recognizes the *M. paratuberculosis* DNA sequence as a promoter in one orientation and *M. smegmatis* recognizes it in the opposite orientation. S1 nuclease mapping of the *galK* transcript produced in *E. coli* identified the location of the 5' end to be 15 bp downstream of the Pribnow box. This transcription start site agrees well with that predicted from Thomas *et al.* (1992) and resembles an *E. coli* consensus promoter more than a mycobacterial consensus promoter. From these data, it was reasoned that the 493 bp fragment contained a fortuitous promoter region that *E. coli* recognizes but is not biologically significant in mycobacteria. Moreover, when the fragment is cloned in the reverse orientation, it drives expression of a reporter gene in *M. smegmatis* and is therefore considered to be a legitimate mycobacterial promoter-containing fragment. Northern hybridization analysis of the *galK* transcript in *E. coli* was performed (data not shown) to show transcriptional activity in that system, however the identified transcript was larger than the expected size of 1.9 kb for the *galK* fusion transcript. This may be explained by a run on transcript where RNA polymerase transcribed through the terminator sequence at the end of *galK.* Transcriptional activity of pAJB60 in *M. smegmatis* was observed and quantitated by slot blot hybridization (Bannantine *et al.*, in preparation).

The quantitative β-galactosidase data shows that the pAG5 fragment does contain promoter activity in a mycobacterial host. One limitation of the β-galactosidase assay for mycobacteria, as stated by Barletta *et al.* (1992), is that toluene permeabilization may not work as efficiently as it does for *E. coli.* Therefore the data presented should be viewed as a relative comparison of activity and not as the actual number of β-galactosidase units. In light of this information, it is probable that *M. smegmatis* cells harboring the pAJB60::*lacZ* fusion produces more β-galactosidase activity than 1.9 units.

With the prospect of successful cloning of *M. paratuberculosis* promoters in different hosts, it may eventually become possible to determine which genes are expressed within the host. This could be accomplished by an analysis of *M. paratuberculosis* promoters that are turned on when grown *in vivo* and turned off when grown *in vitro.* Subsequent studies could be targeted to these promoters or the genes they express, and along with it, the organism's ability to persist within the host.

The vector pJJ2 is a more versatile promoter-probe vector than pKO1 since it contains additional unique restriction sites. It has the added advantage that unnecessary pKO1 plasmid sequences between the cloning sites and the *galK* reporter gene are removed. Furthermore, cloning of *Sau*3AI fragments into the *Bam*HI site, a procedure known to result in imperfect *Bam*HI sites, can still result in easy removal of the insert from the vector using flanking

restrictions sites. The evidence above casts some doubt on the utility of pJJ1 or pJJ2 for cloning mycobacterial transcription signals since *M. paratuberculosis* sequences may be fortuitously recognized as promoters in *E. coli.* However, *M. smegmatis* cells harboring pAJB60 are *lacZ*-positive which is indicative of a legitimate mycobacterial promoter sequence.

The results herein show that *M. paratuberculosis* sequences can serve as promoters in *E. coli* and *M. smegmatis.* While expression of these sequences in a mycobacterial host gives compelling evidence, whether these sequences are in fact legitimate promoters of *M. paratuberculosis* is the subject of future studies.

## ACKNOWLEDGEMENTS

We thank the DNA sequencing and synthesis facility (1184 Molecular Biology building, Iowa State University) for sequencing of the pSEQ4 insert.

## REFERENCES

**Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman,J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. John Wiley & Sons, New York.

**Bagdasarian, M. M., E. Amann, R. Lurz, B. Ruckert, and M. Bagdasarian.** 1983. Activity of the hybrid *trp-lac* (*tac*) promoter of *Escherichia coli* in *Pseudomonas putida.* Construction of broad-host-range, controlled-expression vectors. Gene **26**:273-282.

**Barletta, R. G., D. D. Kim, S. B. Snapper, B. R. Bloom, and W. R. Jacobs, Jr.** 1992. Identification of expression signals of the mycobacteriophages Bxb1, L1 and TM4 using the *Escherichia-Mycobacterium* shuttle plasmids pYUB75 and pYUB76 designed to create translational fusions to the *lacZ* gene. J. Gen. Microbiol. **138:**23-30.

**Kaneda, S., and K. Yabu.** 1983. Purification and some properties of β–lactamases from *Mycobacterium smegmatis.* Microbiol. Immunol. **27:**191-193.

**McFadden, J. J., P. D. Butcher, R. J. Chiodini, and J. Hermon-Taylor.** 1987. Determination of genome size and DNA homology between unclassified *Mycobacterium* species isolated from patients with Crohn's disease and other mycobacteria. J. Gen. Microbiol. **133:**211-214.

**McKenney, K., H. Shimatake, D. Court, U. Schmeissner, C. Brady, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase. *In* J. G. Chirikjian and T. S. Papas (Eds.), Gene Amplification and Analysis, vol. 2. Elsevier, North Holland, pp. 383-414.

**Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

**Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Laboratory press, Cold Spring Harbor, N. Y.

**Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

**Sirakova, T. D., S. S. Bardarov, J. I. Kriakov, and K. I. Markov.** 1989. Molecular cloning of mycobacterial promoters in *E. coli.* FEMS Microbiol. Lett. **59:**153-156.

**Tartof, K. D. and C. A. Hobbs.** 1987. Improved media for growing plasmid and cosmid clones. Bethesda Res. Lab. Focus **9**:12.

**Thomas, T. J., R. E. Andrews, Jr., and C. O. Thoen.** 1992. Molecular cloning and characterization of *Mycobacterium paratuberculosis* promoters in *Escherichia coli.* Vet. Microbiol. **32**:351-362.

**Timm, J. M., G. Perilli, C. Uez, J. Trias, G. Orefici, L. Fattorini, G. Amicosante, A. Oratore, B. Joris, J. M. Fr`ere, 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.

**Tsang, T., V. Copeland, and G. T. Bowden.** 1991. A set of cassette cloning vectors for rapid and versatile adaptation of restriction fragments. Biotechniques **10**:330.

**Whipple, D. L., R. B. Le Febvre, R. E. Andrews, Jr., and A. B. Thiermann.** 1987. Isolation and analysis of restriction endonuclease digestive patterns of chromosomal DNA from *Mycobacterium paratuberculosis* and other *Mycobacterium* species. J. Clin. Microbiol. **25**:1511-1515.

# CHAPTER 3. IDENTIFICATION OF GENE EXPRESSION SIGNALS IN *Mycobacterium paratuberculosis.*

A paper to be submitted to the *Journal of Bacteriology*

John P. Bannantine, R. G. Barletta, C. O. Thoen and R. E. Andrews, Jr.

## ABSTRACT

Ten novel promoter-containing clones were isolated from a *Mycobacterium paratuberculosis* DNA library cloned in the promoter selection vector pYUB76. The promoter-containing inserts were identified by expression of the promoterless *lacZ* reporter gene of pYUB76 and sequenced. The promoters exhibited a wide range of strengths, as indicated by their corresponding β-galactosidase reporter activities. β-galactosidase activity was determined in two hosts, *E. coli* and *M. smegmatis*, to assess diversity in transcription/translation signal recognition. The transcriptional or translational start signals were not recognized at similar levels in the two hosts. Those clones which exhibited the strongest β–galactosidase expression in *M. smegmatis* were studied further. Primer extension was used to locate the transcription start sites within the cloned fragments. Predicted open reading frames and codon usage were identified by computer analysis. Database searching for homologous sequences using the BLAST method revealed limited homologies. The

promoter regions characterized in this study were compared with those of several previously reported *Mycobacterium* promoters and a consensus promoter sequence was established. Hexanucleotide sequences centered approximately 35 and 10 base pairs upstream from the transcription startpoints did not correspond to the consensus hexanucleotides of *Escherichia coli* promoters.

## INTRODUCTION

Characterization of mycobacterial promoter structure has lagged well behind that of other bacterial systems. A detailed analysis of *Mycobacterium paratuberculosis* promoters may provide an increased understanding of genetic expression in mycobacteria and the extent of divergence in gene expression from other bacteria. Despite the importance of members in the genus *Mycobacterium* due to their pathogenic nature, researchers are reluctant to enter this field of study because genetic manipulation of these organisms can be frustrating and dangerous. Mycobacteria are among the most difficult prokaryotes to lyse due to the thick lipid-rich composition of their cell walls. Indeed, chemical or enzymatic lysis is relatively tedious requiring long periods of time, and mechanical disruption may still result in some incomplete lysis (Hurley *et al.*, 1987; Hines *et al.*, 1991). They are also slow-growing and refractile to genetic exchange by laboratory methods. For these reasons, *M. paratuberculosis* is an underdeveloped genetic system. To date, only

four genes have been cloned and sequenced (Stevenson *et al.*, 1991; Gilot *et al.*, 1993; Green *et al.*, 1989; Hance *et al.*, 1989) and only two promoters have been characterized in this species (Murray *et al.*, 1992; Thomas *et al.*, 1992). No composite class transposons or phages have been discovered in this species. A method of genetic exchange has only recently been established (Foley-Thomas *et al.*, 1995). By contrast, there are over 160 characterized *Escherichia coli* promoters that have been assembled and analyzed in detail (Hawley and McClure, 1983; O'Neill, 1989).

Recently, some studies have focused on cloning mycobacterial promoters (Sela and Clark-Curtiss, 1991; Barletta *et al.*, 1992; Das Gupta *et al.*, 1993; Thomas *et al.*, 1992; Timm *et al.*, 1994). The limitations of these studies are that cloned promoter fragments were not sequenced (Barletta *et al.*, 1992; Timm *et al.*, 1994) and transcription start site mapping was either performed with *E. coli* RNA (Sela and Clark-Curtiss, 1991) or not at all (Barletta *et al.*, 1992; Das Gupta *et al.*, 1993).

The promoter selection vector pYUB76 has been used to clone expression signals from mycobacteriophages (Barletta *et al.*, 1992). This vector replicates in *E. coli* as well as *Mycobacterium smegmatis* and has the following features: 1) a truncated reporter β-galactosidase gene devoid of promoter, ribosome-binding site and initiation codon; 2) a mycobacterial replicon (*oriM*), and an *E. coli* replicon (*colE1*); 3) a kanamycin-resistance gene that is expressed in both *E. coli* and mycobacteria. Activation of the reporter gene

depends on providing adequate transcriptional and translational initiation signals upstream. Furthermore, the ATG or GTG initiation codons of the translational signal derived from the insert must be in the same reading frame as *lacZ*.

This study was undertaken to extend the current knowledge of promoter architecture and gene expression in *M. paratuberculosis* by cloning and characterizing additional promoter elements. Using these data, a mycobacterial consensus promoter sequence was assembled.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** *M. paratuberculosis* ATCC 19698 was provided by D. Whipple NADC, Ames, IA and served as the source from which genomic DNA was isolated. *Escherichia coli* DH10B (Gibco-BRL, Gaithersburg, MD) served as the strain for routine plasmid maintenance and isolation. The shuttle vectors pYUB75 and pYUB76 (Barletta *et al.*, 1992) and *Mycobacterium smegmatis* $mc^2155$ were provided by R. Barletta, Department of Veterinary Science, University of Nebraska, Lincoln, NE. These two plasmids have a kanamycin resistance marker for selection, contain a promoterless *lacZ* gene, and can replicate in both *E. coli* and *M. smegmatis*.

**Growth conditions and extraction of *M. paratuberculosis* DNA.** For genomic DNA isolation, *M. paratuberculosis* was grown in Middlebrook 7H9 broth (Difco, Detroit, MI) containing 0.5% Tween-80, OADC enrichment (Difco) and 2 mg/L mycobactin J (Allied Monitor, Fayette, MO). The cells were harvested after 5 weeks of growth in tissue culture flasks (Corning, Inc., Corning, N. Y.) at 37°C. *M. paratuberculosis* pellets (1 to 3 g) were frozen at -70°C and genomic DNA was extracted by the method of Whipple *et al.* (1987).

**DNA cloning.** Genomic DNA from *M. paratuberculosis* was digested with *Sau*3AI and fractionated by agarose gel electrophoresis, and fragments from 100 to 2000 bp were purified by electroelution. Approximately 0.5 μg of purified genomic DNA fragments was cloned into 1 μg of dephosphorylated, *Bam*HI-cleaved pYUB76 in a final volume of 10 μl. The cloning of *Sau*3AI *M. paratuberculosis* chromosomal fragments into the *Bam*HI site 5' to the *lacZ* gene resulted in the construction of translational fusions, where properly positioned *M. paratuberculosis* promoters drive the transcription of *lacZ*. The ligation mixture was used to transform *E. coli* DH10B by electroporation. Potential transformants were plated on Luria Bertani (LB) medium containing 50 μg/ml kanamycin. The LB medium also contained 5-bromo-4-chloro-3-indolyl-β–D galactopyranoside (Xgal) added to each plate as described by Sambrook *et al.* (1989). Blue colonies on this medium were

considered to have a cloned *M. paratuberculosis* promoter sequence. The selective media for *M. smegmatis* electroporations was Middlebrook 7H10 agar (Difco, Detriot, MI) with 50 μg/ml kanamycin and Xgal added as described above. Conditions for the construction of pAJB60 are described elsewhere (Bannantine *et al.*, in preparation).

**Quantitative β–galactosidase assays.** The β–galactosidase assays for *E. coli* were performed as described by Miller (1972) with one modification. The minimal growth medium recommended by Miller was substituted for LB medium. β–galactosidase assays for *M. smegmatis* were performed as described by Barletta *et al.*, 1992. For all experiments, β-galactosidase activity is expressed as Miller units of activity (Miller, 1972).

**RNA extraction.** Total cell RNA was isolated from *E. coli* by routine methods (Ausubel *et al.*, 1989). Total cell RNA was extracted from *M. smegmatis* using the protocol described by Bashyam and Tyagi (1994).

**DNA sequencing.** Sequencing reactions were performed manually with Sequenase® Version 2.0 sequencing kit (United States Biochemical Corporation, Cleveland, OH) for generating molecular weight markers with phage M13 to resolve the size of the primer extension products. The *lacZ* fusion constructs in pYUB76 were

sequenced by using the *Taq* DyeDeoxy® Terminator Cycle Sequencing Kit (Perkin Elmer/Applied Biosystems Division (PE/ABD), Foster City, CA). Each sequencing reaction included 100 ng of template DNA and 10 pmol of sequencing primer in a total volume of 20 µl. Each reaction underwent 25 cycles (98°C for 15 s, 50°C for 2 s, 60°C for 4 min) in the GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer, Norwalk, Conn.). The sequencing products were purified through Centri-Sep columns (Princeton Separations, Adelphia, N.J.) and were loaded onto 6% denaturing polyacrylamide gels in an automated sequencer (373A DNA Sequencer; PE/ABD). Template DNA for automated sequencing was prepared by one of two methods; PEG preparation (Tartof and Hobbs, 1987) or Qiagen extraction. The Qiagen miniprep was performed according to the instructions of the manufacturer (Qiagen).

To sequence *lacZ*+ inserts from pYUB76, synthetic oligonucleotides with the following sequences 5'-CATTATTATCATGACATTAACC-3' (JB807) and 5'-GTGCTGCAAGGCGATTAAGTTG-3' (JB808) were employed. These primers were originally designed by Raul Barletta (University of Nebraska, Lincoln, NE) to PCR amplify inserts from pYUB75 or pYUB76. Other primers used in this project were designed to sequence further into the pAJB86 insert. These primers are shown in Table 1.

Table 1. Deoxyribonucleotide primers

| Primer | Sequence | Nucleotide position in the *lacZ* fusion[a] |
|---|---|---|
| **pAJB86:** | | |
| JB957 | 5'-TGCTCGACCTGGAGCGGATCAG-3' | -1,319 to -1,298 |
| JB1054 | 5'-TGCTCGGTGCCCAGTTCGTCAT-3' | -1,019 to -998 |
| JB958 | 5'-ACACCGAGATCGACCTGGAT-3' | -534 to -515 |
| JB923 | 5'-GATGAGATCCTGCAGGCGTG-3' | -301 to -282 |
| JB924 | 5'-GTCCGCGTGCTCCAGGAG-3' | -10 to +8 |
| **pAJB125:** | | |
| JB499 | 5'-TGATGCCGGCGACCTCCTGGT-3' | -1,111 to -1,091 |
| JB500 | 5'-ATCCGACCACGCCCGGAATTG-3' | +29 to +49 |

[a] Nucleotide position designations are all relative to the transcription start site.

**Computer analysis of DNA sequences.** Sequence alignments were made by using the software package SeqEd™ (Version 1.0.3; PE/ABD). Basic Local Alignment Search Tool (BLAST; Altschul *et al.*, 1990) was used for database homology searches and were performed via email of the sequence queries to blast@ncbi.nlm.nih.gov or through the world wide web using the biologist's control panel (URL address: http://gc.bcm.tmc.edu:8088/bio/bio_home.html). The ORFs were identified with either Codon Usage or Gene Finder in PBRT or Sequencher® 3.0. The Wisconsin package GCG (Devereux *et al.*, 1984) was used to identify conserved regions.

**Slot Blot Hybridization analysis.** Samples of *M. smegmatis* total cell RNA were heat and chemically denatured prior to application on a Hybond-N nylon membrane (Amersham, Arlington heights, IL). RNA samples were applied with a filtration manifold consisting of a lucite block containing a number of slots (slot blotting apparatus; Gibco-BRL). Conditions for prehybridization, hybridization and washes were performed by routine methods (Sambrook, *et al.*, 1989). The radioactive probes used for hybridization consisted of [α-$^{32}$P]dATP-labeled 3.2-kb *Eco*RI fragment from pCClac5 (Cupples and Miller, 1988) and 16S and 23S rRNA from *M. smegmatis* mc$^2$155 (pYUB76). Plasmid pCClac5 contains the 3.2-kb *Eco*RI fragment carrying *lacZ*. Unincorporated nucleotides were separated from labeled probe by two consecutive precipitations with ethanol in the presence of 7.5M ammonium acetate. Slot blot signals were quantitated by computer scaning the autoradiogram into the shareware program NIH image and performing an internal calibration. The levels of RNA used were adjusted by first normalizing the relative amounts of RNA used in the rRNA slots then similarly adjusting the RNA levels in the *lacZ* slots.

**Primer Extention analysis.** Total cell RNA from *M. smegmatis* strains harboring the *lacZ* fusions was used as a template for the synthesis of cDNA products from 5'-end-labeled synthetic oligonucleotide primers. The two primers (JB808 and JB1058, 5'-AGTCACGACGTTGTAAAACGACG-3' which correspond to nucleotides

+69 to +48, and +27 to +5 of *lacZ* in pYUB76, respectively. The "T" nucleotide in the *Bam*HI site is defined as +1) were end labeled by T4 polynucleotide kinase (Promega, Madison, WI) with [$\gamma$-$^{32}$P]ATP (Amersham). The annealing and reverse transcription reactions were performed with the AMV reverse transcriptase primer extension system following the instructions of the manufacturer (Promega). The products of the reactions were visualized by autoradiography after separation in 6% urea-polyacrylamide gels.

## RESULTS

**Cloning *M. paratuberculosis* promoters in pYUB76.** To effectively study promoter structure in *M. paratuberculosis*, a genomic library was cloned in the promoter-probe vector pYUB76 (Barletta *et al.*, 1992). 1,774 kanamycin resistant colonies were obtained after electroporation of the pYUB76 library into *E. coli* DH10B. Of those colonies, 74 were *lacZ*$^+$ (4.2%), indicative of promoter-containing clones. Forty-six randomly selected *lacZ*$^+$ *E. coli* clones were extracted and analyzed with restriction endonucleases. To assess the extent of divergence of *M. paratuberculosis* promoters from those of *E. coli*, the β–galactosidase activities supported by the promoters in *E. coli* were measured and compared with their corresponding activities in mycobacteria (Figure 1). Six of ten mycobacterial promoter elements functioned poorly in *E. coli*. These cloned sequences drove expression of β–galactosidase in *M.*

*smegmatis* at varying levels or strengths, according to histochemical screening and quantitative assays (Figure 1). In addition, the entire 1,774 clone library was pooled, amplified and electroporated into *M. smegmatis*. Twenty-four *lacZ* positive, kanamycin resistant colonies

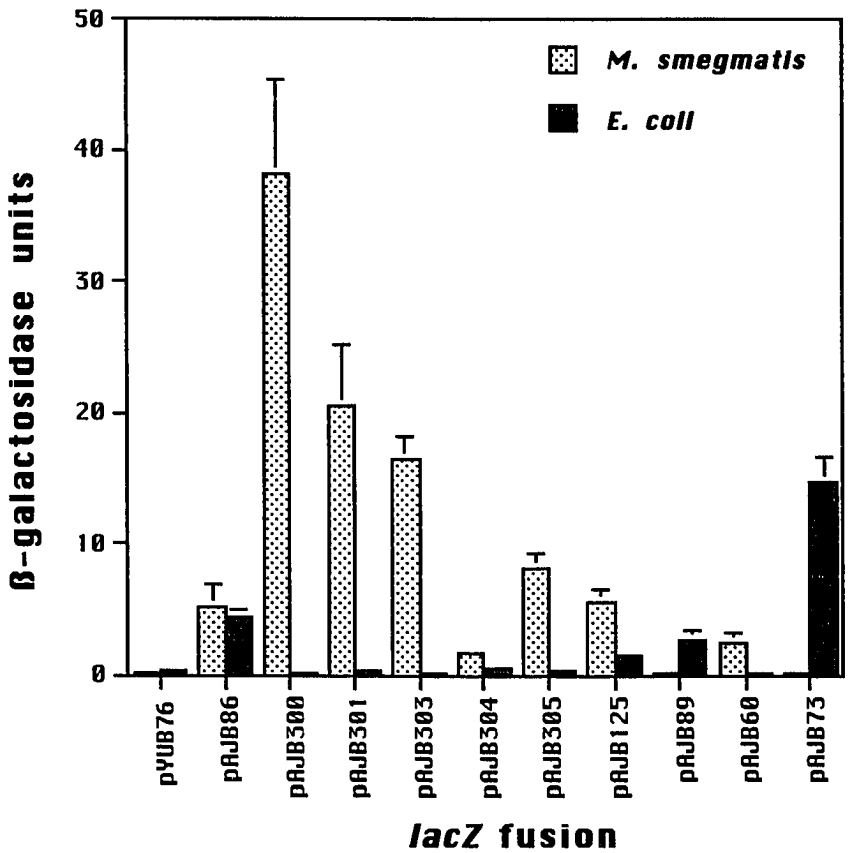



Figure 1. Relative expression rates of the pYUB76::*lacZ* fusions in *E. coli* and *M. smegmatis*. β-galactosidase Units are units of activity defined by Miller (1972). Experiments were performed in triplicate with the error bars representing the standard deviation of the mean.

were obtained. Table 2 summarizes the β-galactosidase activity for the *lacZ* fusions in *M. smegmatis* only. The twelve constructs that displayed the strongest β-galactosidase expression from both experiments were sequenced.

Table 2. β-galactosidase expression levels of pYUB76::*lacZ* fusions in *M. smegmatis.*

| Construct | Colony color | Initial Selection | β-galactosidase units[a] |
|---|---|---|---|
| pYUB76 | White | Both | 0.10±0.12 |
| pAJB303 | Dark Blue | *M. smegmatis* | 16.4±1.7 |
| pAJB300 | Dark Blue | *M. smegmatis* | 38.1±7.3 |
| pAJB304 | Blue | *M. smegmatis* | 1.8±0.47 |
| pAJB301 | Dark Blue | *M. smegmatis* | 20.6±4.5 |
| pAJB73 | Light Blue | *E. coli* | 0.29±0.22 |
| pAJB305 | Blue | *M. smegmatis* | 8.2±1.2 |
| pAJB86 | Blue | *E. coli* | 5.2±1.6 |
| pAJB60 | Blue | *E. coli* | 2.5±0.75 |
| pAJB125 | Blue | *E. coli* | 5.7±0.87 |
| pAJB89 | Light Blue | *E. coli* | 0.29±0.21 |

[a] β-galactosidase units are measured in Miller units of activity (Miller, 1972). The reported figures are the average of three trials ± the standard deviation.

**DNA sequence analysis.** Sequence analysis revealed that two of the twelve constructs were identical to the insert in pAJB300 and one was identical to pAJB304. Of the nine novel *lacZ*$^+$ constructs that have been sequenced, only one (pAJB86) was not completely sequenced. The pAJB86 insert was too large and repeated attempts at sequencing further into its insert did not generate reliable

sequence data. Their sizes, % guanine plus cytosine (%G+C) content, and other information are presented in Table 3. The %G+C of sequenced inserts agrees well with the published figure of 66-67% for *M. paratuberculosis* (McFadden *et al.*, 1987). The putative open reading frames (ORFs) have also been identified by computer analysis and confirmed for conditions specific to *lacZ* expression in pYUB76 (i. e. all cloned sequences must contain a promoter

Table 3. Sequence data from *lacZ*$^+$ *M. smegmatis* contructs.

| Construct name | Insert Size (bp) | %GC content | No. of Mycobacterial amino acids in fusion |
|---|---|---|---|
| pAJB303 | 610 | 67.2 | 95 |
| pAJB300 | 412 | 66.1 | 24 |
| pAJB304 | 663 | 68.2 | 7 |
| pAJB301 | 292 | 70.9 | 61 |
| pAJB73 | 104 | 69.2 | 15 |
| pAJB305 | 559 | 71.0 | 10 |
| pAJB86 | >1,800 | 66.3 | 20 |
| pAJB60 | 531 | 66.5 | ND[a] |
| pAJB125 | 1,660 | 60.8 | 81 |

[a] ND, not determined

structure, ribosome binding site, a start codon, and must be in frame with the ninth codon (GUC) of *lacZ*). The complete nucleotide sequences as well as the deduced ORFs are shown in Appendix A. The potential ribosome binding sites (RBS) are also indicated in Appendix A. BLAST searches of Genbank and EMBL nucleic acid databases showed different degrees of homology to genes of other

organisms (Appendix B). Several of the homologies obtained from the databases were sequences from *Streptomyces* species. This was not suprising since *Streptomyces* is phylogenetically related to *Mycobacterium.* BLAST searches were also performed on the deduced amino acids of *M. paratuberculosis* that participate in the *lacZ* fusions (Appendix C). Among the strongest amino acid matches were the human collagen alpha 1 chain with the *M. paratuberculosis* ORF from pAJB301. This match produced a sum probability of 0.00015 (the smaller the number, the closer the match).

The high G+C content in *M. paratuberculosis* genomic DNA made it particularly difficult to sequence. The best results were obtained when high temperature denaturation and extension conditions were used with *Taq* polymerase and the reactions were combined with 2% final concentration of dimethyl sulfoxide. An average reading length of 328 bp was obtained from successful reactions.

**Codon Usage.** A compilation of the codons employed in the N-terminal mycobacterial ORFs which participated in the *lacZ* fusions is presented in Figure 2. As has been found for other microorganisms, there is a correlation between the %G+C content of the genome and the percent of codons with a guanine or cytosine in the third position (Muto and Osawa, 1987). The data compiled for *M. paratuberculosis* in this study shows that 73% of the codons contain a dG or dC in the third position whereas codons ending in dT and dA are quite rare, at 15 and 12%, respectively.

**Slot Blot Hybridization analysis.** To estimate transcriptional activity from the *lacZ* fusions, slot blot hybridization experiments were performed. Total cell RNA from the *lacZ* positive *M. smegmatis* strains was applied to a nylon membrane and hybridized (Figure 3) as described in Materials and Methods. Based on these data, *M. smegmatis* (pAJB303) showed the highest transcriptional activity. No transcriptional activity was detected in *M. smegmatis* (pYUB76) by either northern hybridization analysis (data not shown) or slot blot analysis (Figure 3). Figure 4 summarizes the transcriptional and translational data in conjunction with schematic maps of all the *lacZ* fusions.

| | | | |
|---|---|---|---|
| F TTT 0 | S TCT 2 | Y TAT 0 | C TGT 0 |
| F TTC 4 | S TCC 2 | Y TAC 0 | C TGC 3 |
| L TTA 0 | S TCA 0 | * TAA 0 | * TGA 0 |
| L TTG 2 | S TCG 4 | * TAG 0 | W TGG 3 |
| L CTT 4 | P CCT 3 | H CAT 4 | R CGT 0 |
| L CTC 10 | P CCC 13 | H CAC 8 | R CGC 11 |
| L CTA 2 | P CCA 2 | Q CAA 3 | R CGA 7 |
| L CTG 15 | P CCG 14 | Q CAG 2 | R CGG 10 |
| I ATT 1 | T ACT 0 | N AAT 2 | S AGT 0 |
| I ATC 7 | T ACC 8 | N AAC 5 | S AGC 5 |
| I ATA 2 | T ACA 3 | K AAA 3 | R AGA 1 |
| M ATG 8 | T ACG 6 | K AAG 1 | R AGG 1 |
| V GTT 5 | A GCT 7 | D GAT 12 | G GGT 6 |
| V GTC 9 | A GCC 14 | D GAC 8 | G GGC 15 |
| V GTA 6 | A GCA 0 | E GAA 7 | G GGA 1 |
| V GTG 10 | A GCG 12 | E GAG 9 | G GGG 7 |

Figure 2. Codon usage within the seven open reading frames found in the pYUB76::*lacZ* fusions. A total of 313 codons went into this compilation.

Figure 3. Quantitation of steady state RNA from *M. smegmatis* harboring pYUB76::*lacZ* fusions. Total cell RNA from *M. smegmatis* was quantitatively spotted to the nylon membrane shown and hybridized with either radiolabeled rRNA from *M. smegmatis* (left column) or a *lacZ*-specific fragment from pCClac5 (right column). The strength of the signals was determined using an internal calibration in the Computer program NIH image. The strain designations are as follows: AJB-217: *M. smegmatis* (pAJB125); AJB-173: *M. smegmatis* (pAJB86); AJB-174: *M. smegmatis* (pAJB89); AJB-112: *M. smegmatis* (pYUB76); AJB-201: *M.smegmatis* (pAJB73); AJB-274: *M. smegmatis* (pAJB304); AJB-275: *M. smegmatis* (pAJB305); AJB-207: *M. smegmatis* (pAJB60); AJB-270: *M. smegmatis* (pAJB300); AJB-271: *M. smegmatis* (pAJB301); AJB-273: *M. smegmatis* (pAJB303).

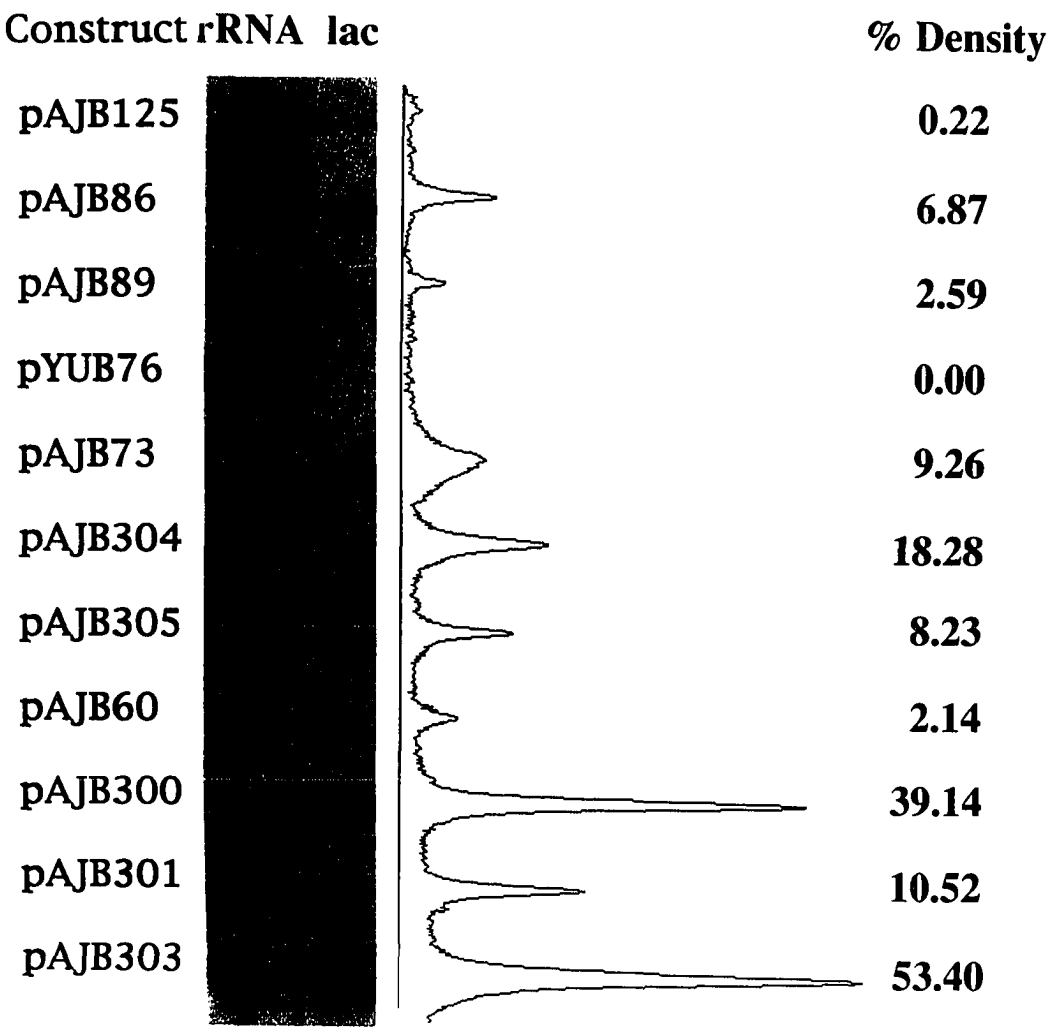
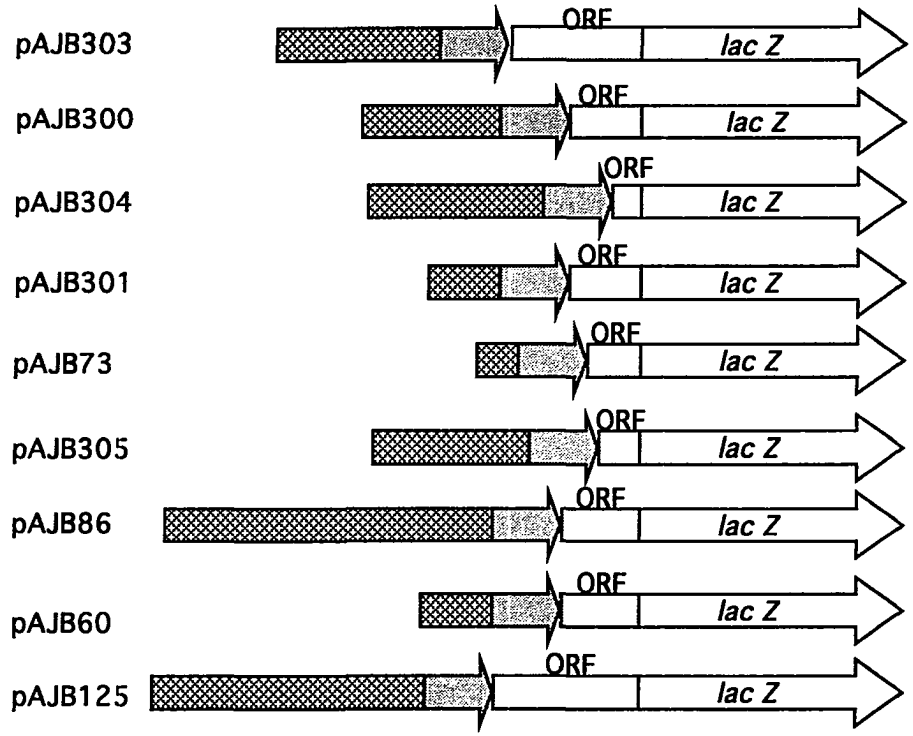

Construct rRNA lac
% Density
pAJB125
0.22
pAJB86
6.87
pAJB89
2.59
pYUB76
0.00
pAJB73
9.26
pAJB304
18.28
pAJB305
8.23
pAJB60
2.14
pAJB300
39.14
pAJB301
10.52
pAJB303
53.40

Figure 4. Schematic diagram of the pYUB76::*lacZ* fusions along with their mRNA and reporter activity levels. The construct maps are drawn to scale and the translational efficiencies are shown for both the *M. smegmatis* and *E. coli* backgrounds. The constructs are listed from highest amount of *lacZ* fusion mRNA to lowest amount as measured by percent density in Figure 3. The pAJB60 *lacZ* fusion is also shown here for the sake of comparison only since the size of the open reading frame and location of a promoter structure could not be determined.




| | Steady state RNA levels in *M. smegmatis* | ß-galactosidase activity *M. smegmatis* | ß-galactosidase activity *E. coli* |
|---|---|---|---|
| pAJB303 | 53.40 | 16.4±1.7 | >0.2 |
| pAJB300 | 39.14 | 38.1±7.3 | >0.2 |
| pAJB304 | 18.28 | 1.8±0.47 | 0.6±0.24 |
| pAJB301 | 10.52 | 20.6±4.5 | 0.4±0.21 |
| pAJB73 | 9.26 | 0.29±0.22 | 14.7±1.98 |
| pAJB305 | 8.23 | 8.2±1.2 | 0.3±0.13 |
| pAJB86 | 6.87 | 5.2±1.6 | 4.5±0.63 |
| pAJB60 | 2.14 | 2.5±0.75 | 0.3±0.33 |
| pAJB125 | 0.22 | 5.7±0.87 | 1.5±0.47 |

**Primer Extention analysis.** To localize the promoter region within the cloned inserts, the transcription start sites were mapped. The mRNA initiation sites of the cloned *lacZ* fusion transcripts were identified by primer extension with reverse transcriptase and a 5'-end-labeled oligonucleotide primer. Comparison of the electrophoretic mobility of the primer extension product with that of the products of the dideoxy sequencing reactions in Figures 5, 6 and 7 was used to map the 5' end of the transcripts. Primer extension products were generated for the pAJB73 (Figure 5), pAJB86, pAJB300, pAJB301, pAJB303, pAJB304, and pAJB305 constructs (Figure 6). The transcripts from pAJB300 and pAJB304 were mapped with the JB808 primer (Figure 7) and the JB1058 primer (Figure 6). Several extension products appeared on the autoradiogram of the primer extensions performed with the JB808 primer (Figure 7). A defined extension product could not be reliably determined for pAJB300 when using JB808, however, a 101 bp product was identified for pAJB304 (Figure 7). The JB1058 primer produced an extension product for pAJB304 that was 5 bp longer than that produced with the JB808 oligonucleotide (Figure 6). The startpoint of the pAJB303 transcript was not precisely determined due to the large size of the cDNA product (too large to comigrate with resolvable DNA sequencing ladders (Figure 6)). Therefore an estimate of the extension product was made using radiolabeled øX174 DNA markers digested with *Hin*fI (data not shown). Primer extension products were not observed for pAJB60 or pAJB89.

Figure 5. Primer extension analysis of pAJB73::*lacZ*. Details of the experiment are described in Materials and Methods. The location of the 81 and 82 nucleotide (nt) extension products of the transcript from pAJB73 (lane 1) are indicated in the left hand margin. The lanes labeled G, A, T, and C contain the dideoxy sequencing reactions with the M13 DNA template. Lanes 2, 3, and 4 contain failed primer extension reactions from pAJB60, pAJB89, and pAJB301, respectively.


1 2 3 4 G A T C
82 nt
81 nt

Figure 6. Primer extension analysis of *lacZ* fusion transcripts from pAJB300 (lane 1), pAJB301 (lane 2), pAJB303 (lane 3), pAJB86 (lane 4), pAJB305 (lane 5), pAJB125 (lane 6), pAJB304 (lane 7), pAJB60 (lane 8), and pAJB89 (lane 9). The position of the primer extension products are indicated by arrows along with the nucleotide (nt) length of the product. An extension product was not detected for either pAJB60 or pAJB89. Details of the experiment are described in Materials and Methods.

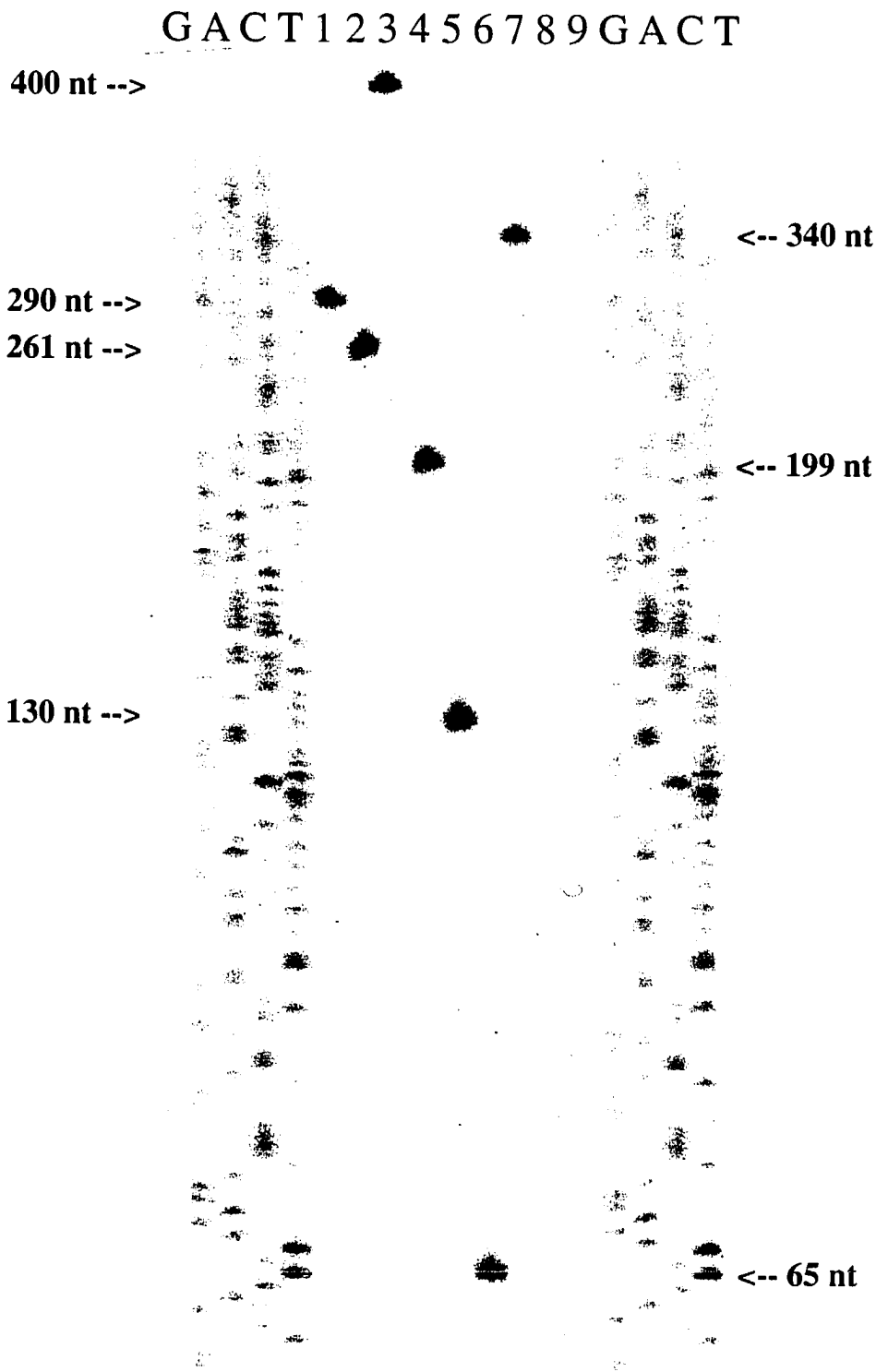
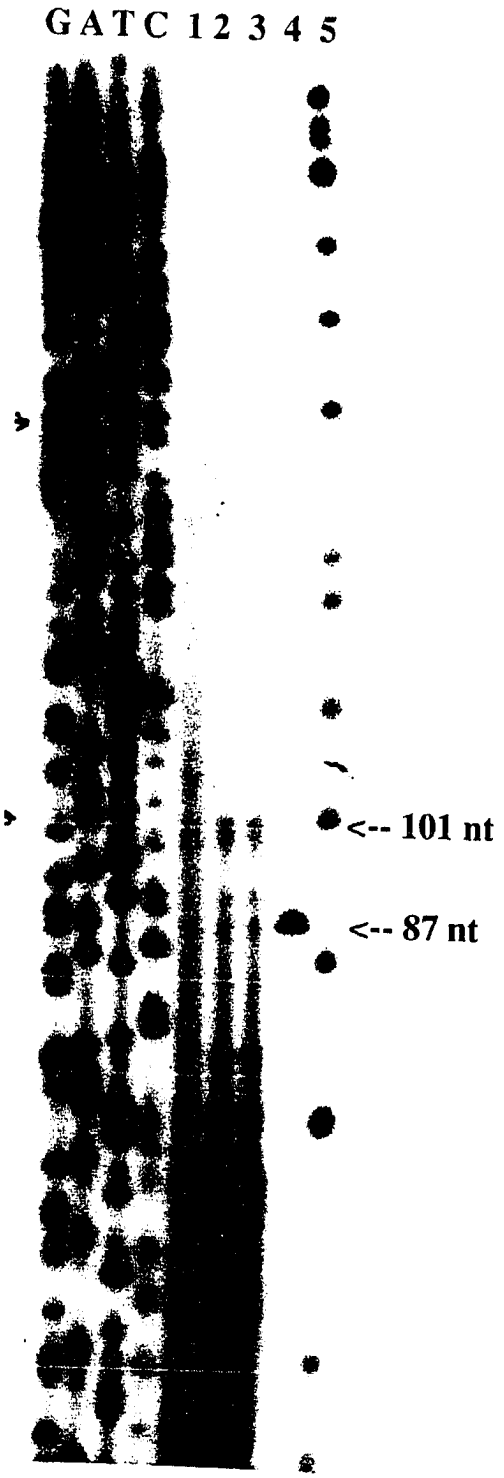

G A C T 1 2 3 4 5 6 7 8 9 G A C T
400 nt -->
290 nt -->
261 nt -->
130 nt -->
<-- 340 nt
<-- 199 nt
<-- 65 nt

Figure 7. Primer extension analysis of pAJB300::*lacZ* and pAJB304::*lacZ* with the JB808 primer. Details of the experiment are described in Materials and Methods. The location of the 101 and 87 nucleotide (nt) extension products of the transcript from pAJB304 (lane 2 and 3) and the control (lane 4) are indicated in the left hand margin. The lanes labeled G, A, T, and C contain the dideoxy sequencing reactions with the M13 DNA template.


G A T C 1 2 3 4 5
<-- 101 nt
<-- 87 nt

**Putative ribosome binding sites in *M. paratuberculosis* mRNA.** McLaughlin *et al.* (1981) observed that the ribosome binding sites of prokaryotic mRNA exhibit complementarity to the 3' region of 16S rRNA. With the exception of pAJB89 and pAJB60, a region of homology to the 3' terminus of *M. leprae* 16S rRNA followed after several bases by an initiation codon and an open reading frame for translation was observed for all the *lacZ* fusions (Appendix A). Nucleotides that may base pair with rRNA are highlighted with oversize letters in Table 4. The predicted free-energy of base pairing ($\Delta G$) of these ribosomal binding sites with 16S rRNA, calculated according to the rules of Tinoco *et al.* (1973), ranged from -3.6 to -19 Kcal.

**Determination of a consensus promoter sequence.** The eight promoters that had their transription start sites identified by primer extension analysis were aligned as shown in Table 5. The frequency at which any one nucleotide occurred at each position extending from +1 to -40 was recorded (Figure 8). From this analysis, no consensus "*E. coli*-like" promoter sequences were found. Instead, single nucleotides at various spacing were shown to occur at greater than 60% frequency.

The -35 and -10 regions of all the mycobacterial promoters that contain an experimentally determined transcription start site were assembled (Table 6). Frequently occurring nucleotides in a

Table 4. Nucleotide sequences of putative *M. paratuberculosis* ribosome binding sites

| Construct | Nucleotide Sequence | ΔG (kcal) | Position of Start Codon |
|---|---|---|---|
| pAJB303 | UGGGCUGGGCGGGCUUGCGGAAUCACCGAAAAGAUAAGUGAAAAUG | -6.0 | +94 |
| pAJB300 | GGACGAUUUCACGAACCUGUGUACGCAAGUUACAGUUAUCUCCAUG | -4.8 | +96 |
| pAJB304 | GAGCUGGACGGGAGGUCGUCGAUG | -15 | +22 |
| pAJB301 | CGACCAGGGUUUCAUGCACGGCGCUGGCGACCAUCAGCCUGCGGUG | -5.6 | +59 |
| pAJB73 | UGCUCGAGCUCGAGCUCGGUGUG | -3.6 | +15 |
| pAJB305 | GCGACAAGGCAAUGUUGUCGGUACCGGAUGAGAUGGUCUCUUCAUG | -10 | +79 |
| pAJB86 | GGCUACCGCGUCGAAGCCGGAACCAUGACUGGCGUUCGGCAGAAUG | -8.0 | +118 |
| pAJB89 | UCCGACAUGGCGCCCGGAUAGAAGGAGGCGAUCCCGUC | -19 | ND[a] |
| pAJB125 | AGGAGGUCGCCGGCAUCAGCGCCGGCACCACGUUGUUGCGCCACGUG | -19 | +76 |
| 16S rRNA 3' | AUUCCUCCACUA | | |

[a] ND = Not determined

Table 5. Alignment of the transcription start sites determined using the JB1058 primer in the pYUB76::*lacZ* fusions.

| -35 region | -10 region | +1 | Construct |
|---|---|---|---|
| AGGGCGG**TGGCGT**CGCCGGTGTAGCCGAA**CGGCAC**GTGCGCGT | | | pAJB303 |
| AGACCC**TGACGC**TGGCCGACCTCGGCGCG**CAGCCG**ACCGCGCA | | | pAJB300 |
| ACGCC**AAGGAC**AACGGCCGTATCCGGTA**CCAACG**GGTGTGCG | | | pAJB304 |
| GG**TCCAGT**ACACCGCGAGTTCGCGCACG**CTGGCC**GGCAGCGTCT | | | pAJB301 |
| GTG**TGCCGC**TTGAACCGGCCCAGCTCCCG**CTCCAG**GGTGACGT | | | pAJB73 |
| GGCAGT**TGTTGG**AGTTTCTGTCGGACGGT**TGGTTG**GCGGCATT | | | pAJB305 |
| ACTCCCGA**TGACGT**TGCACGGCTGGGATTAA**CGGTCC**GCGTGCT | | | pAJB86 |
| GCATCA**TTAAAG**ATCGANGGCGCCGGGNT**CATGTC**CCTTCACC | | | pAJB125 |

given position are shown in bold in Table 6. Based on these data, a consensus for the -35 region was established as T(G/T)G(A/C)GT. The -10 region was established as CAGCCG. The -35 region is typically defined by a thymine in the first position of the hexamer, which occurs in 82% of the sequences compiled in Table 6. The guanine nucleotide was never found to occupy this first position. A guanine nucleotide at the fifth position of the -35 hexameric region occurred in 7 of the 9 *M. paratuberculosis* promoters but was not observed for any of the promoters in other mycobacterial species. A cytosine or adenine had an 88% occurrence in the fourth position of the -35 hexamer. The begining of the -10 region is typically defined by a pyrimidine (cytosine or thymine) which had an occurrence of 82%. For *M. paratuberculosis* promoters defined in this study, both of the -10 and -35 regions contained a guanine or cytosine in the last position of the hexamer with 78% frequency. An adenine was never observed in the sixth position of either region for *M. paratuberculosis.*

Figure 8. Distribution of nucleotide bases relative to the transcription start site. Shown is a histogram plot of each nucleotide position from -40 to +1 with the relative frequency at each location.

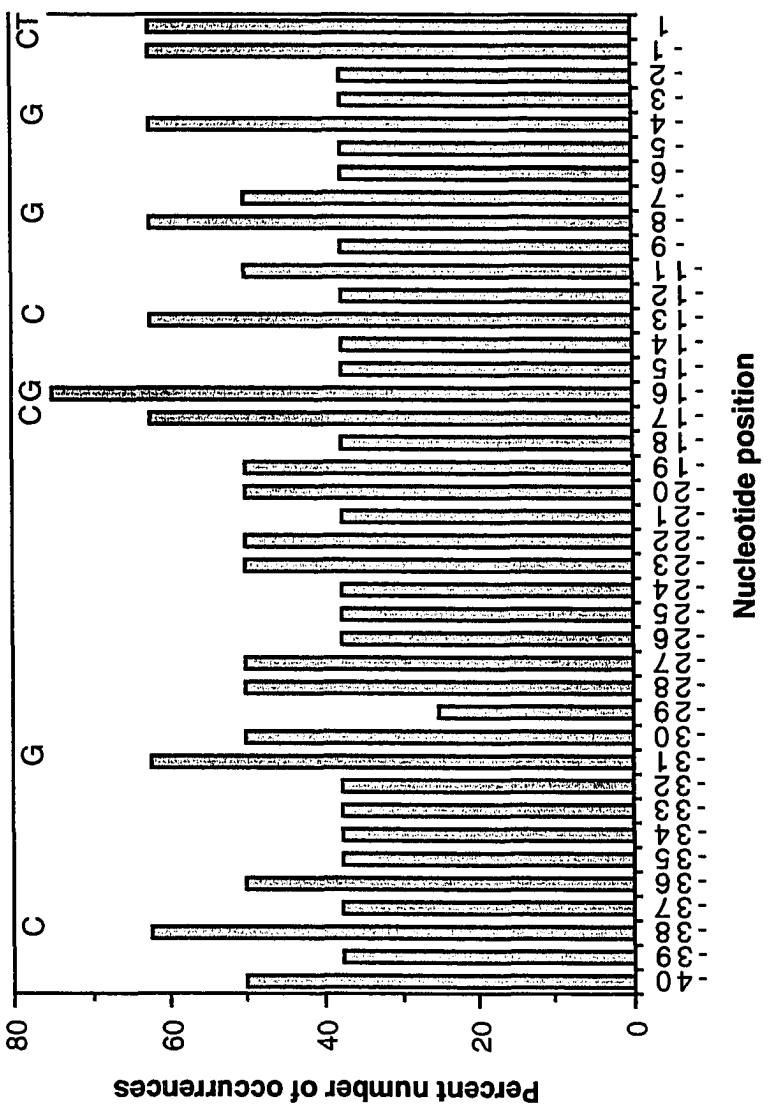

Percent number of occurrences
0
20
40
60
80
Nucleotide position
C
G
CG
C
G
G
CT

Table 6. -35 and -10 promoter regions and their spacings in mycobacteria.

| Promoter | -35 Region | | -10 Region | | +1 | | Initial Codon |
|---|---|---|---|---|---|---|---|
| Promoters from this study: | | | | | | | |
| pAJB303 | TGGCGT | 16 | CGGCAC | 7 | T | 92 | ATG |
| pAJB300 | TGACGC | 17 | CAGCCG | 7 | A | 195 | ATG |
| pAJB304 | AAGGAC | 17 | CCAACG | 7 | G | 21 | ATG |
| pAJB301 | TCCAGT | 20 | CTGGCC | 9 | T | 55 | GTG |
| pAJB73 | TGCCGC | 20 | CTCCAG | 7 | T | 13 | GTG |
| pAJB305 | TGTTGG | 17 | TGGTTG | 7 | T | 77 | ATG |
| pAJB86 | TGACGT | 17 | CGGTCC | 6 | T | 116 | ATG |
| pAJB125 | TTAAAG | 17 | CATGTC | 7 | C | 74 | GTG |
| Other mycobacterial promoters:[a] | | | | | | | |
| *hsp60* BCG | TTGCAC | 17 | TAAGAA | 6 | T | 183 | ATG |
| $P_{AN}$ *M. ptb.* | TCGACA | 17 | TACACT | 7 | A | 40 | ATG |
| 16SrRNA *leprae* | TTGACT | 16 | ATTAAT | 7 | G | | |
| 16SrRNA *M. tub.* | TTGACT | 18 | TAGACT | 7 | T | | |
| *cpn60 M. tub.* | TGCTCA | 17 | GGCGGC | 7 | A | 29 | ATG |
| *mpb70* BCG | CCGATC | 17 | CATCAG | 6 | G | 176 | ATG |
| *bla fortuitum* | TTCAAA | 19 | TACGCT | 7 | A | 0 | ATG |
| *ask smegmatis* | CCCACG | 17 | ACGCTG | 7 | G | 250 | GTG |
| 85A *M. tub.* | TTGACT | 22 | CGCCTG | 7 | A | 63 | ATG |
| *Mycobacterium* consensus | T(G/T)G(A/C)GT | | CAGCCG | | | | |
| *E. coli* consensus | TTGACA | 17-19 | TATAAT | | | | |

[a] *Mycobacterium* sp. promoter regions as compiled by Kremer *et al.* (1995). Only promoters for which the transcription start sites are known were included.

# DISCUSSION

These studies have established that *M. paratuberculosis* expression signals are capable of driving the expression of β-galactosidase in *M. smegmatis* and *E. coli*. Based on β-galactosidase expression levels in the pYUB76::*lacZ* fusion constructs, *M. smegmatis* was the preferred host for initial selection of promoter-containing inserts when compared with *E. coli*. The work described herein effectively doubles the current state of knowledge regarding promoter structure in the entire mycobacterial genus since only nine mycobacterial promoters have been previously characterized in the detail described here (Kremer *et al.*, 1995). With the exception of pAJB60 and pAJB89, all of the sequenced promoters showed clear transcriptional initiation signals. Moreover, typical start codons were not identified in pAJB60 and pAJB89, even though the reporter gene is expressed. Thus, it is unlikely these two constructs contain *bona fide* promoter sequences. Importantly, the combination of screening in mycobacterial hosts and use of primer extension analysis distinguished false reporter gene activity from *bona fide* promoter-containing fragments.

The ability to recognize *M. paratuberculosis* gene expression signals was variable in *E. coli* and *M. smegmatis* depending on which host was employed for initial selection of the *lacZ* fusions. The constructs initially selected in *M. smegmatis*, displayed strong expression of the reporter gene in *M. smegmatis* but not in *E. coli*.

For constructs whose initial selection was in *E. coli*, the reverse was observed. Reporter gene expression in both hosts was observed only with pAJB86, pAJB89, pAJB73 and pAJB125. It is important to note that plasmid copy number between the two hosts was not tested. While it is likely that the two hosts maintain different copy numbers for the pYUB76::*lacZ* fusions, there is no *a priori* reason to suspect variable copy number within a single host.

Strong promoter activity was observed in the *lacZ* fusion constructs of pAJB300, pAJB301, and pAJB303. When comparing the Miller units obtained in the quantitative β-galactosidase assay with those reported by Barletta *et al.* (1992) for cloning mycobacteriophage promoters in pYUB76, it was discovered that pAJB300 is stronger and pAJB301 and pAJB303 are slightly weaker than the cloned mycobacteriophage promoters. These data make pAJB300 a particularly interesting clone since bacteriophage promoters are typically classified as strong promoters. The cloned pAJB300 or pAJB303 inserts may be good candidates for expression of heterologous antigens in *M. bovis* BCG for recombinant vaccine development (Stover *et al.*, 1991; Stover *et al.*, 1993) because of their high level expression of *lacZ* in this study. The β–galactosidase assays indicate that pAJB300 was translationally the most active with pAJB303 and pAJB301 a comparable second.

The slot blot hybridization of *M. smegmatis* RNA was performed to quantitate steady state levels of the *lacZ* fusion message. Thus, pAJB303 might contain the strongest promoter.

Alternatively, the pAJB303 message might be more stable. The 5' leader sequence of some stable transcripts such as the T4 bacteriophage gene 32 or the *E. coli ompA* can stabilize other sequences fused downstream from them (Ehretsmann *et al.*, 1992). For example, the *lacZ* transcript has a half-life of 90 seconds, but a gene 32::*lacZ* fusion has a half-life of 20 min (Ehretsmann *et al.*, 1992).

The use of the *Taq* method for sequencing the fragments cloned in pYUB76 gave the best results since it enabled high temperature polymerization of GC-rich templates. Even though the complete nucleotide sequence was not determined for pAJB86, the promoter region was clearly identified.

The ribosome binding site or Shine-Dalgarno sequence (Shine and Dalgarno, 1974), which orients the 30S ribosomal subunit to the transcript, is a purine-rich region based on the sequence of the 16S rRNA of *M. leprae* (Estrada-G *et al.*, 1988). Although we have not identified initiation codons directly by ammino acid sequencing, eight of the ten *lacZ* fusion mRNAs display regions of complementarity to the 3' region of *M. leprae* 16S rRNA followed by a possible initiation codon and an open reading frame. The free-energies of interaction ($\Delta G$) for the proposed ribosome binding sites vary from -3.6 to -19 Kcal as compared to -9.4 Kcal for the prototype *E. coli* ribosome binding sequence AGGA (McLaughlin *et al.*, 1981).

The codon usage data compiled for *M. paratuberculosis* in this study shows that 73% of the codons contain a dG or dC in the third

position whereas codons ending in dT and dA are quite rare, at 15 and 12%, respectively. These figures are consist with codon usage in *M. leprae* (Honoré *et al.*, 1993) where 75% of the codons end in dG or dC and codons ending in dT and dA are 17 and 8%, respectively.

Some ambiguity exists with the primer extension data regarding pAJB304. The JB808 primer produced a cDNA product 101 bp in length that was five base pairs downstream from the start site identified using the JB1058 primer. Reasons for this discrepency have yet to be determined, but it is most likely a premature stop with the JB808 primer since the extension product obtained with the JB1058 primer is much "cleaner" with no premature stops. Direct evidence that the sequence upstream from the mapped transcription start point acts as the promoter will be obtained by deletion analysis, RNA polymerase binding studies, or by promoter mutations for all of the *lacZ* fusion constructs.

The cDNA product obtained for the pAJB303 construct was too large to comigrate with resolvable sequencing ladders. Therefore the length of that extension product could not be determined to single base pair accuracy. A long extension product from pAJB303 is consistent with the long mycobacterial ORF that was cloned in this construct. Primer extension products have been shown to be as long as 1 kb (Kroll *et al.*, 1991).

All eight of the transcriptionally mapped promoters display striking conformity in their -35 and -10 regions with the corresponding consensus nucleotides (T, G/T, G, A/C, G, T and C, A, G,

C, C, G, respectively). Indeed, bases at four out of twelve positions are invariant in all eight promoters and bases at three additional positions are conserved in five out of eight promoters (Table 6). Moreover, the "spacer" between the -35 and -10 regions (16-20 base pairs) corresponds to the preferred interval for prokaryotic promoters. Both the conserved nucleotides and the spacing between them is consistent with the compilation of other sequenced promoters from *Bacillus subtilis* (Moran *et al.*, 1982) and *E. coli* (Hawley and McClure, 1983). Additional promoter sequences must be compared to test or refine the consensus regions. The -35 region of the mycobacterial consensus resembles that of *E. coli* whereas the -10 region does not. Upstream regions of mycobacterial genes have been shown to have a higher G+C content than their *E. coli* counterparts (Dale and Patki, 1990). This observation was confirmed in the present study as the conserved regions possess a G+C content of 66.7% compared with 16.7% for the *E. coli* consensus. As was pointed out by Dale and Patki (1990), analysis of the sequence of many mycobacterial genes does not reveal any regions upstream from the putative translational start position that resemble a consensus prokaryotic promoter. The results of this study have, for the first time, revealed a mycobacterial consensus promoter.

With the consensus sequence information established, it became of interest to compare experimentally determined *M. paratuberculosis* promoter strengths with the degree of sequence conservation in the conserved promoter regions. Theoretically, a

DNA sequence containing all of the consensus nucleotides with the proper spacing should exhibit the strongest promoter activity. The pAJB300 construct has been shown to produce the most reporter gene product. Examination of the conserved regions of the pAJB300 construct shows that the -10 region exactly matches the -10 consensus sequence. Also four of six positions in the -35 region matched consensus nucleotides. Similarly, the pAJB303 construct was shown to be the most transcriptionally active based on quantitative hybridization experiments. The -35 region in pAJB303 exactly matches the -35 consensus sequence and three positions in the -10 region contain consensus nucleotides.

This study has provided the framework for many future studies of promoters in mycobacteria. With these cloned promoters, mutations can be created that disrupt spacing and add or remove consensus base pairs to observed how this affects promoter strength. Deletion analysis or RNA polymerase binding studies could be performed to determine the minimum sequence requirement for a functional promoter.

Conditions such as osmolarity, pH, temperature, water concentration, etc. could upregulate or downregulate expression of the *lacZ* reporter gene. These laboratory stresses, including exposure to oxidative radicals, may mimic the type of environmental changes encountered by these intracellular parasites and may provide clues as to patterns of gene expression *in vivo*. Iron-regulated promoters could be cloned from mycobactin-dependent strains. To accomplish

this, a promoter library would be constructed using pYUB76. *M. smegmatis* clones carrying promoter signals would be identified by β-galactosidase production on iron-deficient medium. The final screen would involve those promoters that are repressed (white on Xgal plates) on iron sufficient medium. These studies could then directly contribute to understanding of mechanisms involved in mycobacterial pathogenicity. Finally, more mycobacterial promoters need to be cloned and characterized in order to further refine the consensus sequence.

The results herein show that *M. paratuberculosis* sequences can serve as promoters in *M. smegmatis* and occasionally in *E. coli.* Promoter-probe studies do not prove that a cloned sequence actually has promoter activity in its original context. The fact that transcriptional activity in *E. coli* may be attributed to a fortuitous "promoter-like" sequence and not a legitimate mycobacterial promoter cannot be ruled out. Future experiments will be focused on deletion analysis and ultimately RNA polymerase binding assays to presicely define the sequence required for promoter activity. Little is known about mycobacterial gene experssion, gene regulation, and the influence of environmental conditions on expression levels of these cloned promoters. This study represents the most comprehensive and detailed analysis of *M. paratuberculosis* promoter elements to date.

## ACKNOWLEDGEMENTS

We thank the DNA sequencing and synthesis facility (1184 Molecular Biology building, Iowa State University) for performing the automated sequencing of the pYUB76::*lacZ* fusion inserts.

## REFERENCES

**Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403-410.

**Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman,J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. John Wiley & Sons, New York.

**Barletta, R. G., D. D. Kim, S. B. Snapper, B. R. Bloom, and W. R. Jacobs, Jr.** 1992. Identification of expression signals of the mycobacteriophages Bxb1, L1 and TM4 using the *Escherichia-Mycobacterium* shuttle plasmids pYUB75 and pYUB76 designed to create translational fusions to the *lacZ* gene. J. Gen. Microbiol. **138**:23-30.

**Bashyam, M. D. and A. K. Tyagi.** 1994. An efficient and high yielding method for isolation of RNA from mycobacteria. BioTechniques **17**:834-836.

**Cupples, C. G. and J. H. Miller.** 1988. Effect of amino acid substitutions at the active site in *Escherichia coli* β-galactosidase. Genetics **120**:637-644.

**Dale, J. W., and A. Patki.** 1990. Mycobacterial gene expression and regulation. *In* J. McFadden (ed.), Molecular biology of mycobacteria. Surrey University Press, London, pp. 173-198.

**Das Gupta, 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.

**Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387-395.

**Ehretsmann, C. P., A. J. Carpousis, and H. M. Krisch.** 1992. mRNA degradation in procaryotes. FASEB J. **6:**3186-3192.

**Estrada-G, I. C. E., F. I. Lamb, M. J. Colston, and R. A. Cox.** 1988. Partial nucleotide sequence of 16S ribosomal RNA isolated from armadillo-grown *Mycobacterium leprae.* J. Gen. Microbiol. **134:**1449-1453.

**Foley-Thomas, E., D. L. Whipple, L. E. Bermudez, and R. G. Barletta.** 1995. Phage infection, transfection, and transformation of *Mycobacterium avium* complex and *Mycobacterium paratuberculosis.* Microbiology (in press).

**Gilot, P., M. De Kesel, L. Machtelinckx, M. Coene, and C. Cocito.** 1993. Isolation and sequencing of the gene coding for an antigenic 34-kilodalton protein of *Mycobacterium paratuberculosis.* J. Bacteriol. **175:**4930-4935.

**Green, E. P., M. L. V. Tizard, M. T. Moss, J. Thompson, D. J. Winterbourne, J. J. McFadden, and J. Hermon-Taylor.** 1989. Sequence and charateristics of IS*900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis.* Nucleic Acids Res. **17:**9063-9073.

**Hance, A. J., B. Grandchamp, V. Levy-Frebault, D. Lecossier, J. Rauzier, D. Bocart, and B. Gicquel.** 1989. Detection and identification of mycobacteria by amplification of mycobacterial DNA. Mol. Microbiol. **3:**843-849.

**Hawley, D. K. and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

**Hines, M. E. II, A. R. Long, T. G. Snider III, and S. A. Barker.** 1991. Lysis and fractionation of *Mycobacterium paratuberculosis* and *Escherichia coli* by matrix solid-phase dispersion. Anal. Biochem. **195**:197-206.

**Hurley, S. S., G. A. Splitter, and R. A. Welch.** 1987. Rapid lysis technique for mycobacterial species. J. Clin. Microbiol. **25**:2227-2229.

**Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389-409.

**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.

**Kroll, J. S., P. R. Langford, and B. M. Loynds.** 1991. Copper-zinc superoxide dismutase of *Haemophilus influenzae* and *H. parainfluenzae*. J. Bacteriol. **173**:7449-7457.

**McFadden, J. J., P. D. Butcher, R. J. Chiodini, and J. Hermon-Taylor.** 1987. Determination of genome size and DNA homology between unclassified Mycobacterium species isolated from patients with Crohn's disease and other mycobacteria. J. Gen. Microbiol. **133**:211-214.

**McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283-11291.

**Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

**Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186:**339-346.

**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 *Echerichia 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.

**Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. U.S.A. **84:**166-169.

**O'Neill, M. C.** 1989. *Escherichia coli* promoters I. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. J. Biol. Chem. **264:**5522-5530.

**Saier, M. H., Jr.** 1995. Differential codon usage: a safeguard against inappropriate expression of specialized genes? FEBS lett. **362:**1-4.

**Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

**Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71:**1342-1346.

**Sela, S., and J. Clark-Curtiss.** 1991. Cloning and characterization of the *Mycobacterium leprae* ribosomal RNA promoter in *Escherichia coli.* Gene **98:**123-127.

**Stevenson, K., N. F. Inglis, B. Rae, W. Donachie, and J. M. Sharp.** 1991. Complete nucleotide sequence of a gene encoding the 70 kd heat shock protein of *Mycobacterium paratuberculosis.* Nucleic Acids Res. **19:**4552.

**Stover, K., G. P. Bansal, M. S. Hanson, J. E. Burlein, S. R. Palazinski, J. F. Young, S. Koenig, D. B. Young, A. Sadziene, and A. G. Barbour.** 1993. Protective immunity elicited by recombinant bacille-Calmette-Guérin (BCG) expressing outer surface protein A (Osp A) lipoprotein: a candidate lyme disease vaccine. J. Exp. Med. **178:**197-209.

**Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs, Jr., and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. Nature (London) **351:**456-460.

**Tartof, K. D. and C. A. Hobbs.** 1987. Improved media for growing plasmid and cosmid clones. Bethesda Res. Lab. Focus **9:**12.

**Thomas, T. J., R. E. Andrews, Jr., and C. O. Thoen.** 1992. Molecular cloning and characterization of *Mycobacterium paratuberculosis* promoters in *Escherichia coli.* Vet. Microbiol. **32:**351-362.

**Timm, J. M., G. Perilli, C. Uez, J. Trias, G. Orefici, L. Fattorini, G. Amicosante, A. Oratore, B. Joris, J. M. Fr`ere, 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.

**Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. **246:**40-41.

**Whipple, D. L., R. B. Le Febvre, R. E. Andrews, Jr., and A. B. Thiermann.** 1987. Isolation and analysis of restriction endonuclease digestive patterns of chromosomal DNA from *Mycobacterium paratuberculosis* and other *Mycobacterium* species. J. Clin. Microbiol. **25:**1511-1515.

# GENERAL CONCLUSIONS

The ability to genetically manipulate the mycobacteria, some of the first microbial pathogens identified, is much less than a decade old. The new tools that make possible the study of gene expression in members of this genus will ultimately enhance our understanding of the molecular basis of the pathogenicity of *M. paratuberculosis.*

The most important developments in the molecular analysis of *M. paratuberculosis* have been the characterization of restriction site length polymorphisms (Whipple *et al.*, 1989), the characterization and sequence determination of the insertion sequence IS*900* (Green *et al.*, 1989), and the studies on *M. paratuberculosis* taxonomy which indicate a very close relationship with *Mycobacterium avium* (Thorel *et al.*, 1990). New advances in mycobacterial molecular genetics in the last few years (Jacobs *et al.*, 1991; Barletta *et al.*, 1990; Cirillo *et al.*, 1991) have opened up the possibility of applying new technologies to the direct manipulation of *M. paratuberculosis.* This may result in new and more effective vaccines, a more comprehensive understanding of the pathogenesis of *M. paratuberculosis,* and the development of new diagnostic tools.

This work has focused on the cloning and characterization of gene expression signals from *M. paratuberculosis.* We have demonstrated that pJJ2, a promoter selection vector, can be used to clone bacterial promoters in *E. coli.* To test this vector, several novel promoters from *M. paratuberculosis* were cloned. One of these

transcriptionally active constructs (pTB28) was sequenced and shown to contain potential -35 and -10 promoter regions that closely resemble the consensus promoter sequence established for mycobacteria in this study. In addition, a previously characterized *M. paratuberculosis* promoter sequence was cloned into pJJ2 and shown to express the *galK* reporter gene.. This enabled a direct comparison of promoter selection abilities in pJJ2 versus the parent vector pKO1 in *E. coli.*

Nine novel *M. paratuberculosis* promoter-containing fragments were identified and sequenced from a DNA library cloned in pYUB76. Expression of the reporter gene *lacZ* was measured transcriptionally by slot blot hybridization analysis with steady state RNA from *M. smegmatis*, and translationally by β-galactosidase assays. The transcription start sites were experimentally determined for eight promoters and their sequences were aligned relative to the start sites. From this alignment, a consensus nucleotide sequence for mycobacterial promoters was established.

During these studies, at least two strong *M. paratuberculosis* promoters (pAJB300 and pAJB303) were isolated when an initial selection for promoter activity was performed in *M. smegmatis.* These same constructs show no β-galactosidase activity in *E. coli.* The strength and specificity of these promoters make them excellent potential candidates for production of recombinant antigens in BCG. These data provide the impetus for several new studies of gene expression in mycobacteria.

# LITERATURE CITED

**Abbas, B., H. Riemann, and D. E. Behymer.** 1983. Evaluation of the fluorescent antibody test for the diagnosis of paratuberculosis. Am. J. Vet. Res. **44:**720-721.

**Aldovini, A., and R. A. Young.** 1991. Humoral and cell-mediated immune responses to live recombinant BCG-HIV vaccines. Nature (London) **351:**479-482.

**Ambrosio, R. E., Y. Harris, and H. F. A. K. Huchzermeyer.** 1991. A DNA probe for the detection of *Mycobacterium paratuberculosis.* Vet. Microbiol. **26:**87-93.

**Barletta, R. G., B. Snapper, J. D. Cirillo, N. D. Connell, D. D. Kim, W. R. Jacobs, and B. R. Bloom.** 1990. Recombinant BCG as a candidate oral vaccine vector. Res. Microbiol. **141:**931-939.

**Barletta, R. G., D. D. Kim, S. B. Snapper, B. R. Bloom, and W. R. Jacobs, Jr.** 1992. Identification of expression signals of the mycobacteriophages Bxb1, L1 and TM4 using the *Escherichia-Mycobacterium* shuttle plasmids pYUB75 and pYUB76 designed to create translational fusions to the *lacZ* gene. J. Gen. Microbiol. **138:**23-30.

**Bashyam, M. D. and A. K. Tyagi.** 1994. An efficient and high yielding method for isolation of RNA from mycobacteria. BioTechniques **17:**834-836.

**Bercovier, H., O. Kafri, and S. Sela.** 1986. Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. Biochem. Biophys. Res. Comm. **136:**1136-1141.

**Billman, J., M. Carrigan, F. Cockram, L. Corner, I. Gill, J. Hill, T. Jessep, A. Milner, and P. Wood.** 1992. A comparison of the interferon gamma assay with the absorbed ELISA for the diagnosis of Johne's disease in cattle. Aust. Vet. J. **69:**25-28.

**Bloom, B. R., and P. E. M. Fine.** 1994. The BCG experience: implications for future vaccines against tuberculosis. *In* B. R. Bloom (ed.), Tuberculosis: pathogenesis, protection, and control. American Society for Microbiology, Washington, D. C., pp. 531-557.

**Burlein, J. E., C. K. Stover, S. Offutt, and M. S. Hanson.** 1994. Expression of foreign genes in mycobacteria. *In* B. R. Bloom (ed.), Tuberculosis: pathogenesis, protection, and control. American Society for Microbiology, Washington, D. C., pp. 239-252.

**Calmette, A., B. Weil-Halle, A. Saenz, and L. Costil.** 1933. Demonstration experimentale du passage des bacilles vaccins BCG a travers la muqueuse intestinale chez l' enfant et chez le singe. Bull. Acad. Med. **110:**203-206.

**Camphausen, R. T., R. L. Jones, and P. J. Brennan.** 1985. A glycolipid antigen specific to *Mycobacterium paratuberculosis*: structure and antigenicity. Proc. Natl. Acad. Sci. U.S.A. **82:**3068-3072.

**Chan, J., T. Fujiwara, P. Brennan, M. McNeil, S. J. Turco, J.-C. Sibille, M. Snapper, P. Aisen, and B. R. Bloom.** 1989. Microbial glycolipids, possible virulence factors that scavenge oxygen radicals. Proc. Natl. Acad. Sci. U.S.A. **86:**2453-2457.

**Chiodini, R. J., H. J. Van Kruiningen, and R. S. Merkal.** 1984. Ruminant paratuberculosis (Johne's disease): The current status and future prospects. Cornell Vet. **74:**218-262.

**Christensen, D.** 1993. Another look at the TB vaccine: should the United States use BCG widely? Science News **145:**393.

**Cirillo, J. D., R. G. Barletta, B. R. Bloom, and W. R. Jacobs.** 1991. A novel transposon trap for mycobacteria: isolation and characterization of IS*1096*. J. Bacteriol. **173:**7772-7780.

**Clark-Curtiss, J. E., W. R. Jacobs, M. A. Doherty, L. R. Ritchie, and R. Curtiss, III.** 1985. Molecular analysis of DNA and

construction of genomic libraries of *M. leprae*. J. Bacteriol. **161**:1093-1102.

**Cocito, C., P. Gilot, M. Coene, M. De Kesel, P. Poupart, and P. Vannuffel.** 1994. Paratuberculosis. Clin. Microbiol. Rev. **7**:328-345.

**Cole, S. T., and D. R. Smith.** 1994. Toward mapping and sequencing the genome of Mycobacterium tuberculosis. *In* B. R. Bloom (ed.), Tuberculosis: Pathogenesis, protection, and control. ASM press, Washington, D. C., pp. 227-238.

**Das Gupta, 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.

**Dellagostin, O. A., S. Wall, E. Norman, T. O'Shaughnessy, J. W. Dale, and J. McFadden.** 1993. Construction and use of integrative vectors to express foreign genes in mycobacteria. Mol. Microbiol. **10**:983-993.

**Fine, P. E. M.** 1988. BCG vaccination against tuberculosis and leprosy. British Med. Bull. **44**:691-703.

**Foley-Thomas, E., D. L. Whipple, L. E. Bermudez, and R. G. Barletta.** 1995. Phage infection, transfection, and transformation of *Mycobacterium avium* complex and *Mycobacterium paratuberculosis*. Microbiology (in press).

**Gilot, P., M. De Kesel, L. Machtelinckx, M. Coene, and C. Cocito.** 1993. Isolation and sequencing of the gene coding for an antigenic 34-kilodalton protein of *Mycobacterium paratuberculosis*. J. Bacteriol. **175**:4930-4935.

**Green, E. P., M. L. V. Tizard, M. T. Moss, J. Thompson, D. J. Winterbourne, J. J. McFadden, and J. Hermon-Taylor.** 1989. Sequence and charateristics of IS*900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. Nucleic Acids Res. **17**:9063-9073.

**Hance, A. J., B. Grandchamp, V. Levy-Frebault, D. Lecossier, J. Rauzier, D. Bocart, and B. Gicquel.** 1989. Detection and identification of mycobacteria by amplification of mycobacterial DNA. Mol. Microbiol. **3**:843-849.

**Hatfull, G. F.** 1995. Mycobacteriophage L5: A toolbox for tuberculosis. ASM News **60**:255-260.

**Hawley, D. K. and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

**Hill, E. B., L. G Wayne, and W. M. Gross.** 1972. Purification of mycobacterial DNA. J. Bacteriol. **112**:1033-1039.

**Hines, M. E. II, A. R. Long, T. G. Snider III, and S. A. Barker.** 1991. Lysis and fractionation of *Mycobacterium paratuberculosis* and *Escherichia coli* by matrix solid-phase dispersion. Anal. Biochem. **195**:197-206.

**Hinshelwood, S. and N. G. Stoker.** 1992. An *Escherichia coli-Mycobacterium* shuttle cosmid vector, pMSC1. Gene **110**:115-118.

**Horn, G. T., and R. D. Wells.** 1981a. The leftward promoter of bacteriophage λ. Structure, biological activity, and influence by adjacent regions. J. Biol. Chem. **256**:2003-2009.

**Horn, G. T., and R. D. Wells.** 1981b. The leftward promoter of bacteriophage λ. Isolation on a small restriction fragment and deletion of adjacent regions. J. Biol. Chem. **256**:1998-2002.

**Hurley, S. S., G. A. Splitter, and R. A. Welch.** 1987. Rapid lysis technique for mycobacterial species. J. Clin. Microbiol. **25**:2227-2229.

**Hurley, S. S., G. A. Splitter, and R. A. Welch.** 1989. Development of a diagnostic test for Johne's disease using a DNA hybridization probe. J. Clin. Microbiol. **27**:1582-1587.

**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.

**Jöhne, H. A., and L. Frothingham.** 1895. Ein eigenthumlicher Fall von Tuberculose beim Rind. Deut. Z. Tiermed. vergl. Path. **21**:438-454.

**Jørgensen, J. B.** 1982. An improved medium for culture of *Mycobacterium paratuberculosis* from bovine faeces. Acta. vet. scand. **23**:325-335.

**Kenney, T. J., and C. P. Moran, Jr.** 1991. Genetic evidence for interaction of $\sigma^A$ with two promoters in *Bacillus subtilis*. J. Bacteriol. **173**:3282-3290.

**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.

**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.

**Kroos, L., B. Kunkel, R. Losick.** 1989. Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science **243**:526-529.

**Kunze, Z. M., F. Portaels, and J. J. McFadden.** 1992. Biologically distinct subtypes of *Mycobacterium avium* differ in possession of insertion sequence IS*901*. J. Clin. Microbiol. **30**:2366-2372.

**Langermann, S., S. Palaszynski, A. Sadziene, C. K. Stover, and S. Koenig.** 1994. Systemic and mucosal immunity induced by BCG vector expressing outer-surface protein A of *Borrelia burgdorferi*. Nature (London) **372**:552-555.

**Lee, M. H., and G. F. Hatfull.** 1993. Mycobacteriophage L5 integrase-mediated site specific integration *in vitro*. J. Bacteriol. **175:**6836-6841.

**Losick, R., and J. Pero.** 1981. Cascades of sigma factors. Cell **25:**582-584.

**Mahan, M. J., J. M. Slauch, J. J. Mekalanos.** 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. Science **259:**686-688.

**Mahan, M. J., J. W. Tobias, J. M. Slauch, P. C. Hanna, R. J. Collier, and J. J. Mekalanos.** 1995. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. Proc. Natl. Acad. Sci. U.S.A. **92:**669-673.

**Martin, C., J. Timm, J. Rauzier, R. Gomez-Lus, J. Davies, and B. Gicquel.** 1990. Transposition of an antibiotic resistance element in mycobacteria. Nature (London) **345:**739-743.

**McAdam, R. A., C. Guilhot, and B. Gicquel.** 1994. Transposition in mycobacteria. *In* B. R. Bloom (ed.), Tuberculosis: Pathogenesis, protection, and control. ASM press, Washington, D. C., pp. 199-216.

**McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54:**171-204.

**McFadden, J. J., P. D. Butcher, R. J. Chiodini, and J. Hermon-Taylor.** 1987. Determination of genome size and DNA homology between unclassified Mycobacterium species isolated from patients with Crohn's disease and other mycobacteria. J. Gen. Microbiol. **133:**211-214.

**McFadden, J. J., J. Collins, B. Beaman, M. Arthur, and G. Gitnick.** 1992. Mycobacteria in Crohn's disease: DNA probes identify the Wood Pigeon strain of *Mycobacterium avium* and

*Mycobacterium paratuberculosis* from human tissue. J. Clin. Microbiol. **30:**3070-3073.

**Mekalanos, J. J.** 1992. Environmental signals controlling expression of virulence determinants in bacteria. J. Bacteriol. **174:**1-7.

**Merkal, R. S., and B. J. Curran.** 1974. Growth and metabolic characteristics of *Mycobacterium paratuberculosis.* Appl. Microbiol. **28:**276-279.

**Merkal, R. S., and W. G. McCullough.** 1982. A new mycobactin, mycobactin J, from *Mycobacterium paratuberculosis.* Current Microbiol. **7:**333-335.

**Merkal, R. S.** 1984. Paratuberculosis. *In* G. P. Kubica and L. G Wayne (eds.), The Mycobacteria: A Sourcebook. Marcel Dekker, Inc. New York, N. Y., pp. 1237-1249.

**Merkal, R. S., D. L. Whipple, J. M. Sacks, and G. R. Snyder.** 1987. Prevalence of *Mycobacterium paratuberculosis* in ileocecal lymph nodes of cattle culled in the United States. JAVMA **190:**676-680.

**Molina, A., L. Morera, and D. Llanes.** 1991. Enzyme-linked immunosorbent assay for detection of antibodies against *Mycobacterium paratuberculosis* in goats. Am. J. Vet. Res. **52:**863-868.

**Momotani, E., D. Whipple, A. Thiermann, and N. Cheville.** 1988. Role of M cells and macrophages in the entrance of *Mycobacterium paratuberculosis* into domes of ileal Peyer's patches in calves. Vet. Pathol. **25:**131-137.

**Moran, C. P., Jr.** 1989. Sigma factors and the regulation of transcription. *In* I. Smith, R. Slepecky, and P. Setlow (eds.), Regulation of procaryotic development. American Society for Microbiology, Washington, D. C., pp. 167-184.

**Moss, M. T., E. P. Green, M. L. Tizard, Z. P. Malik, and J. H. Taylor.** 1991. Specific detection of *Mycobacterium paratuberculosis* by DNA hybridisation with a fragment of the insertion element IS900. Gut **32:**395-398.

**Mulligan, M. E., D. K. Hawley, R. Entriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res. **12:**789-800.

**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 *Echerichia 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.

**NcNab, B. W., A. H. Meek, J. R. Duncan, S. W. Martin, and A. A. Van Dreumel.** 1991. An epidemiological study of paratuberculosis in dairy cattle in Ontario: study design and prevalence estimates. Can. J. Vet. Res. **55:**246-251.

**O'Neill, M. C.** 1989. *Escherichia coli* promoters I. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. J. Biol. Chem. **264:**5522-5530.

**O'Neill, M. C. and F. Chiafari.** 1989. *Escherichia coli* promoters II. A spacing class-dependent promoter search protocol. J. Biol. Chem. **264:**5531-5534.

**Patel, B. K. R., D. K. Banerjee, and P. D. Butcher.** 1991. Extraction and characterization of mRNA from mycobacteria: implication for virulence gene identification. J. Microbiol. Methods **13:**99-111.

**Patel, R. J., J. T. Kvach, and P. Mounts.** 1986. Isolation and restriction endonuclease analysis of mycobacterial DNA. J. Gen. Microbiol. **132:**541-551.

**Poupart, P., M. Coene, H. Van Heuverswyn, and C. Cocito.** 1993. Preparation of a specific RNA probe for detection of *Mycobacterium paratuberculosis* and diagnosis of Johne's disease. J. Clin. Microbiol. **31:**1601-1605.

**Qin, M., H. Taniguchi, and Y Mizuguchi.** 1994. Analysis of the replication region of a mycobacterial plasmid, pMSC262. J. Bacteriol. **176:**419-425.

**Rather, P. N., R. E. Hay, G. L. Ray, W. G. Haldenwang, and C. P. Moran, Jr.** 1986. Nucleotide sequences that define promoters that are used by *Bacillus subtilis* sigma-29 RNA polymerase. J. Mol. Biol. **192:**557-565.

**Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Gross.** 1985. The regulation of transcription initiation in bacteria. Ann. Rev. Genet. **19:**355-387.

**Rosenberg, M. and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13:**319-353.

**Saier, M. H., Jr.** 1995. Differential codon usage: a safeguard against inappropriate expression of specialized genes? FEBS lett. **362:**1-4.

**Satola, S., P. Kirchman, and C. P. Moran, Jr.** 1991. Spo0A binds to a promoter used by σA RNA polymerase during sporulation in *Bacillus subtilis.* Proc. Natl. Acad. Sci. U.S.A. **88:**4533-4537.

**Savioz, A., A. Zimmermann, D. Haas.** 1993. *Pseudomonas aeruginosa* promoters which contain a conserved GG-$N_{10}$-GC motif but appear to be RpoN-independent. Mol. Gen. Genet. **238:**74-80.

**Sela, S., and J. Clark-Curtiss.** 1991. Cloning and characterization of the *Mycobacterium leprae* ribosomal RNA promoter in *Escherichia coli.* Gene **98:**123-127.

**Sirakova, T. D., S. S. Bardarov, J. I. Kriakov, and K. I. Markov.** 1989. Molecular cloning of mycobacterial promoters in *E. coli.* FEMS Microbiol. Lett. **59:**153-156.

**Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis.* Mol. Microbiol. **4:**1911-1919.

**Staden, R.** 1984. Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Res. **12:**505-519.

**Stevenson, K., N. F. Inglis, B. Rae, W. Donachie, and J. M. Sharp.** 1991. Complete nucleotide sequence of a gene encoding the 70 kd heat shock protein of *Mycobacterium paratuberculosis.* Nucleic Acids Res. **19:**4552.

**Stover, K., G. P. Bansal, M. S. Hanson, J. E. Burlein, S. R. Palazinski, J. F. Young, S. Koenig, D. B. Young, A. Sadziene, and A. G. Barbour.** 1993. Protective immunity elicited by recombinant bacille-Calmette-Guérin (BCG) expressing outer surface protein A (Osp A) lipoprotein: a candidate lyme disease vaccine. J. Exp. Med. **178:**197-209.

**Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs, Jr., and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. Nature (London) **351:**456-460.

**Sugden, E., B. Brooks, N. Young, D. Watson, K. Nielsen, A. Corner, C. Turcotte, A. Michaelides, and R. Stewart.** 1991. Chromatographic purification and characterization of antigens A and D from *Mycobacterium paratuberculosis* and their use in enzyme-linked immunosorbent assays for diagnosis of paratuberculosis in sheep. J. Clin. Microbiol. **29:**1659-1664.

**Talkington, C., and J. Pero.** 1979. Distinctive nucleotide sequences of promoters recognized by RNA polymerase

containing a phage-coded "σ-like" protein. Proc. Natl. Acad. Sci. U.S.A. **76**:5465-5469.

**Thoen, C. O. and K. H. Baum.** 1988. Current knowledge on paratuberculosis. JAVMA **192**:1609-1611.

**Thole, J. E. R., H. G. Dauwerse, P. K. Das, D. G. Groothjuis, L. M. Schouls, and J. D. A. van Embden.** 1985. Cloning of *Mycobacterium bovis* BCG DNA and expression of antigens in *Escherichia coli.* Infect. Immun. **50**:800-806.

**Thole, J. E. R., W. J. Keulen, A. H. J. Kolk, D. G. Groothuis, L. G. Berwald, R. H. Tiesjema, and J. D. A. van Embden.** 1987. Characterization, sequence determination, and immunogenicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli.* Infect. Immun. **55**:1466-1475.

**Thomas, T. J., R. E. Andrews, Jr., and C. O. Thoen.** 1992. Molecular cloning and characterization of *Mycobacterium paratuberculosis* promoters in *Escherichia coli.* Vet. Microbiol. **32**:351-362.

**Thorel, M. F., M. Krichevsky, and V. V. Levy-Frebault.** 1990. Numerical taxonomy of mycobactin-dependent mycobacteria, amended description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. Int. J. of Syst. Bacteriol. **40**:254-260.

**Thoresen, O. F., and F. Saxegaard.** 1991. Gen-probe rapid diagnostic system for the *Mycobacterium avium* complex does not distinguish between *Mycobacterium avium* and *Mycobacterium paratuberculosis.* J. Clin. Microbiol. **29**:625-626.

**Timm, J. M., G. Perilli, C. Uez, J. Trias, G. Orefici, L. Fattorini, G. Amicosante, A. Oratore, B. Joris, J. M. Fr`ere, 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.

**Vary, P. H., P. R. Andersen, E. Green, J. Hermon-Taylor, and J. J. McFadden.** 1990. Use of highly specific DNA probes and the polymerase chain reaction to detect *Mycobacterium paratuberculosis* in Johne's disease. J. Clin. Microbiol. **28:**933-937.

**Wayne, L. G., and G. P. Kubica.** 1986. The mycobacteria. *In* P. H. A. Sneath and J. G. Holt (eds.), Bergey's Manual of Systematic Bacteriology, vol. 2. The Williams and Williams Co., Baltimore, pp.1435-1457.

**Whipple, D. L., R. B. Le Febvre, R. E. Andrews, Jr., and A. B. Thiermann.** 1987. Isolation and analysis of restriction endonuclease digestive patterns of chromosomal DNA from *Mycobacterium paratuberculosis* and other *Mycobacterium* species. J. Clin. Microbiol. **25:**1511-1515.

**Whipple, D. L., P. A. Kapke, and R. E. Andrews, Jr.** 1989. Analysis of restriction endonuclease fragment patterns of DNA from *Mycobacterium paratuberculosis.* Vet. Microbiol. **19:**189-194.

**Young, R. A., B. R. Bloom, C. M. Grosskinsky, J. Ivanyi, D. Thomas, and R. W. Davis.** 1985. Dissection of *M. tuberculosis* antigens using recombinant DNA. Proc. Natl. Acad.Sci. U.S.A. **82:**2583-2587.

## ACKNOWLEDGEMENTS

Automated DNA sequencing was performed by the DNA sequencing and synthesis facility, 1184 Molecular Biology building, Iowa State University.

I thank my major professor Bob Andrews for giving me the space I needed to grow as a scientist. His mentoring style was a welcome relief from the rigid instruction of Peter Pattee. I also wish to acknowledge Greg Phillips, Alan DiSpirito, and Susan Carpenter for their advice and assistance in obtaining a postdoctoral position.

Most importantly, I wish to thank my wife Joyce who put up with all my training, research, and teaching; the three things that really soaked up my time away from her.

Finally, I wish to acknowledge Jesus Christ who sacrificed himself for us all.

# APPENDIX A: SEQUENCES UPSTREAM OF *LACZ* IN THE pYUB76::*lacZ* FUSIONS

GATCTGGCTGAAGGGTCCGCTGGTGCTCGGGCTGGGCGCCATCCAGCTGGTGTTGCTGGCCCT

GACCGTGGTGATCTCGACGCCGGTGGACCAGATCGAGTTCAAGACCCT**GACGC**TGGCCGACC
-35

+1
TCGGCGCG**CAGCCG**ACCGCGC<u>A</u>GCGGTGCACGGTCGCGATCTGGCGGCGCTGCGGGACCGGG
-10

TCACCCGGCTGGCCAGGGATTTCCCGCTGTACGAGGGCCTCGAAGACTGGGCGCTGGTCGGCC

***
GGTAGCCGGCCGCTGGGGACCAAAGTCACCACTTTGCCCGCGGCGGCGCGGACGATTTCACG

AACCTGTG**TACGCAAGTTA**CAGTTATCTCC <u>ATG</u> GCC GAA AGA CCT GTC GCT AAT ACG
RBS Met Ala Glu Arg Pro Val Ala Asn Thr

CTG ACC CTG GAA CTC GAG CCG GTT GTC GAA GCC AAC ATG GAT CCC GTC GTT TTA
Leu Thr Leu Glu Leu Glu Pro Val Val Glu Ala Asn Met Asp Pro <u>Val</u> <u>Val</u> <u>Leu</u>

CAA CGT CGT GAC
<u>Gln</u> <u>Arg</u> <u>Arg</u> <u>Asp</u>

Figure 1A. Nucleotide sequence upstream of *lacZ* in pAJB300. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide. A stop codon is indicated by three "*" above the sequence.

```
                                                                    +1
GATCTGGCGGGCGGTCCAGTACACCGCGAGTTCGCGCACGCTGGCCGGCAGCGTCTTGGAC
               -35                              -10

GCCCGGTCGACCAGGGTTTCATGCACGGCGCTGGCGACCATCAGCCTGCG GTG ATG CGC
                         RBS                          Val Met Arg

GGC GAC GGC GAA CGG TTC GGG AAC CTG GCC GTA GCG CCG CTT GGC GAT CGA GCG
Gly Asp Gly Glu Arg Phe Gly Asn Leu Ala Val Ala Pro Leu Gly Asp Arg Ala

GCT GGG CAA CTT CCT GGT CGG CTA TCG ACA GTA AGC ACC CAG CCG GGC GCC TCT
Ala Gly Gln Leu Pro Gly Arg Leu Ser Thr Val Ser Thr Gln Pro Gly Ala Ser

ACG GTG GAG CGC GTG ACG CAC TCC CAC TCC CAC GGC CTG CCG TCG GGC CCG GCC
Thr Val Glu Arg Val Thr His Ser His Ser His Gly Leu Pro Ser Gly Pro Ala

CCC GTC GAT CCC GTC GTT TTA CAA CGT CGT GAC
Pro Val Asp Pro Val Val Leu Gln Arg Arg Asp
```

Figure 2A. Nucleotide sequence upstream of *lacZ* in pAJB301. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide.

GATCAACACCGTGACCAGCGGGCCGAGCTGGGTGACCGCACCGAACGCGGCGCCCGCGCCGG

ACAGATCGGCACCAACTGCCTGGCCGACGGCAGCAGCGCCATCGCCACCGCGCTCTCGGTGG

CGGGGCCGGCCAAGATCTCGGAGTTGCCCTGCCCCAAGCCAAGTGCGTTCGTGATCTGGTCGA

+1

CGACGAGGGCGG**TGGCGT**CGCCGGTGTAGCCGAA**CGGCAC**GTGCGCG<u>T</u>AGGCCCAGATCGG

-35 -10

TGCGGTCATCGCGGNCACTGTCGTCACGGCAACTCCTGGGCTGGGCGGGCTTGCGGAA**TCACC**

***

**GAAAAGAT**AA<u>GTG</u>AAA <u>ATG</u> ATT ATC GTT ATC AGC AGC GAC CGG GAG GCT CAT
RBS Met Ile Ile Val Ile Ser Ser Asp Arg Glu Ala His

ATC GAA GGC ATA GAA TTC GCT CGG TGC CGC GCT TTC GCT GCG GGC CTC GGG GGC
Ile Glu Gly Ile Glu Phe Ala Arg Cys Arg Ala Phe Ala Ala Gly Leu Gly Gly

CGG GGC CGG CGG AGG CGG CCA CTA GAC TTC TGG CGT GGC CGA ATC CCC CCC CAC
Arg Gly Arg Arg Arg Arg Pro Leu Asp Phe Trp Arg Gly Arg Ile Pro Pro His

CGA GGT GCC CGA ACC GCC CGC TCG ACC CGT GTT GGT GGT CGA CTT CGG TGC CCA
Arg Gly Ala Arg Thr Ala Arg Ser Thr Arg Val Gly Gly Arg Leu Arg Cys Phe

GTA CGC CGA GCT GAT CGC GCC GGA CTG CGG CAG GAA CAC CAA CGC CTG CTC GAC
Val Arg Arg Ala Asp Arg Ala Gly Leu Arg Gln Glu His Gln Arg Leu Leu Asp

GCC GTT GTG ATA GCG CTT GAG CAT GTC GAT CCC GTC GTT TTA CAA CGT
Ala Val Val Ile Ala Leu Glu His Val Asp Pro <u>Val Val Leu Gln Arg</u>

Figure 3A. Nucleotide sequence upstream of *lacZ* in pAJB303. The *lacZ* and potential in frame initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide. A stop codon is indicated by three "*" above the sequence.

GATCCGACGAACCGTGCACGATCATAGCTGTGGTCCTCGCGGCGGGTCGGTTTGTCCGAACGG

CCGAACCCGACCAGGGTATCAGCGTGGCGAGCGGGTCGACGGTGTGGCCGCCGGCGACGGCC

AGCGCGGTGGACCCGTCGACGATGCGGGACAACTCGGGGCCCAGTCCGCCGACCGCCCGGTTG

GTCTCGTCGATCTGGACCGGATGGTCAATCAACGGCAGTTCGTCGCGGCGCTGCTGCACCGTGC

GGNCAGCCCCACCGTCTGGCTCAACCCGTGGCGCTGGTACGCGGCGCCGCGCGCGGTCGCCGAC

GCGCTGACCGTGGACCGCGCCGACCACGCCTGGGACCTGGCCCGGCTCGGCTGGGCGCTNNAC

GGCTCCCCCGCCACGGTGACGGTCCCGATCGGGCCGCTGTGCAGACCGCTTTGCGTCGTTAGAC

GCTTATCCGGGGGTGGTGTGGGCATACTGACTTACATGCGCTCCATCTGGAAGGGTTCCATCG

CGTTCGGGCTGGTGAACGTCCCGGTCAAGGTGTACAGCGCGACCGAAGACCACGACATCAAG

+1

TTCCACCAGGTACACGCC**AAGGAC**AACGGCCGTATCCGGTA**CCAACG**GGTGTGC<u>G</u>AGCTG

-35 -10

GACGG**GGAGGT**CGTCG <u>ATG</u> ATC GAG GCC CTG GAT CCC GTC GTT TTA CAA

RBS Met Ile Glu Ala Leu Asp Pro <u>Val</u> <u>Val</u> <u>Leu</u> <u>Gln</u>

Figure 4A. Nucleotide sequence upstream of *lacZ* in pAJB304. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide.

GATCCGCAGGACTGCGTCTGCGCGGCGGTCGCCAAGCGGCGCTACCTGGGCAAGCTGTCCGGG

CCGCTGCTGGACCGGGTGGACCTGCGGGTGCAGATGCATCCGATGCGGGCCGGCGCGTTCGGC

GGCGGCGACGGCGAATCCACCGCGCAGGTCCGCCGGCGGGTCGCGCAGGCGCGGCAGGCCGCG

GCGCAGCGCTGGCTGCCGTACGGNTTTCGCACCAACGCCGAGGTGAGCGGGCCGCTGCTGCGC

CGCAAGTTCCGGCTGGGCAGCGCGGCGATGGATCTGTTCGAGGTGCGGCCCGACGGGCAGGAC

CGGTTCACCGCCGACACCGCCGGTGATATCGGGGCCGGGGGCGACGAACGCCAGGTGGTCGA

GGGCACCCAGGTGCTCGCGCAGGCGATCCGGACGGGCAGT**TGTTGG**AGTTTCTGTCGGACGG
-35

+1
**TTGGTTG**GCGGCATTTCCGGCGAGGCCGTATTCGGTCGGCCGGATGCGGGCGACAAGGCAA
-10

***
TGTTGTCGGTACC**GGATGAGAT**GGTCTCTTC ATG GAG ACG GCG CTC GGT CGA AGC GAT
RBS Met Glu Thr Ala Leu Gly Arg Ser Asp

CCC GTC GTT TTA CAA CGT
Pro Val Val Leu Gln Arg

Figure 5A. Nucleotide sequence upstream of *lacZ* in pAJB305. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide. A stop codon is indicated by three "*" above the sequence.

```
                                                          +1
GATCGGTGTGCCGCTTGAACCGGCCCAGCTCCCGCTCCAGGGTGACGTGCTCGAGCTCGGT
          -35                              -10                RBS

GTG CGC GGT GAT GAG CAC CAT GGC GCC GGG CGC CTC GTA GAT CCC GTC GTT TTA
Val Arg Gly Asp Glu His His Gly Ala Gly Arg Leu Val Asp Pro Val Val Leu

CAA
Gln
```

Figure 6A. Nucleotide sequence upstream of *lacZ* in pAJB73. The *lacZ* and initiation codons are underlined. The potential RBS, -35 and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide.

GATCAGGTTGCGGGTGGTTTCGTTGCCGGCCACCAGCAGCAGAAAGGCGAAGTTGAGCAGGT
CTTCGTCGGTCAACCGGTGCTCGTCGATCTCGGCCTCGGCCAGCACCGAGAGCAGGTCGGCGCG
CGGTCGCGCGCGGCGTGCCGCGATCAGCCGCTGGAAGTACTCGTAGAGTTGCCCGGCCGCCACC
GCCGGGTCCAGCTCGATCTCCGGGTCGGCCGTCCCGGTGGCCGCATCCGACCACGCCCGGAATT
GCTCCCAGTCGTCGGGCGGGGCCCCGATCAATTCGGCGATCATCCGGGTGGGCAGGGGAGCGG
TGATCTGTTCGGTGAACTCGTGCACGGCGCCGNGTTCGATGCCGTCGAGAATGNCGTNTACG
ATCTTGCGGATCTTCGGTTCCAGCACCGAAACACGCCGNCNGNTGAATCCGGAGTTGATCAG
CTTGCGTAGCTGNCGGTGNCGNGGCGGATCGGTGAAGATCAGGCTGCCCTGTTGCACCGGGTT
GGNCAGTTGCGGATCCNGNCATCGTNATGCCCCGGNTCGAGGTGAACAGAGNCGGATTGTT
TCGACACGAACCGGATGTCCTCGTATTTCAGCAGCGCCCAGNAGTTGGTGACGTCGTTCCAGC
ACAACGGTTGCCGTTCGNGCGNAGNTTCNCGNNTAGGCGNGNATACCGGGTTTACCCCTCNT
CGGAAGTCCCGGNGNAATGCAAGCTGTAANGCTTCGNCNANCNCAGNCNGCGAGTTTGAN
TCTTTAACGGCAAACCCCCCTCTTCCGGAGGAAAACCCGCAATCCCGACAGAATTGGGTTCA
ATCCTGTTTCTTGCCGCACTTTGTTTTTTCTATCCCTTCGCCCGCCGCGTGCTGTTTTTCTAGACA
GGTGGGCCTAAAGTAGNGGATANCTCGATTTTCACTCGAAGGATGATTGNCGGNCNANAC
GNGGGGGGTATACACCGAGTGATCAAGATGACCGGAGGAANCNAANATGTCCGATCCCAN
CCGCTCTTGTCCAGACCTACCAGCTCTACATCGACGACCAGTGGGTCGAGCCACAGGACGNTC
GCTACGACGANCTGTCTCCGTCCNCCGAGGCCGTCATCGCCACCGCANCCGACGCGAGTGTGG
NNCAGGTCGACGACGCGATACGCGTCCGCACACCGCGCCTTCGACAACGGNCCGTGGGGCCG
GATGTCCGCCGAGCAGCGCGCCCGGTGTCTGAACCAACTGGGCGAGGCGCTCACCAAGCACA
CCGACGACTTCNTCGCGCTGTCCCAAACGGAGTGGGGCTGCATCGGCAACGAGCGCATCATT

+1
**AAAG**ATCGANGGCGCCGGGNT**CATGTC**CCTTCAC<u>C</u>CCGCCCAGCTCGCGACCCAACTGGTC
-35 -10

GACC**AGGAGGT**CGCCGGCATCAGCGCCGGCACCACGTTGTTGCGCCAC <u>GTG</u> CCG CTG GGC
RBS Val Pro Leu Gly

GTG GTA GCG GTG CTC ACA CCC TGG AAC TTG CCG CAC TGC CTC AAC GTC ATG AAG
Val Val Ala Val Leu Thr Pro Trp Asn Leu Pro His Cys Leu Asn Val Met Lys

CTG AAT CAT GCG CTG GCC GCG GGC AAA ACG GTC GTG CTC AAA CCG TCT CCG CTG
Leu Asn His Ala Leu Ala Ala Gly Lys Thr Val Val Leu Lys Pro Ser Pro Leu

ACA CCG CTG GCC GGG CTG GCG CTG GCG CGC CTC ATC GAC GAG CAC ACC GAC ATC
Thr Pro Leu Ala Gly Leu Ala Leu Ala Arg Leu Ile Asp Glu His Thr Asp Ile

CCG CCG GGG GTG GTC AAC GTC GTC ACG CCC TCG GGG GTT GAG GCG GCC AAA CTG
Pro Pro Gly Val Val Asn Val Val Thr Pro Ser Gly Val Glu Ala Ala Lys Leu

CTC ACC ACC GAT CCC GTC GTT TTA CAA
Leu Thr Thr Asp Pro <u>Val</u> <u>Val</u> <u>Leu</u> <u>Gln</u>

Figure 7A. Nucleotide sequence upstream of *lacZ* in pAJB125. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide.

GATCGTACGGTAAATGAGTGGAGGTAAACATGGAAGCGGCTCAGCTCGCGGCAAGAGAGA
TTAGCGACGGCCCGTGGCCGGAGACCGGATTCGGCTGGTACGACCAGGCGACGGGTTGTCGGG
TCACCGCGTCGACGCCGGAAGCGGAGCCCGAATTGTGGAATCAATATCTGGACGGCGCACTC
GAGAGTTATCGCAAGCACGGCCTGGACTGGGTGCTCGACCTGGAGCGGATCAGGGATGGCGC
CGACACGGGCCTGTTCTTTGCCGCGCTTGACCCCCAAGGTCGGGTCGTCGGGGGGACGCGGGTG
GTGGGGCCGTTGCGGTCTCCGGAGGATTCGCACGCCCTGCAGGAGTGGGAAGGAAACCCGGG
GCTTCCGCTGCTGCGCTACATGATTGCCAACCGGATTCCCTTTGGCGTCGTGGAAGTGAAAAG
CGCATGGACGAATTCCCGCGAATACGGCAATCGGGCCCTTTCCACCGTGCTGGCCCGCACCGC
GTTGTCGACGATGACGCTGCTCGGTGCCCAGTTCGTCATGGCCACCGCGGCGGNGCACGTGCTC
GAGCAGTGGCAGTCCTCCGGCGGTGTGGTGGTGTCTCGGGTGCCGNCGGTGGCCTATCCCAAC
GAGAACTTCCGAACCATGGTGATGT.....TAAGGGCCATCAAGGAATACCACACCGGTGCT
AACAAGCTTCTCGATGCGGGCGGAGCCGTCACCCGGGTCACCCTGGGACCCAGGTGGCTGATT
TCACCGATCGTTTTGGCGACGTCCCCGCAGGGCATCCGCGACATCGTCAGCGTGCGGGACGGTT
CGATTGACAAGACGAGCACGGTAGCCACCGAGCTGCGCCGACTGCTCGGTCCGAACCTGTTCG
TCCTGCCCCACACCGAGTGGTTGCCGCGCCGACGCACCCTGCAACCGGTGTTCACCCGTCAGCG
TGTCCGTGAGTTTGGGGGACACATGTCCGAGACCGCGGAGAGCGTGTGCTCCGGGTGGCCCGA
AGACACCGAGATCGACCTGGATGCACAGTGCCGCACACTGACCTTGCGCGCGCTGGGCCGCTC
GGTATTGGGTTTGGACCTCGACGAGCGCTCCGACGCGATCACCGAGCCCCTGCGCGTGGCGAC
GAGCTACGCGGTGCGGCGGGCGCTGCGGCCGCTGCGCGCACCGGAGTGGCTGCCCACACCGTCC
CGGCGGCGCGCCCGCGCCGCGGCCGGCGCCATTCGCGCGCTGGCGGATGAGATCCTGCAGGCGT
GCCGCGCCGACCCCGACCGGGAGGCCCCGTTGGTGGATGCGCTCATCGCGGCCACCGACCCGGA
GACCGGGCAGGCGTTGTCGGACAAGGAAATCCGCGACGAAATGATCATCTTCCTGTTCGCCG
GCCACGACACCACGGCGACGACGCTGACCTACGCGCTGTGGGCGCTGGGCCGCCATCCCGAGT

ATCAGGCGCGGGTCGCGGCCGAAGTGTCCGAGTTACCGGACCGCCACCTTACTCCCGA**TGAC**
-35

+1
GTTGCACGGCTGGGATTAA**CGGTCC**GCGTGCTCCAGGAGACACTACGACTGTGTCCGCCCG
-10

GGCCGACCGGAACCCGGATGGCCACCCACGATGTCGAAGTCGCCGGCTACCGCGTCG**AAGCCG**

***
**GAACCATGA**CTGGCGTTCGGCAGA ATG GCG GTA CAA ACC GAT CCG AGC TTG TGG
RBS Met Ala Val Gln Thr Asp Pro Ser Leu Trp

GAC GCC CCT CTG CGC GGC GAC CCG GAT CCC GTC GTT TTA CAA
Asp Ala Pro Leu Arg Gly Asp Pro Asp Pro Val Val Leu Gln

Figure 8A. Nucleotide sequence upstream of *lacZ* in pAJB86. The *lacZ* and initiation codons are underlined. The potential ribosome binding site (RBS), -35, and -10 regions are indicated beneath the bolded sequence. The transcription start site is indicated by a +1 above the underlined nucleotide. A stop codon is indicated by three "*" above the sequence.

TAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCCCGGGGATCCCCGGGTACCGAGCTCG
AATTCACTGGCCGCCTGCATCACCGGCGTCATCTGCGCGCTCATCGGTCAGCCTGGCGGCTCCG
GCGGGCTTGCGGGTCGAAGCGCAGCACCAGCCCGAAGATCAACACGATCAATCCGAACAGC
GTGCCGAGCATGATGACAGCGCCCACGGCCGCACCCCTTCCGATCCGAGCTCGACTCATTAGA
TAGTGCACACTATCCTAAAGTCAACCATCGACGAGAGGACACCACGTGGCCCAGCGGCCGTC
CAGGCGGCGGCGCGGCAGCGGCGCCGGCCTCGATCCGCCGGCCCCGCGGCGCGCCGCGCCAGCT

GCTGCTCGACACGGCCCGGGCCCGCTTTGCCCGGCAGGACTACCGAAGCACCACGACACGCGA

***

AATCGCCCAGGCCGCAGGGG TAA CCG AGC ATC TGC TGT TCC GCC ACT TCG GTT CCA
Pro Ser Ile Cys Cys Ser Ala Thr Ser Val Pro

AGG CGG CGC TGT TCC GCG AGG CGC TGG TGC TGC CGT TCA CCG ATT TCG TCG CCG
Arg Arg Arg Cys Ser Ala Arg Arg Trp Cys Cys Arg Ser Pro Ile Ser Ser Pro

AAT TCG AGC TCG GTA CCC GGG GAT CCC GTC GTT TTA CAA CGT
Asn Ser Ser Ser Val Pro Gly Asp Pro Val Val Leu Gln Arg

Figure 9A. Nucleotide sequence upstream of *lacZ* in pAJB60. The *lacZ* codons are underlined. A stop codon is indicated by three "*" above the sequence. No typical translational start signal was identified and the transcriptional start could not be experimentally determined.

GATCACCAGCTCCGGCTTGCCCAGCGACGACACCAGGTCGGCGAGCTTGATGTTGCGGGCCTC
-35 -10

CGACATGGCGCCCGGATAGAAGGAGGC GAT CCC GTC GTT TTA CAA
RBS Pro Val Val Leu Gln

Figure 10A. Nucleotide sequence upstream of *lacZ* in pAJB89. The *lacZ* codons are underlined. No typical translational stop or start signals were identified for this sequence. The potential -35, -10 and ribosome binding site are shown in bold.

# APPENDIX B: NUCLEIC ACID BLAST HOMOLOGY RESULTS

## BLAST Search Results

| pAJB89: | % Recogntion |
|---|---|
| >gb\|U17898\|BSU17898 Pseudomonas solanacearum exopolysaccharide EPS I synthesis and export locus, epsA, epsB, epsC, epsD, epsE, and epsF genes, complete cds. (Length = 9382) | Positives = 31/38 (81%) |
| >emb\|X16300\|SVPKS Streptomyces violaceoruber polyketide synthase gene cluster. Length = 5820 | Positives = 42/66 (63%) |
| >emb\|X16144\|SVGRA S. violaceoruber DNA for granaticin polyketide synthase. Length = 6793 | Positives = 42/66 (63%) |
| >emb\|X74768\|SCCYAAC S.coelicolor cya gene for adenylate cyclase. Length = 1324 | Positives = 48/80 (60%) |
| >gb\|L13419\|CVNFCCAB Chromatium vinosum tetraheme cytochrome c gene, 3' end, bacterial ankyrin homologue, flavocytochrome c heme subunit fccA (complete cds) ,and flavin subunit, fccB (3' end). Length = 2562 | Positives = 31/42 (73%) |
| **pAJB304:** | |
| >gb\|M63677\|SERERYAB S.erythraea second and third ORF's of eryA gene, complete cds. Length = 20,235 | Positives = 60/90 (66%) |
| >emb\|X62569\|SEERYABS S.erythraea eryA gene for 6-deoxyerythronolyde B synthase II & III . Length = 20,444 | Positives = 60/90 (66%) |
| >emb\|D13614\|SGORF S.griseus genes encoding proteins similar to the regulators of two-component regulatory systems and membrane translocators, complete cds. Length = 5428 | Positives = 36/44 (81%) |
| >emb\|X72787\|SGNUSG S.griseus nusG, rplKAJL gene cluster. Length = 7235 | Positives = 32/45 (71%)<br>Positives = 33/49 (67%)<br>Positives = 36/56 (64%)<br>Positives = 32/51 (62%)<br>Positives = 35/58 (60%) |
| >emb\|X65467\|SLMRO S.lividans mercury resistance operon Length = 5082 | Positives = 35/58 (60%)<br>Positives = 49/76 (64%) |
| >gb\|M37378\|STMAPHE S.griseus streptomycin-3'- | Positives = 51/80 (63%) |

phosphotransferase gene, complete cds.
Length = 2854

>gb|U02635|SLCPOL Streptomyces lividans TK64 chloroperoxidase (cpoL) gene, complete cds. Length = 1903 — Positives=63/112 (56%) Positives = 22/28 (78%)

>emb|Z19549|MTBCARBCP M.tuberculosis gene for biotin carboxyl carrier protein. Length = 2340 — Positives = 49/76 (64%)

>emb|Z11929|SCMETALLO S.coelicolor genes for metallo-proteinase and LysR-type transcriptional activator. Length = 4673 — Positives = 37/52 (71%) Positives = 25/35 (71%)

>emb|X62518|SCPKSGC S.curacoi polyketide synthase gene cluster. Length = 5469 — Positives = 39/59 (66%) Positives = 27/37 (72%)

>emb|X65195|SVPTT S.viridochromogenes genes pms, phsA, pat and dea for phosphinomethylmalic-acid-synthase, phosphinothricin-tripeptide-synthetase A, phosphinothricin-N-acetyltransferase and N-acetylphosphinothricin-tripeptide deacetylase, respectively Length = 4053 — Positives = 39/56 (69%)

>emb|X53527|SGAPHE Streptomyces griseus aphE gene for streptomycin phosphotransferase. Length = 819 — Positives = 50/80 (62%)

>gb|M16482|STMSMK S.griseus streptomycin 6-phosphotransferase gene, complete cds. Length = 1627 — Positives = 50/80 (62%)

>emb|X74904|GGLRPA2MR G.gallus mRNA for LRP/alpha-2-macroglobulin receptor Length = 15,598 — Positives = 35/48 (72%)

>emb|D13775|SLCA Streptomyces lividans chiA gene for chitinase A. Length = 2000 — Positives = 30/47 (63%) Positives = 28/34 (82%)

>emb|X15867|MLUVRA Micrococcus luteus homolog of the E.coli uvrA gene. Length = 3739 — Positives = 49/79 (62%)

>dbj|D12470|STMGASP S. fradiae DNA for glutamic acid-specific protease, complete cds. complete cds. Length = 2064 — Positives = 48/77 (62%)

>gb|L13455|STMVALDEHY Streptomyces coelicolor (clone pWHM1051) valine dehydrogenase (vdh) gene, complete cds; 3' end ORF1. Length = 3007 — Positives =71/130 (54%)

>gb|L32017|USMTOI Ustilago maydis topoisomerase I gene, complete cds. Length = 3740 — Positives = 42/64 (65%)

pAJB303:

| | |
|---|---|
| >gb\|U00015\|U00015 Mycobacterium leprae cosmid B1620. Length = 42,325 | Positives = 71/95 (74%) |
| >emb\|Y00142\|SLXP55 Streptomyces lividans DNA for xp55 protein and p49 protein. Length = 4540 | Positives = 37/50 (74%)<br>Positives = 23/30 (76%) |
| >gb\|L26338\|STMFPS Streptomyces roseofulvus frenolicin polyketide synthase genes, complete cds. Length = 5530 | Positives = 45/66 (68%) |
| >gb\|M15050\|AMSXYLI Ampullariella xylose isomerase gene, complete cds; xylulose kinase gene, putative partial cds. Length = 1892 | Positives = 52/86 (60%)<br>Positives = 23/30 (76%) |
| >emb\|X72857\|SCPSCP2 S.coelicolor plasmid SCP2 DNA. Length = 3267 | Positives = 50/85 (58%)<br>Positives = 29/40 (72%) |
| >emb\|Z14056\|PSHRPB P.solanacearum DNA for hrp gene locus. Length = 18,300 | Positives = 37/49 (75%) |
| >emb\|X55794\|SAGDDPEP S.albus G gene for extracellular metallo (Zn) DD-peptidase. Length = 1234 | Positives = 49/81 (60%)<br>Positives = 20/24 (83%) |
| >gb\|M77739\|ECOBETB E.coli betaine aldehyde dehydrogenase (betB) gene, complete cds. Length = 1854 | Positives = 36/48 (75%) |
| >emb\|X52905\|ECBET Escherichia coli betT, betI, betB and betA genes. Length = 7412 | Positives = 36/48 (75%) |
| >gb\|U14662\|BHU14662 Baboon herpesvirus HVP2 gB glycoprotein (UL27) gene, complete cds. Length = 2817 | Positives = 45/74 (60%)<br>Positives = 26/35 (74%)<br>Positives = 31/47 (65%) |
| >gb\|L14320\|HSBBICP4A Bovine herpesvirus type 1 early-intermediate transcription control protein (BICP4) gene, complete cds. Length = 8113 | Positives = 27/33 (81%)<br>Positives = 25/32 (78%) |
| >emb\|Z18946\|MLCGA Mycobacteriophage L5 complete genome. Length = 52,297 | Positives = 34/43 (79%)<br>Positives = 29/45 (64%) |
| >gb\|U12276\|BPU12276 Bordetella pertussis BrkA (brkA) gene and BrkB (brkB) genes, complete cds. Length = 4300 | Positives = 40/61 (65%)<br>Positives = 48/87 (55%)<br>Positives = 33/54 (61%) |
| >gb\|M64834\|STMLATACV S.clavuligerus lysine-epsilon-amino-transferase (lat) gene, complete cds and delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine | Positives = 37/52 (71%) |

synthetase (pcbAB) gene, 5' end. Length = 2632

>gb|U12015|SCU12015 Streptomyces clavuligerus lysine e-aminotransferase (lat) gene, complete cds, and delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase (pcbAB) gene, partial cds. Length = 4029 — Positives = 37/52 (71%)

>emb|X59793|ECRP4TRA E.coli broad-host-range IncP-alpha plasmid RP4 traA, traB, traC-1, traC-2 & traD genes Length = 4412 — Positives = 33/51 (64%) Positives = 45/79 (56%) Positives = 21/27 (77%)

>emb|X62287|SCWHIB S.coelicolor whiB gene Length = 3488 — Positives = 30/43 (69%) Positives = 30/43 (69%)

>gb|M34651|SH1PROIE Pseudorabies virus with upstream and downsteam sequences. Length = 15,141 — Positives = 34/47 (72%) Positives = 25/35 (71%)

>gb|U08602|SSU08602 Streptomyces sp. WL6 alpha-amylase (amy) and (urf2) genes, complete cds and (urf1) gene, partial cds. Length = 3088 — Positives = 34/46 (73%)

**pAJB305:**

>emb|X79146|SLLINC S.lincolnensis (78-11) Lincomycin production genes. Length = 36,270 — Positives = 37/50 (74%) Positives = 25/30 (83%) Positives = 30/42 (71%) Positives = 40/65 (61%)

>gb|L34880|STMDAUABCE Streptomyces sp. aklaviketone reductase (dauE) gene; cyclase/dehydrase (dauA) gene; polyketide reductase (dauB) gene; oxygenase, polyketide synthase, acyltransferase (dauA) genes; aklanonic acid methyltransferase (dauC) gene. Length = 8089 — Positives = 70/120 (58%)

>gb|M37181|PSEALG76 P.aeruginosa alg76 gene, complete cds. Length = 2328 — Positives = 35/49 (71%) Positives = 33/49 (67%)

>gb|L29417|PSEGAS Pseudomonas aeruginosa (clone pGL2B5) gyrase A subunit (gyrA) gene, complete cds. Length = 3150 — Positives = 71/123 (57%)

>gb|L13455|STMVALDEHY Streptomyces coelicolor (clone pWHM1051) valine dehydrogenase (vdh) gene, complete cds; 3' end ORF1. Length = 3007 — Positives = 41/56 (73%)

>gb|L23970|AVIHOXHYP Azotobacter vinelandii hox and hyp operons, complete cds. Length = 13,914 — Positives = 41/62 (66%) Positives = 28/41 (68%) Positives = 26/37 (70%) Positives = 39/67 (58%) Positives = 34/56 (60%)

| | |
|---|---|
| >gb\|L35560\|STMDNRDPS Streptomyces peucetius daunorubicin-doxorubicin polyketide synthase (dnrH, dpsA-dpsF) genes, ORF5 and ORF8. Length = 8737 | Positives = 68/118 (57%) |
| >emb\|X62666\|PAXCPSP P.aeruginosa genes xcpR, xcpS, xcpT, xcpU, xcpV, xcpW and xcpX for secretion proteins. Length = 5940 | Positives = 32/44 (72%)<br>Positives = 26/35 (74%) |

**pAJB300:**

| | |
|---|---|
| >emb\|X59610\|AVNFRX A.vinelandii nfrX gene for uridylyl transferase. Length = 2790 | Positives = 70/121 (57%) |
| >emb\|D11136\|HHHAL Halobacterium halobium DNA for mex halorhodopsin, partial sequence. Length = 619 | Positives = 64/108 (59%) |
| >emb\|Z46913\|SAPKSGENE S.ambofaciens gene for hypothetical polyketide gene. Length = 1664 | Positives = 42/62 (67%) |
| >emb\|D90223\|PABRACG P.aeruginosa braCDEFG operon for branched-chain amino acid transport (LIV-I) components. Length = 5302 | Positives = 38/53 (71%) |
| >gb\|M28303\|STMLACBG S.albus beta-lactamase gene, complete cds. Length = 1410 | Positives = 25/31 (80%)<br>Positives = 33/51 (64%) |
| >emb\|Z14056\|PSHRPB P.solanacearum DNA for hrp gene locus. Length = 18,300 | Positives = 57/98 (58%)<br>Positives = 19/23 (82%) |
| >gb\|L29299\|FRANIFX Frankia alni nitrogen fixation gene cluster (including nifX, nifW, nifZ, nifB, and nifU), complete cds's. Length = 4567 | Positives = 47/76 (61%) |

**pAJB301:**

| | |
|---|---|
| >emb\|X14112\|HE1CG Herpes simplex virus type 1 complete genome. Length = 152,260 | Positives=53/87 (60%)<br>Positives=52/93 (55%)<br>Positives=18/19 (94%) |
| >emb\|XO2499\|RRATP Rhodospirillum rubrum atp operon. Length = 8775 | Positives=42/63 (66%) |

# APPENDIX C: AMINO ACID BLAST HOMOLOGY RESULTS

pAJB300:

```
sp|P36756|VL2_HPV30  Probable L2 protein                    Positives=14/17  (82%)
      Query:    5  PVANTLTLELEPVVEAN  21
                   P+A+T  +E++P++ AN
      Sbjct:  325  PIAHTEEIEMQPLLSAN  341

sp|P23870|HHA_ECOLI  Haemolysin expression modulating      Positives=7/9  (77%)
                     protein                                Positives=6/8  (75%)
      Query:    1  MAERPVANT  9    14  LEPVVEAN  21
                   M+E+P+  T          LE V +E N
      Sbjct:    1  MSEKPLTKT  9    24  LERVIEKN  31
```

pAJB301:

```
sp|Q02388|CA17_HUMAN  Collagen alpha 1(VII) chain (Long    Positives=15/25  (60%)
                      chain collagen)                      Positives=6/11  (54%)
                                                           Positives=9/12  (75%)
      Query:   12  LAVAPLGDRAAGQLPGRLSTVSTQP  36
                   L V+P+ D  A +LPG+   VS  P
      Sbjct:  379  LPVSPVTDLQATELPGQRVRVSWSP  403

      Query:   34  TQPGASTVERVT  45      51  GLPSGPAPVDP  61
                   T PG S+ +RVT            GLP  P P  P
      Sbjct:  878  TLPGISSSQRVT  889  1300  GLPGSPGPQGP  1310

sp|P35569|IRS1_MOUSE  Insulin receptor substrate-1         Positives=9/15  (60%)
                                                           Positives 15/27  (55%)
      Query:   13  AVAPLGDRAAGQLPG  27
                   A  P G  + G+LPG
      Sbjct:  552  AAYPPGGGSGGRLPG  566

      Query:   30  STVSTQPGASTVERVTHSHSHGLPSGP  56
                   S+ S  P  +T E+ THS   G P GP
      Sbjct: 1036  SSASVTPQGATAEQATHSSLLGGPQGP  1062

sp|P16917|RHSB_ECOLI  RHSB protein precursor               Positives=11/19  (57%)
                                                           Positives=8/12  (66%)
                                                           Positives=10/27  (37%)
      Query:    1  VMRGDGERFGNL  12
                   V R D +RFG +
      Sbjct:  549  VTRYDHDRFGQM  560

      Query:   14  VAPLGDRAAGQLPGRLSTVSTQPGAST  40
                   +AP G R   Q       V T  G  T
      Sbjct:  606  IAPDGSRNGTQYDAWGKAVRTTQGGLT  632

      Query:   36  PGASTVERVTHSHSHGLPS  54
                   P   + E+V+  H+HG  S
      Sbjct: 1331  PCPNGTEKVSAYHTHGADS  1349
```

```
sp|P31574|FIXB_ECOLI  FixB protein                          Positives=20/39  (51%)
      Query:     8  RFGNLAVAPLGDRAAGQLPGRLSTVSTQPGASTVERVTH  46
                    R G L  A LG R     +   STVS Q G +TV+ + +
      Sbjct:    54  RRGKLLAAKLGYRLKAAVSNDASTVSVQDGKATVKHMVY  92

sp|P28691|FTSH_ECOLI  Cell division protein ftsH            Positives=9/15  (60%)
                                                            Positives=15/45(33%)
      Query:     6  GERFGNLAVAPLGDRAAGQLPGRLSTVSTQPGASTVERVTHSHSH  50
                    G   G     P GD  +       S +ST  G    E  +  +  H
      Sbjct:   440  GRALGVTFFLPEGDAISASRQKLESQISTLYGGRLAEEIIYGPEH  484

      Query:    46  HSHSHGLPSGPAPVD  60
                    +S  +G P  P PVD
      Sbjct:   612  NSGDNGSPKAPRPVD  626
```

pAJB303:

```
sp|P32536|ENV_HV2S2  Envelope polyprotein GP160 precursor  Positives=12/18  (66%)
                                                            Positives=7/13  (53%)
                                                            Positives=7/18  (38%)
      Query:    27  GLGGRGRRRRPLDFWRGR  44   45  IPPHRGARTARST  57
                    G  G     R   +W GR          +PP  G   T  ST
      Sbjct:   264  GFNGTRAENRTYIYWHGR  281   432  LPPREGQLTCNST  444

      Query:    69  RADRAGLRQEHQRLLDAV  86
                    R   AG+ Q+ Q+LLD V
      Sbjct:   535  RTSLAGIVQQQQQLLDVV  552

sp|P33485|VNUA_PRVKA  Probable nuclear antigen              Positives=14/25  (56%)
                                                            Positives=12/28  (42%)
      Query:    26  AGLGGRGRRRRPLDFWRGRIPPHRG  50
                    AG G R RRRR  D   G + P RG
      Sbjct:  1001  AGGGARRRRRRRWDDEAGLLGPERG  1025

      Query:    46  PPHRGARTARSTRVGGRLRCFVRRADRA  73
                    P  RG R A     GGR  C   RA  A
      Sbjct:  1646  PSPRGCRGAGRAGGGGRGGCGGGRAPGA  1673

sp|P18040|ENV_HV2G1  Envelope polyprotein GP160 precursor  Positives=11/14  (78%)
                                                            Positives=7/13  (53%)
                                                            Positives=7/18  (38%)
      Query:    27  GLGGRGRRRRPLDFWRGR  44   45  IPPHRGARTARST  57
                    G  G     R   +W GR          +PP  G  T  ST
      Sbjct:   261 GFNGTRAENRTYIYWHGR  278   429  LPPREGQLTCNST  441

      Query:    73  AGLRQEHQRLLDAV  86
                    AG+ Q+ Q+LLD V
      Sbjct:   536 AGIVQQQQQLLDVV  549
```

pAJB305:

```
sp|P14581|CP47_RABIT  Cytochrome P450 IVA7 (EC 1.14.15.3)  Positives=9/10  (90%)
      Query:     1  METALGRSDP  10
                    M+T LGRSDP
      Sbjct:   105  MKTILGRSDP  114
```