

# Gene Knockout Reveals a Novel Gene Cluster for the Synthesis of a Class of Cell Wall Lipids Unique to Pathogenic Mycobacteria\*

(Received for publication, April 17, 1996, and in revised form, May 12, 1997)

Abul K. Azad, Tatiana D. Sirakova, Norvin D. Fernandes, and Pappachan E. Kolattukudy‡

From the Neurobiotechnology Center and Departments of Biochemistry and Medical Biochemistry, The Ohio State University, Columbus, Ohio 43210

Surface-exposed unusual lipids containing phthiocerol and phenolphthiocerol are found only in the cell wall of slow-growing pathogenic mycobacteria and are thought to play important roles in host-pathogen interaction. The enzymology and molecular genetics of biosynthesis of phthiocerol and phenolphthiocerol are unknown. We postulate the domain organization of a set of multifunctional enzymes and a cluster of genes (*pps*) that would encode these enzymes for the biosynthesis of phthiocerol and phenolphthiocerol. A cosmid containing the postulated *pps* gene cluster was identified by screening a genomic library of *Mycobacterium bovis* BCG with the postulated homologous domains from mycocerosic acid synthase and fatty acid synthase genes as probes. Homologous cosmids were also identified in the genomic libraries of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *M. bovis* BCG was transformed with a *pps* disruption construct, made from the BCG cosmid by introducing the hygromycin resistance gene as the positive-selectable marker and the *sacB* gene as the counter-selectable marker. Gene disruption by homologous recombination with double crossover was confirmed by polymerase chain reaction and Southern hybridization. Chromatographic analysis showed that the phenolphthiocerol derivative, mycoside B, and phthiocerol dimycocerosates were not produced by the gene knockout mutants. This result confirms the identity of the *pps* genes. With the identification of the *pps* gene clusters in both *M. tuberculosis* and *M. leprae*, it should be possible to test the postulated roles of these unique lipids in tuberculosis and leprosy.

Tuberculosis, caused by *Mycobacterium tuberculosis*, claims twice as many human lives as AIDS and more lives than any other infectious disease (1). The rapid spread of tuberculosis among the AIDS patients and the emergence of multidrug-resistant strains of *M. tuberculosis* demand the development of new drugs (2, 3). Unfortunately, no new antimycobacterial target for drug development has been characterized in the past 25 years. Since the unique cell wall lipids of mycobacteria play a major role in survival of the pathogen in the hostile environment within the host (4–6), inhibition of synthesis of wall components can be an effective antimycobacterial therapy (7). Surface-exposed lipids such as phthiocerol esters and mycosides have been found to be unique to pathogenic mycobacteria (8–10) and play important functions in mycobacterial interac-

tion with the host (11, 12). These lipids contain two diols, phthiocerol and phenolphthiocerol, respectively, esterified with multiple methyl-branched long chain fatty acids called mycocerosic acids (8, 12–14). Thus, biosynthesis of mycocerosic acids, phthiocerol, or phenolphthiocerol may provide new antimycobacterial drug targets and clues to pathogenesis of tuberculosis. We previously purified the mycocerosic acid synthase (MAS)<sup>1</sup> (15) and cloned and sequenced the gene encoding MAS (*mas*) (16).

Labeling studies using <sup>14</sup>C-labeled propionate and acetate indicated that both are incorporated into the long chain diols (17–19). Therefore, it seems likely that the aliphatic chains are derived from both malonyl-CoA and methylmalonyl-CoA. However, nothing is known about the enzymes involved in their biosynthesis or the genes that encode them. Based on current biochemical concepts, here we postulate a unique biosynthetic pathway for phthiocerol and phenolphthiocerol and a genetic organization for a cluster of genes (*pps*) that would encode a set of enzymes to catalyze the synthesis of phthiocerol from docosanoyl and phenolphthiocerol from *p*-hydroxyphenylheptadecanoyl primers. We report the identification of the postulated *pps* gene cluster in *Mycobacterium bovis* BCG by demonstrating that disruption of this gene cluster abolishes the production of both phthiocerol and phenolphthiocerol derivatives. We also report that the same genes are present in *M. tuberculosis* and *Mycobacterium leprae*, the two most pathogenic species of slow-growing mycobacteria.

## EXPERIMENTAL PROCEDURES

**Screening of *M. bovis* BCG Genomic DNA Library**—A cosmid DNA library of BCG genome was screened by colony-blot hybridization at low stringency with active site motifs of acyl transferase (AT; nucleotides 3259–3753) and ketoacyl synthase (KS; nucleotides 2061–2370) domains of the *mas* gene (16) as probes. The positive clones were re-screened with AT (nucleotides 103–354) and KS (nucleotides 7534–7842) domains of the fatty acid synthase gene (*fas*) (20). DNAs from the cosmids, found positive with both sets of probes, were analyzed by Southern hybridization with AT, KS, ketoreductase (KR; nucleotides 6847–7164), enoyl reductase (ER; nucleotides 6040–6622), and acyl carrier protein (ACP; nucleotides 7789–8079) domains of *mas*, and AT and KS domains of *fas*. Two clones containing the original *mas* were discarded. The positive fragments from another clone were subcloned into pUC19 and sequenced. Sequences were searched for homology by BLAST server. A total of 8.2 kb was sequenced from 12 segments of BCG cosmid 49, and on the basis of homology search, this was selected as a candidate cosmid for the *pps* gene cluster and found to be the homolog to *M. leprae* cosmid L518 (GenBank™ accession number U00023).

**Construction of a Plasmid for *pps* Gene Disruption**—An 8.2-kb *Bam*HI fragment was isolated from the BCG cosmid 49 by hybridization with a 6.6-kb coding segment, obtained by PCR from *M. leprae* cosmid

\* This work was supported in part by Grant R01 AI35272 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 614-292-5682; Fax: 614-292-5379; E-mail: kolattukudy.2@osu.edu.

<sup>1</sup> The abbreviations used are: MAS, mycocerosic acid synthase; AT, acyl transferase; KS, ketoacyl synthase; KR, ketoreductase; ER, enoyl reductase; ACP, acyl carrier protein; kb, kilobase pair(s); PCR, polymerase chain reaction; FAS, fatty acid synthase; ORF, open reading frame.

L518. The 8.2-kb fragment was cloned into pUC19 to produce plasmid 49-8B. An internal 1.9-kb fragment was deleted from 49-8B with *Nru*I and replaced with a 1.5-kb Klenow-filled *Bgl*II-*Sma*I fragment, containing the hygromycin resistance gene (*hyg*), from plasmid pIJ963 (21) to generate pAKD9. A plasmid pAZ1.1 was created by cloning a 1.95-kb PCR fragment from pCVD442 (22), containing the *sacB* gene of *Bacillus subtilis* (23), into the *Sma*I site of pUC19. A 7.9-kb *Bam*HI fragment was isolated from pAKD9 and ligated to the unique *Bam*HI site of pAZ1.1 to construct pAKD10 (Fig. 2D), which thus contains the *sacB*, and the *hyg* flanked by 3.8- and 2.5-kb mycobacterial DNA in pUC19, a vector non-replicative in mycobacteria.

**Transformation of BCG with pAKD10 and Screening of Transformants**—Construct pAKD10 was linearized with *Xba*I and electroporated into BCG competent cells as described previously (24). Transformant colonies grown on 7H11 agar (Difco) medium containing hygromycin B (50 µg/ml; Calbiochem) were individually transferred to 3 ml of M-ADC-TW broth (25) in 15-ml tubes and grown with shaking in a rotator (Thomas Scientific, Swedesboro, NJ). Culture from each tube was then streaked onto 7H11 agar containing hygromycin B and 10% sucrose. BCG transformants grown on sucrose were regrown in 100 ml of broth in the presence of hygromycin B for extraction of genomic DNA for use in Southern hybridization and PCR.

**Genomic DNA Preparation and Southern Hybridization**—Genomic DNA from BCG cultures was extracted as described (25) after disrupting the cells with 425–600-µm glass beads (Sigma) in a Mini-Bead-beater cell disruptor (Biospec-Products, Bartlesville, OK). Southern hybridization of cosmid or genomic DNA was performed according to standard procedures as described previously (24).

**PCR Methods**—Standard PCR was done to amplify 1.95-kb *sacB* segment with primers: 5'-GATCCTTTTAAACCCATCAC-3' and 5'-AATGCCAATAGGATATCGGC-3', using *Taq* DNA polymerase (Boehringer Mannheim) and 10% dimethyl sulfoxide ( $\text{Me}_2\text{SO}$ ). The Expand long template PCR system (Boehringer Mannheim) was employed to amplify the 2.6- and 3.9-kb mycobacterial fragments, using supplier's buffer 1 and 2%  $\text{Me}_2\text{SO}$  with the following pairs of primers: 5'-GGACCTCGACGACCTGCAGGCATGCAAGC-3' (*H1*), 5'-GCTGTTGGCCGGCGCCGCGCAGGCGCTG-3' (*D1*), and 5'-CGCAGTTCCTCTGGGGGCGCCGGACAC-3' (*S2*), 5'-CGGTCGTGGTGGCCATGCCCGCTCCGGC-3' (*D2*), respectively. The same PCR system was also used to amplify 6.6-kb fragment from cosmid L518 with primers: 5'-ATGACCGCCTCGACACCTGATC-3' and 5'-TCATGACTCGCCTCTCGTTGCA-3'. PCR conditions in both systems were as described (24).

**Radiolabeling Studies and Biochemical Analysis of Cell Wall Lipids from BCG Wild Type and *pps* Mutant Strains**—To 100 ml of the 12-day-old culture of each of BCG wild type and its *pps*-disrupted mutant, 42 µCi of Na[1- $^{14}\text{C}$ ]propionate (specific activity, 55 Ci/mol; American Radiolabeled Chemicals) was added, and incubation of the cultures was continued in roller bottles for a further 16 h. The cells were collected by centrifugation and extracted with an excess of a 2:1 mixture of chloroform and methanol. The total lipids were recovered by the Folch method (26), and aliquots were assayed for  $^{14}\text{C}$ . The total lipid was subjected to thin-layer chromatography (TLC) on 20 × 20 cm × 0.5-mm layers of silica gel G with a 19:1 (v/v) mixture of chloroform and methanol as the solvent. Chromatograms were scanned for  $^{14}\text{C}$  with Berthold Tracemaster 20 TLC-linear analyzer. The silica gel from the labeled area, representing phthiocerol and other wax esters, was scrapped off, and lipids were recovered by elution with ethyl ether. This lipid fraction was applied to TLC plate and developed in a 9:1 (v/v) mixture of *n*-hexane and diethyl ether as the solvent. Lipids from non-radioactive cells from 12-day-old wild type and gene-disrupted mutant cultures were also analyzed using the same methods. For detection of lipids, chromatograms were sprayed with 5%  $\text{K}_2\text{Cr}_2\text{O}_7$  in 50%  $\text{H}_2\text{SO}_4$  and heated for 15 min at 180 °C. Identification of mycoside B and dimycocerosyl phthiocerol was done by comparison of their  $R_f$  values with those of authentic standards obtained from Professor P. J. Brennan, Colorado State University, and by hydrolysis followed by radio-gas chromatography as described previously (24).

**Extraction of Total Proteins and Electrophoresis**—Cells from 200 ml of culture of each of BCG wild type and *pps* mutant were harvested by centrifugation, disrupted in a French press, and the total proteins were extracted as described previously (15). Proteins were analyzed by SDS-polyacrylamide gel electrophoresis with a 3% stacking gel and a 5% running gel. For detection of protein bands, the gel was silver-stained by using the Bio-Rad silver stain kit (Bio-Rad) according to the manufacturer's instructions.

**MAS Activity Determination**—The cell-free extract was fractionated by DEAE-Sephadex chromatography, and MAS activity was measured with labeled methylmalonyl-CoA as described (15).

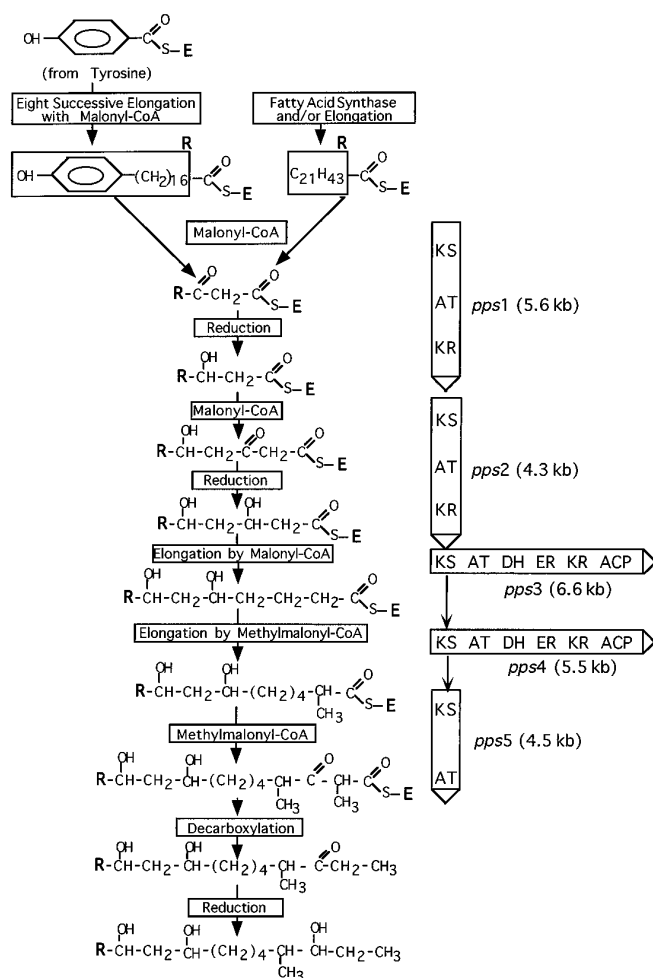
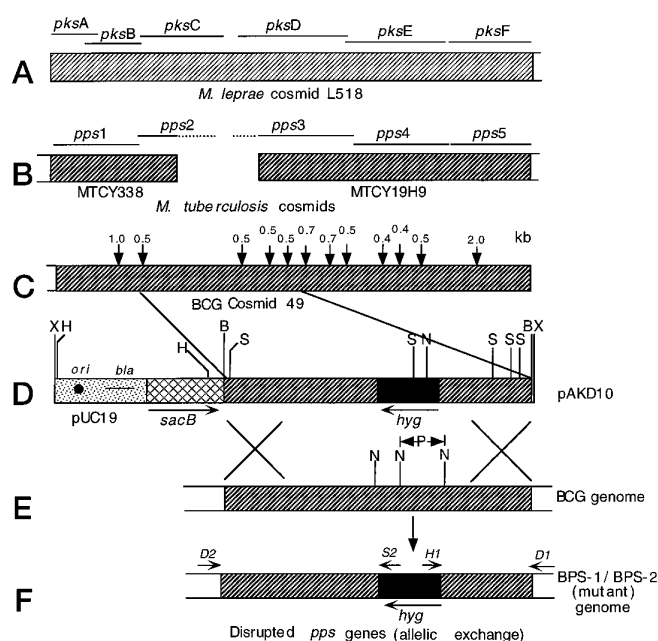


FIG. 1. Postulated pathways and organization of a gene cluster for the biosynthesis of phthiocerol and phenolphthiocerol. *E* represents enzymes to which the intermediates may be bound during biosynthesis. The gene and the domain organization of the enzymes postulated to catalyze the various biosynthetic steps are shown at the right side. The cluster of five genes, *pps1*–*pps5*, of different sizes would encode the five multifunctional enzymes. KS, ketoacyl synthase; AT, acyl transferase; KR, keto reductase; DH, dehydratase; ER, enoyl reductase; ACP, acyl carrier protein.

## RESULTS

**Identification of the Cosmids Containing the Postulated *pps* Gene Cluster in Pathogenic *Mycobacteria***—The structures of the closely related diols, phthiocerol and phenolphthiocerol, suggest that they can be produced from two different starter molecules using identical series of biochemical reactions as indicated in Fig. 1. Based on the current knowledge of fatty acid synthase (FAS), MAS, and polyketide synthase, the enzyme domains required to catalyze each biosynthetic step can be postulated as shown in Fig. 1. The open reading frames (ORFs) required to encode these domains could be organized in the mycobacterial genome in a cluster, designated *pps1*, *pps2*, *pps3*, *pps4*, and *pps5* (Fig. 1).

The biosynthesis of both diols involves synthesis of a straight chain segment followed by synthesis of methyl-branched segments. FAS-like domains might be involved in the synthesis of the straight chain segment by incorporating malonyl-CoA, whereas methylmalonyl-CoA-specific MAS-like domains might synthesize the branched segments. The AT and KS domains of MAS show high degree of homology to the corresponding domains of erythronolide synthase (*eryA*) that uses methylmalonyl-CoA as substrate (27, 28), whereas the other domains of MAS showed equal homology to the corresponding domains of



**FIG. 2. Schematic diagram for mycobacterial cosmids containing the *pps* gene cluster and the strategy for generation of the *pps* gene-disrupted BCG mutants.** A, *M. leprae* cosmid L518. The nucleotide positions of the ORFs are: *pksA* (56–1963), *pksB* (1549–4395), *pksC* (4396–8736), *pksD* (9016–15621), *pksE* (15,618–21,065), and *pksF* (21,094–25,563); B, *M. tuberculosis* cosmids MTCY338 and MTCY19H9. The nucleotide positions of the ORFs are as follows: *pps1* (25,005–30,635), *pps2* (30,632 to >31,307), *pps3* (<1–5457), *pps4* (5454–10,937), and *pps5* (10,943–15,409). In both cases the bars above the cosmids represent the relative positions and sizes of the designated ORFs, which transcribe in the same direction. C, *M. bovis* BCG cosmid 49. The numbers and arrows above indicate the size and location of the sequenced segments. D, the *pps* gene disruption construct, pAKD10, used for transformation. Hatched boxes represent mycobacterial DNA, the stippled box represents the vector pUC19, the black boxes represent the *hyg* gene as a disruption element as well as positive selectable marker, and the cross-hatched box represents the *sacB* gene as a counter-selectable marker. E, BCG genome representing part of the *pps* genes and sites of integration. F, genome organization of the allelic exchange mutant at the *pps* locus. Restriction enzymes: B, *Bam*HI; H, *Hind*III; N, *Nru*I; S, *Sst*I; X, *Xba*I. P indicates the 1.3-kb *Nru*I deleted fragment from the *pps* genes used as the probe for Southern blot in Fig. 3. H1-D1 and S2-D2 represent the primers for PCR used to test for homologous recombination.

both FAS and *eryA* (16). Therefore the AT and KS domains of MAS were suggested to be selective for methylmalonyl-CoA, and this suggestion was recently experimentally verified with expressed AT and KS domains of MAS.<sup>2</sup> Thus, AT and KS domains of FAS and MAS might selectively recognize the *pps* domains that encode enzymes involved in the synthesis of *n*- and methyl-branched segments of the diols, respectively. Therefore, a cosmid library of the *M. bovis* BCG genomic DNA was screened with AT and KS domains of *mas*, and the positive clones were rescreened with AT and KS domains of *fas* to select clones that hybridize with both sets of domains. This approach yielded three cosmid clones (D, G, and 49), which were further analyzed by Southern hybridization with AT, KS, KR, ER, and ACP domains of *mas* and AT and KS domains of *fas*. DNA sequence analysis of segments of these cosmids revealed that the overlapping cosmids D and G contain the original *mas* gene. The nucleotide sequences of the 12 segments of cosmid 49, indicated in Fig. 2, showed 85% identity with sequences in *M. leprae* cosmid L518 (GenBank<sup>TM</sup> accession number U00023). Analysis of L518 sequence showed that nucleotides 56–25,563

of this cosmid contain domains organized as postulated for *pps* genes and are depicted in Fig. 2A. DNA sequence from nucleotide 25,005–31,307 of the *M. tuberculosis* cosmid MTCY338 (GenBank<sup>TM</sup> accession number Z74697) contained *pps1* and part of *pps2*; whereas nucleotide 1–15,409 of cosmid MTCY19H9 (GenBank<sup>TM</sup> accession number Z83857) contained part of *pps3* and the entire *pps4* and *pps5* (Fig. 2B). The nucleotide sequence of the 12 different segments totalling 8.2 kb of BCG cosmid 49 showed 96–100% identity with the corresponding segments of the *M. tuberculosis* genome, thus revealing the location of the putative *pps* gene cluster in cosmid 49. The alignment of the *pps* gene cluster contained in *M. leprae* cosmid L518, *M. tuberculosis* cosmids MTCY338 and MTCY19H9, and in the corresponding segment of *M. bovis* BCG cosmid 49 is shown in Fig. 2, A–C. Mapping and sequencing studies of the BCG cosmids D, G, and 49, and the sequence information from the *M. tuberculosis* cosmids MTCY19H9 and MTCY24G1 (GenBank<sup>TM</sup> accession number Z83858), showed that the *pps* gene cluster is about 4 kb downstream from the *mas* locus in *M. bovis* BCG and *M. tuberculosis* genomes and that they are transcribed in the opposite direction.

**Disruption of the *pps* Genes by Allelic Exchange.**—To test whether the postulated *pps* gene cluster actually encodes the proteins responsible for phthiocerol and/or phenolphthiocerol synthesis, the effect of disrupting the *pps* gene cluster in *M. bovis* BCG was investigated. To achieve allelic exchange, a construct (pAKD10) was made that contained a segment of the *pps* gene cluster, a small portion of the 3' region of *pps1*, the entire *pps2*, and the 5'-half portion of *pps3* (Fig. 2, C and D). In this gene segment, parts of *pps2* and *pps3* were deleted and replaced with *hyg*. As a counter-selectable marker, the *sacB* gene (23) was introduced into the vector adjacent to the cloned *pps* gene for positive selection of the allelic exchange mutants. Allelic exchange is only possible by homologous recombination with double crossover, which would insert the *hyg* gene but not the *sacB* gene into the mycobacterial genome. Since the expression of the intact *sacB* gene would not allow growth on sucrose (29), transformants that have undergone allelic exchange would be hygromycin-resistant and would grow on sucrose. On the other hand, nonhomologous or single-crossover recombinants should be hygromycin-resistant but unable to grow on sucrose. Twenty-five hygromycin-resistant BCG transformants generated by electroporation of the construct were screened for growth on sucrose. Five transformants grew on sucrose, and they were tested for homologous recombination using PCR to amplify both sides of the inserted *hyg* gene. With two sets of primers (H1-D1 and S2-D2; Fig. 2F), each representing a *hyg* primer and a primer in the mycobacterial genome outside the *pps* gene used for making the disruption construct, three transformants (BPS-1, BPS-2, and BPS-3) yielded a 2.6-kb and a 3.9-kb PCR product as expected from homologous recombination with double crossover (Fig. 3A; shown for BPS-1 and BPS-2). When the deleted fragment (P in Fig. 2E), the *sacB* gene, and the *hyg* gene (Fig. 2D) were used as probes in Southern hybridization of the genomic DNA from two of these transformants, they showed the absence of the genomic segment deleted in making the disruption construct, the absence of *sacB* gene, and the presence of *hyg* gene (Fig. 3B). These results confirmed that the *pps* gene cluster was disrupted by allelic exchange.

**Biochemical Analysis of Lipids from the *pps* Mutant.**—If the postulated gene cluster is responsible for the production of phthiocerol and/or phenolphthiocerol, the mutants should be unable to produce their naturally occurring derivatives, phthiocerol dimycocerosates and/or mycoside B, respectively. Thin-layer chromatographic analysis of the lipids from the mutant

<sup>2</sup> N. Fernandes and P. E. Kolattukudy, manuscript in preparation.

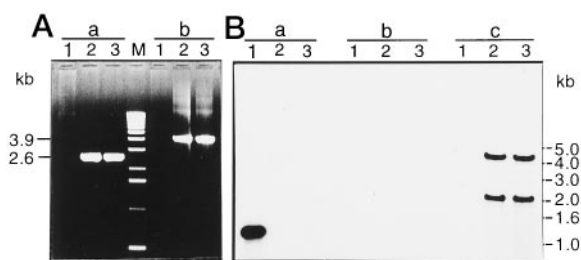


FIG. 3. Demonstration of allelic exchange in *M. bovis* BCG by PCR and Southern hybridization. A, agarose gel of PCR products showing allelic exchange event at the *pps* locus. Genomic DNAs from wild type BCG (lane 1), mutant BPS-1 (lane 2), and mutant BPS-2 (lane 3) were used as templates in PCR with two sets of primer shown in Fig. 2F: *H1-D1*, yielding 2.6 kb product (a) and *S2-D2*, yielding a 3.9-kb product from mutants only (b). M, 1-kb DNA ladder. B, autoradiogram of Southern blot of genomic DNA from the wild type and *pps*-disrupted mutants of BCG. Genomic DNA was digested with either *NruI* (a) or with *SstI* (b and c), and the Southern blots were hybridized with three different probes: 1.3-kb internal *pps* fragment (P in Fig. 2E) that was replaced with *hyg* (a), 1.95-kb *sacB* gene (b), and 1.5-kb *hyg* gene (c). Size markers are from the 1-kb DNA ladder.

showed that mycoside B was absent in the mutant (Fig. 4A). To test for the production of phthiocerol dimycocerosates, the wax fraction recovered from the above chromatographic step was subjected to TLC in a less polar solvent system. The results showed that the wild type contained a component that had an  $R_F$  identical to that of phthiocerol dimycocerosates, whereas the mutant did not contain this lipid (Fig. 4B). Incubation of the cultures with [ $^{14}\text{C}$ ]propionate and analysis of the lipids also confirmed that the mutant was incapable of synthesizing mycoside B and phthiocerol dimycocerosates (data not shown). Radio gas-liquid chromatographic analysis of the fatty acids showed that the mutant did not incorporate labeled propionate into the long chain mycocerosic acids found in the wild type *M. bovis* BCG (data not shown). To test whether the inability of the *pps* gene-disrupted mutant to produce mycocerosic acids is caused by some indirect effect resulting in interference with the *mas* gene functions, we examined the MAS level. SDS-polyacrylamide gel electrophoresis analysis showed that the mutant produced as much MAS as the wild type (data not shown). DEAE-Sephadex fractionation (15) and enzymatic activity measurements showed that the *pps* gene-disrupted mutant contained as much MAS activity as the wild type.

#### DISCUSSION

Molecular strategies used for the synthesis of fatty acids and polyketides allow us to postulate reasonable mechanisms for the biosynthesis of the two diols, phthiocerol and phenolphthiocerol, found uniquely in pathogenic mycobacteria. Both diols have a segment that should be derived from malonyl-CoA and a segment that should be derived from methylmalonyl-CoA. The position of the two hydroxyl groups in both suggests that they could be left during the synthesis of the aliphatic chain rather than by introduction of hydroxyl groups into preformed chains. Thus, an enzyme module that synthesizes this segment would have an AT, KS, KR, but no DH and ER, domains. Only when the synthesis of the entire product involves repetition of all of the steps, such as in the case of FAS or MAS using malonyl-CoA and methylmalonyl-CoA, respectively, a single enzyme module (as dimer) is used. When the synthesis involves multiple dissimilar modules, each enzyme usually catalyzes only one cycle and does not repeat the process. We therefore postulate that two modules each with only AT, KS, and KR sequentially catalyze the condensation with malonyl-CoA each time followed by reduction to generate the 1,3-diol function. Since the next step would involve extension by a fully reduced

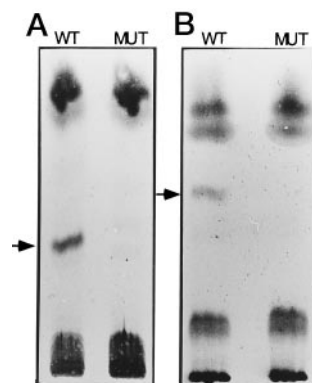


FIG. 4. Thin-layer chromatograms of lipids from wild type (WT) and *pps* gene-disrupted mutant (MUT). A, total lipids analyzed with a 19:1 (v/v) mixture of chloroform and methanol as the developing solvent. The arrow indicates the location of authentic mycoside B. B, the nonpolar wax fraction recovered from A analyzed with a 9:1 (v/v) mixture of hexane and diethyl ether as the developing solvent. The arrow indicates the location of phthiocerol dimycocerosates. In both A and B, 20 × 20 cm × 0.5-mm layers of silica gel G were used and chromatograms were sprayed with 5% (w/v)  $\text{K}_2\text{Cr}_2\text{O}_7$  in 50%  $\text{H}_2\text{SO}_4$  and heated for 15 min at 180 °C.

two-carbon unit, we postulate the need for a full FAS-like module that would use malonyl-CoA as the substrate. The next step should be chain extension by a fully reduced three carbon unit with a methyl branch by a complete MAS-like module. The final process could be condensation by methylmalonyl-CoA requiring a module containing KS and AT domains resembling the corresponding domains of MAS. The resulting 3-keto acid could undergo decarboxylation yielding a keto derivative. Reduction of the carbonyl and methylation of the resulting hydroxyl group would complete the process. The postulated keto intermediates in the synthesis of both phthiocerol and phenolphthiocerol occur naturally in mycobacteria (8, 14, 30).

Finding of the postulated specific and unique combination of domains organized in a modular manner in the mycobacterial genome supports our hypothesis that this cluster represents phthiocerol and/or phenolphthiocerol synthase genes. The postulated need for *mas*-like and *fas*-like AT and KS domains to synthesize the straight chain and methyl-branched segments prompted us to search for mycobacterial genomic segments that would hybridize with both types of domains. This approach led us to the BCG cosmid 49 as the one containing the postulated *pps* genes. The sequence of cosmid 49 that we had generated helped us to recognize the *pps* gene cluster when *M. tuberculosis* and *M. leprae* genomic sequences became available as part of the mycobacterial genome project.

Allelic exchange or gene knockout by homologous recombination is an efficient tool for generating defined mutants to study gene function(s). Gene knockout proved difficult in slow-growing mycobacteria such as *M. tuberculosis* and *M. bovis* BCG (31). So far, only three reports on gene disruption by allelic exchange in this group of mycobacteria are available (24, 32, 33). One of these reports involved disruption of the *ureC* gene with pH indicator screening for urease production (32). Another involved the *leuD* gene, and disruption resulted in leucine auxotrophic mutants (33). Since no general convenient method is available for screening the gene-disrupted mutants, we used PCR to identify *mas* gene-disrupted mutants (24). In the present work, we used the *sacB* gene from *B. subtilis* (23) as a sucrose sensitivity counter-selectable marker to avoid the need for extensive screening to isolate the allelic exchange mutants. Although some of the mutants that grow in sucrose may represent the mutated *sacB* gene and not the desired knockout mutants, this approach should be useful as a general

method for screening for the homologous recombinants with double crossover in the slow-growing mycobacteria, because the sucrose-grown transformants are highly enriched in such mutants.

Disruption of the gene encoding some of the modules involved in the biosynthesis of phthiocerol and/or phenolphthiocerol should prevent the synthesis of the diol(s) if another set of gene cluster in the genome cannot substitute for the disrupted one. Even though *pks*-like genes appear to be widely distributed in mycobacterial genomes, we could not find another cluster by our search by hybridization with *fas* and *mas* domains. Our disruption construct was designed in such a way that *pps2* and *pps3* would be disrupted. The gene-disrupted mutant was incapable of producing mycoside B and phthiocerol dimycocerosates, the known end products of the diol synthase, showing that the disrupted genes encode enzymes involved in the biosynthesis of both diols. This cluster of genes is therefore designated phthiocerol/phenolphthiocerol synthase genes, *pps*. These results show that the same set of enzymes synthesize both diols using two different starting acyl moieties as shown in Fig. 1. The relative amounts of the diols are probably determined by the availability of the starting acyl moieties.

Radio gas-liquid chromatographic analysis of fatty acids generated from [ $1^{14}\text{C}$ ]propionic acid by the *pps*-disrupted mutant showed no mycocerosic acids, whereas such acids are the major labeled fatty acids in the wild type (24). That the lack of such acids was not due to any indirect interference on MAS synthesis was shown by the fact that the mutant had normal levels of enzymatically active MAS. Mycocerosic acids occur mainly as esters of phthiocerol and phenolphthiocerol (8, 12, 14). In the absence of these diols, mycocerosic acids are either not synthesized or are degraded, as they cannot be esterified to the diols.

The *pps*-disrupted mutant should be useful in testing the postulated roles for mycoside B and phthiocerol dimycocerosates in mycobacterial interaction with the host. Comparison of this mutant, which lacks both mycoside B and phthiocerol dimycocerosates, with the *mas*-disrupted mutant, which only lacks the former, should provide information on the function of phthiocerol dimycocerosates. The rapidly expanding information on the genomic sequence of pathogenic mycobacteria can be used if the functional genes can be identified. The present results constitute an example of how the sequence emerging from the genome project can be used to identify functional genes. The gene-disrupted mutants should be useful in testing the postulated roles of the phenolic glycolipids such as inhibiting the production of reactive oxygen species by mononuclear cells (34) and promoting mycobacterial multiplication inside the host (35). Since the *pps* genes are unique to pathogenic mycobacteria, they may be used as targets in high throughput

screening for antimycobacterial drugs desperately needed to combat a major threat to human health.

**Acknowledgments**—We thank W. R. Jacobs for providing the BCG cosmid library and plasmid pCVD442, Doug Smith for cosmid L518, and D. A. Hopwood for plasmid pIJ963.

#### REFERENCES

1. World Health Organization (1996) *The World Health Report*, World Health Organization, Geneva
2. Bloom, B. R., and Murray, C. (1992) *Science* **257**, 1055–1067
3. Young, D. B., and Cole, S. T. (1993) *J. Bacteriol.* **175**, 1–6
4. Mehra, V., Brennan, P. J., Rada, E., Convit, J., and Bloom, B. R. (1984) *Nature* **308**, 194–196
5. Vachula, M., Holzer, T. J., and Anderson, B. R. (1989) *J. Immunol.* **142**, 1696–1701
6. Chan, J., Fujiwara, T., Brennan, P., McNeil, M., Turco, S. J., Sibille, J. C., Snapper, M., Aisen, P., and Bloom, B. R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2453–2457
7. Young, D. B. (1994) in *Tuberculosis: Pathogenesis, Protection and Control* (Bloom, B. R., ed) pp. 559–567, ASM Press, Washington, D. C.
8. Minnikin, D. E. (1982) in *The Biology of the Mycobacteria* (Ratledge, C., and Stanford, J., eds) pp. 95–184, Academic Press, NY
9. Thurman, P., and Draper, P. (1989) *Acta Leprol.* **7**, 74–76
10. Daffé, M., and Lanéelle, M.-A. (1989) *Biochim. Biophys. Acta* **1002**, 333–337
11. Daffé, M., and Lanéelle, M.-A. (1988) *J. Gen. Microbiol.* **134**, 2049–2055
12. Brennan, P. J., and Nikaido, H. (1995) *Annu. Rev. Biochem.* **64**, 29–63
13. Takayama, K., and Qureshi, N. (1984) in *The Mycobacteria: A Sourcebook* (Kubica, G. P., and Wayne, L. G., eds) pp. 315–344, Marcel Dekker, Inc., NY
14. Asselineau, C., and Asselineau, J. (1984) in *The Mycobacteria: A Sourcebook* (Kubica, G. P., and Wayne, L. G., eds) pp. 345–359, Marcel Dekker, Inc., NY
15. Rainwater, D. L., and Kolattukudy, P. E. (1985) *J. Biol. Chem.* **260**, 616–623
16. Mathur, M., and Kolattukudy, P. E. (1992) *J. Biol. Chem.* **267**, 19388–19395
17. Gastambide-Odier, M. (1966) *C. R. Hebd. Acad. Sci. C Sci. Chim.* **262**, 1401–1404
18. Gastambide-Odier, M., Delaumeny, J. M., and Lederer, E. (1963) *Chem. Ind.* **1285–86**
19. Gastambide-Odier, M., Sarda, P., and Lederer, E. (1967) *Bull. Soc. Chim. Biol.* **49**, 849–863
20. Fernandes, N. D., and Kolattukudy, P. E. (1996) *Gene (Amst.)* **170**, 95–99
21. Lydiate, D. J., Ashby, A. M., Henderson, D. J., Kieser, H. M., and Hopwood, D. A. (1989) *J. Gen. Microbiol.* **135**, 941–955
22. Donnenberg, M. S., and Kaper, J. B. (1991) *Infect. Immun.* **59**, 4310–4317
23. Steinmetz, M., Le Coq, D., Aymerich, S., Gonzy-Treboul, G., and Gay, P. (1985) *Mol. Gen. Genet.* **200**, 220–228
24. Azad, A. K., Sirakova, T. D., Rogers, L. M., and Kolattukudy, P. E. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4787–4792
25. Jacobs, W. R., Jr, Kalpana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G., and Bloom, B. R. (1991) *Methods Enzymol.* **204**, 537–555
26. Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) *J. Biol. Chem.* **226**, 497–509
27. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadley, P. F. (1990) *Nature* **348**, 176–178
28. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) *Science* **252**, 675–679
29. Pelicic, V., Reytrat, J.-M., and Gicquel, B. (1996) *J. Bacteriol.* **178**, 1197–1199
30. Puzo, G. (1990) *Crit. Rev. Microbiol.* **17**, 305–312
31. McFadden, J. (1996) *Mol. Microbiol.* **21**, 205–211
32. Reytrat, J.-M., Berthet, F.-X., and Gicquel, B. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8768–8772
33. Balasubramaniam, V., Pavelka, M. S., Bardarov, S. S., Martin, J., Weisbrod, T. R., McAdam, R. A., Bloom, B. R., and Jacobs, W. R. (1996) *J. Bacteriol.* **178**, 273–279
34. Launois, P., Blum, L., Dieye, A., Millan, J., Sarthou, J. L., and Bach, M. A. (1989) *Res. Immunol.* **140**, 847–855
35. Thurman, P., and Draper, P. (1989) *Acta Leprol.* **7**, 74–76