JOURNAL OF BACTERIOLOGY, Mar. 1996, p. 1274–1282 Vol. 178, No. 5
0021-9193/96/$04.00+0

Molecular Analysis of Genetic Differences between *Mycobacterium bovis* BCG and Virulent *M. bovis*

GREGORY G. MAHAIRAS, PETER J. SABO, MARK J. HICKEY, DEVINDER C. SINGH,† AND C. KENDALL STOVER*

Laboratory of Tuberculosis and Molecular Microbiology, PathoGenesis Corp., Seattle, Washington 98119

Received 20 October 1995/Accepted 14 December 1995

The live attenuated bacillus Calmette-Guérin (BCG) vaccine for the prevention of disease associated with *Mycobacterium tuberculosis* was derived from the closely related virulent tubercle bacillus, *Mycobacterium bovis*. Although the BCG vaccine has been one of the most widely used vaccines in the world for over 40 years, the genetic basis of BCG's attenuation has never been elucidated. We employed subtractive genomic hybridization to identify genetic differences between virulent *M. bovis* and *M. tuberculosis* and avirulent BCG. Three distinct genomic regions of difference (designated RD1 to RD3) were found to be deleted from BCG, and the precise junctions and DNA sequence of each deletion were determined. RD3, a 9.3-kb genomic segment present in virulent laboratory strains of *M. bovis* and *M. tuberculosis*, was absent from BCG and 84% of virulent clinical isolates. RD2, a 10.7-kb DNA segment containing a novel repetitive element and the previously identified *mpt-64* gene, was conserved in all virulent laboratory and clinical tubercle bacilli tested and was deleted only from substrains derived from the original BCG Pasteur strain after 1925. Thus, the RD2 deletion occurred after the original derivation of BCG. RD1, a 9.5-kb DNA segment found to be deleted from all BCG substrains, was conserved in all virulent laboratory and clinical isolates of *M. bovis* and *M. tuberculosis* tested. The reintroduction of RD1 into BCG repressed the expression of at least 10 proteins and resulted in a protein expression profile almost identical to that of virulent *M. bovis* and *M. tuberculosis*, as determined by two-dimensional gel electrophoresis. These data indicate a role for RD1 in the regulation of multiple genetic loci, suggesting that the loss of virulence by BCG is due to a regulatory mutation. These findings may be applicable to the rational design of a new attenuated tuberculosis vaccine and the development of new diagnostic tests to distinguish BCG vaccination from tuberculosis infection.

Mycobacterium tuberculosis is the most prolific and poorly understood pathogen of humans. It is estimated that one-third of the world's population is infected with *M. tuberculosis* and that each year 3 million people die of this disease (28). In an effort to control the threat of tuberculosis, attenuated bacillus Calmette-Guérin (BCG) has been used as a live attenuated vaccine. BCG is an attenuated derivative of *Mycobacterium bovis*, a virulent tubercle bacillus very closely related to *M. tuberculosis* (13, 22). BCG has been used for over 5 decades to immunize over 3 billion people in immunization programs against tuberculosis. While its protective efficacy against tuberculosis has been highly variable, by recent meta-analysis estimates, on average the BCG vaccine reduced the overall risks of tuberculosis by 50% and serious forms of this disease by 70 to 80% (7). As a safe, inexpensive vaccine with potent nonspecific immunostimulatory properties, BCG has more recently been proposed and developed as a live recombinant vehicle for new multivalent vaccines against other diseases (36, 37). The original BCG Pasteur strain was developed by 230 serial passages in liquid culture and has never been shown to revert to virulence in animals, indicating that the attenuating mutation(s) in BCG are stable deletions and/or multiple mutations which do not readily revert. While physiological differences between BCG and *M. tuberculosis* and *M. bovis* have been noted (31, 38, 40), the attenuating mutations which arose during serial passage of the original BCG strain have never been identified.

In this study, we used genomic subtraction (39) to identify gross genomic differences between avirulent BCG and virulent *M. bovis* and *M. tuberculosis* strains. By this approach, three genomic regions of difference were identified, cloned, and sequenced. Complete sequence comparisons of each region from a virulent *M. bovis* strain and the BCG Connaught substrain revealed the precise junctions for each region of difference. RD3, a 9.3-kb genomic segment absent from all BCG substrains, was found in virulent laboratory strains of *M. bovis* and *M. tuberculosis*, but only in a minority of *M. tuberculosis* clinical isolates tested. RD2, a 10.7-kb DNA segment containing a novel repetitive element and the previously identified *mpt-64* gene, was found to be conserved in all virulent laboratory and clinical *M. bovis* and *M. tuberculosis* strains tested and was deleted only from BCG substrains derived from the BCG Pasteur strain after 1925, indicating that the deletion of RD2 was not a feature of the original BCG.

A 9.5-kb DNA segment, designated RD1, was absent from all BCG substrains but conserved in all virulent laboratory and clinical strains of *M. bovis* and *M. tuberculosis* tested, suggesting that the deletion of RD1 was an original attenuating mutation which arose in the derivation of BCG. The reintroduction of RD1 into BCG resulted in the repression of at least 10 proteins and induced expression of at least three additional proteins, yielding a protein expression profile almost identical to that of virulent *M. bovis* and *M. tuberculosis*. These data suggest that BCG lacks regulatory control over multiple genetic loci and is consistent with the involvement of RD1-encoded gene product(s) in a global regulatory mechanism associated with virulence. These findings may facilitate new approaches for the discovery and development of new antitu-

* Corresponding author. Mailing address: Laboratory of Molecular Microbiology, PathoGenesis Corp., 201 Elliott Ave. West, Seattle, WA 98119. Electronic mail address: kstover@bart.path.com.

† Present address: Darwin Molecular Corp., Bothell, Wash.

TABLE 1. Bacterial strains, cosmids, and plasmids used in this study

Strain, cosmid, or plasmid	Genotype and/or clone description	Source or reference[a]
Strains		
M. bovis	Wild type, virulent	ATCC 19210
M. tuberculosis Erdman	Wild type, virulent	ATCC 35801
M. tuberculosis H37Rv	Wild type, virulent	ATCC 27294
M. tuberculosis H37Ra	Attenuated derivative of H37Rv	ATCC 25177
M. bovis BCG Connaught	Avirulent derivative of *M. bovis*	ATCC 35745
M. bovis BCG Pasteur	Avirulent derivative of *M. bovis*	ATCC 35734
M. bovis BCG Brazil	Avirulent derivative of *M. bovis*	ATCC 35736
M. smegmatis mc^2155	High-efficiency transformation mutant of ATCC 19420	35
M. avium	Wild type	ATCC 25291
E. coli DH5α		Gibco-BRL
E. coli NM554		Stratagene
Cosmids and plasmids		
sCOS	Ap^r Km^r	Stratagene
pGM540	sCOS containing RD1 from *M. bovis*	This study
pGM538	sCOS containing RD2 from *M. bovis*	This study
pGM542	sCOS containing RD3 from *M. bovis*	This study
pGM592	*Bam*HI fragment left end of RD1 in pGEM5z; Ap^r	This study
pGM896	14,973-bp *Eco*RI fragment from pGM540 containing the right end of RD1 in pGEM7z; Ap^r	This study
pPS1003	*Eco*RI-*Xba*I fragment of RD1 in pGEM7z; derived by *Xba*I digestion and religation of pGM896	This study
pPS1000	9.5-kb *Eco*RI fragment containing the left-right junction of RD1 from BCG Connaught in pGEM7z; Ap^r	This study
pGM889	5.0-kb *Xho*I fragment from RD1 in pGEM5z; Ap^r	This study
pGM590	9.0-*Bam*HI fragment containing the left end of RD2 in pGEM5z; Ap^r	This study
pGM897	10.0-kb *Eco*RI fragment from pGM538 containing the right end of RD2 in pGEM7z; Ap^r	This study
pGM898	10.0-kb *Eco*RI fragment from pGM538 containing the left end and junction of RD2 in pGEM7z; Ap^r	This study
pGM895	*Pst*I digestion and religation of pGM590 containing a 2.4-kb *Bam*HI-*Pst*I fragment carrying the repetitive element of RD2 in pGEM5z; Ap^r	This study
pPS1007	4.0-kb *Xba*I fragment of the BCG chromosome left-right junction of RD2; Ap^r	This study
pGM800	3.2-kb *Eco*RI fragment from pGM542 containing the left end and repetitive element of RD3 in pGEM7z; Ap^r	This study
pGM801	2.8-kb *Eco*RI fragment from pGM542 containing a central fragment of RD3 in pGEM7z; Ap^r	This study
pGM802	3.2-kb *Not*I fragment from pGM542 containing a central fragment of RD3 in pGEM5z; Ap^r	This study
pGM821	1.8-kb *Eco*RI fragment from pGM542 containing a central fragment of RD3 in pGEM7z; Ap^r	This study
pGM893	3.0-kb *Not*I fragment from pGM542 containing the left end and junction of RD3 in pGEM5z; Ap^r	This study
pGM906	*Eco*RI fragment from BCG Connaught containing the left and right junctions of RD3 in pGEM7z; Ap^r	This study
pGM910	*Hin*dIII-*Eco*RI fragment (bases 109 to 17,494) of RD1 in pMV306; Km^r	This study

[a] ATCC, American Type Culture Collection.

berculosis drugs, vaccines, and novel diagnostic tests capable of distinguishing tuberculosis-infected individuals from BCG-vaccinated persons.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. Mycobacteria were maintained either in 7H9 medium (Difco, Detroit, Mich.) supplemented with OADC (BBL, Cockeysville, Md.) and 0.04% Tween 80 or on 7H10 agar supplemented with OADC. *Escherichia coli* strains were maintained on L agar (Difco) or in Luria broth (Difco). As necessary, media were supplemented with 20 μg of kanamycin per ml or 100 μg of carbenicillin per ml. Bacteria were stored as frozen cultures at −80°C in either 7H9 or Luria broth containing 20% glycerol.

Preparation of nucleic acids and subtracted DNA probes. High-molecular-weight (HMW) chromosomal DNA was prepared from mycobacteria by diluting a late-log-phase culture 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing incubation at 37°C for 4 to 5 days. Cells were harvested by centrifugation (5,000 × *g*), washed once in Tris-EDTA buffer (TE; pH 8.0) (29), and resuspended in 4 ml of 25% sucrose in 10× TE. Lysozyme (100 μg/ml) was added, and the preparation was incubated at 37°C for 2 h, with the subsequent addition of 100 μg of proteinase K and *N*-lauroylsarcosine to final concentration of 1%. After overnight incubation at 65°C, the mixture was extracted four times with chloroform-isoamyl alcohol (24:1), once with phenol-chloroform (1:1), and twice more with chloroform-isoamyl alcohol. The resulting HMW DNA was centrifuged through a CsCl gradient by the method of Hull et al. (19) and subsequently dialyzed against 4 1-liter volumes of TE.

Subtracted DNA probes were generated by the method of Straus and Ausubel (39). BCG Connaught HMW chromosomal DNA was sheared by passage through a 22-gauge needle to an average size of 3 to 10 kb, as determined by agarose gel electrophoresis. This DNA was biotinylated with Photobiotin (Clontech Laboratories, Inc., Palo Alto, Calif.) by the method of Straus and Ausubel (39). Sheared and biotinylated BCG Connaught DNA was used in a 10:1 excess for each round of subtraction. Wild-type *M. tuberculosis* H37Rv HMW DNA was digested with *Sau*3AI to an average size of 1 kb and hybridized to biotinylated BCG Connaught DNA for 18 h at 65°C in 1 M NaCl. Biotinylated DNA was removed from the hybridization mixture binding to avidin-coated beads, with subsequent filtration as previously described (39). After 5 cycles of subtraction, *Sau*3AI adapters were ligated to the subtraction products and amplified by 35 cycles of PCR as previously described (39). Radiolabeled probes were generated by using an adapter oligonucleotide primer and incorporating [α-^{32}P]dCTP (New England Nuclear, Boston, Mass.) with Klenow enzyme.

Recombinant DNA methods. An *M. bovis* cosmid library was constructed by the insertion of partially *Sau*3AI-digested HMW *M. bovis* chromosomal DNA into the *Bam*HI site of sCOS. Recombinant clones were packaged in vitro with Packagene extracts (Promega Corp., Madison, Wis.) and used to infect *E. coli* NM554 (Stratagene, La Jolla, Calif.). Recombinant cosmid clones were transferred to charge-modified nylon circular membranes, and membranes were prepared by the method of Grunstein and Hogness (14). Then these membranes were hybridized with radiolabeled products generated by subtraction between BCG and H37Rv DNAs. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly (5). Restriction fragments from each of the three cosmids representing individual regions were subcloned into pGEM7z or pGEM5z for deletion construction and DNA sequencing. Plasmid DNA for sequencing was prepared by using Qiagen mini columns (QIAGEN, Inc., Chatsworth, Calif.). Nested deletions were prepared by using the Erase-a-Base system (Promega Corp.) (16). Membrane filters for the analysis of clinical isolates were obtained from B. Kreiswirth at the Public Health Research Institute, New York, N.Y.

The construction of recombinant BCG carrying RD1 was made by electroporation of pGM910 into BCG Brazil (34). pGM910 was made by two successive cloning steps. The first was the cloning of a *Hin*dIII-*Eco*RI fragment from pGM592 into the *Hin*dIII-to-*Eco*RI sites of pMV306 (36). The *Eco*RI fragment of pGM896 was inserted into the only remaining *Eco*RI site, and the orientation

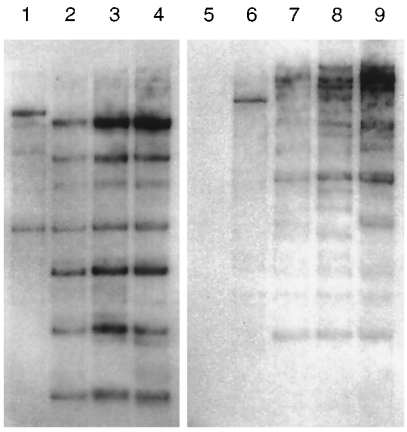

FIG. 1. Southern blot examining the specificity of the BCG Connaught-H37Rv-subtracted probe. *Eco*RI digestion (lanes 1 to 4) of chromosomal DNAs from *M. bovis* and *M. tuberculosis* strains. Lanes: 1, *M. bovis*; 2, H37Ra; 3, H37Rv; 4, Erdman. Strains were probed with the amplified and labeled product of the BCG Connaught-H37Rv subtraction. *Bam*HI digestion (lanes 5 to 9) of chromosomal DNAs hybridized with the same labeled probe. Lanes: 5, BCG Connaught; 6, *M. bovis*; 7, H37Ra; 8, H37Rv; 9, Erdman.

of the fragment was confirmed by restriction endonuclease digestion with *Hin* dIII and *Xba*I.

Labeling and 2-D electrophoretic analysis of mycobacterial proteins. Ten-milliliter cultures of mycobacteria (optical density at 540 nm = 0.1) were labeled by the addition of 125 μCi (1 Ci = 37 GBq) of L-[^{35}S]methionine (NEN). After incubation at 37°C for 1 h, unlabeled L-methionine was added to 10 mM, and the culture was cooled on ice for 10 min. The culture was harvested by centrifugation and resuspended in 300 μl of distilled H_2O with 1 mM phenylmethylsulfonyl fluoride. The culture was transferred to a 0.5-ml tube containing 300 mg of 0.1-mm-diameter glass beads and lysed by shaking in a Mini Bead-Beater apparatus (BioSpec Products, Bartlesville, Okla.) at high speed for 3 min. Lysates were mixed with isoelectric focusing buffer and analyzed by two-dimensional (2-D) gel electrophoresis (11, 27).

DNA sequencing and computer analysis. DNA sequencing was carried out on an Applied Biosystems, Inc., automated sequencer by the dye termination method, and the sequence was edited by using the manufacturer's software. Sequence assembly and further editing were done with Sequencher DNA analysis software (Genecodes, Ann Arbor, Mich.). TBLASTN and FASTA sequence homology analyses were performed by using the National Center for Biotechnology Information BLAST network service.

Nucleotide sequence accession number. The sequence data from which the information presented was derived were submitted to GenBank. A small gap of several hundred base pairs which remained refractory to sequencing exists toward the 3′ end of RD3; therefore, there are two accession numbers for it. The sequences of RD3 and the adjacent BCG homologous sequence have been deposited under accession numbers U35017 and U35018 and U35021, respectively. The sequence of *M. bovis* RD2 and the flanking BCG homologous sequence have been deposited under accession numbers U34849 and U35020, respectively. The sequence of *M. bovis* RD1 and the BCG homologous adjacent sequence have been deposited under accession numbers U34848 and U35019, respectively.

RESULTS

Genomic subtraction between BCG and *M. tuberculosis*. Subtractive hybridization between avirulent BCG Connaught and virulent *M. tuberculosis* H37Rv was used to isolate sequences present in virulent mycobacteria and absent from BCG. The specificity of the subtracted and amplified H37Rv DNA as a probe was examined by Southern blot analysis with BCG and virulent *M. tuberculosis* and *M. bovis* chromosomal DNAs (Fig. 1). The subtracted radiolabeled probe failed to hybridize with BCG (Fig. 1, lane 5), indicating that biotinylated BCG DNA and BCG homologous sequences were efficiently removed and that sequences of difference were both enriched for and amplified. The subtracted probe did hybridize strongly with *M. bovis* (Fig. 1, lanes 1 and 6) and *M. tuberculosis* (lanes 2 to 4 and 7 to 9). *M. tuberculosis* H37Ra (Fig. 1, lanes 2 and 7) produced a banding pattern similar to that of virulent *M. tuberculosis* Erdman (lanes 4 and 9) and H37Rv (lanes 3 and 8). Subtractive hybridization between H37Rv and H37Ra did not yield a probe that could distinguish between these two strains, suggesting that the genetic differences between H37Rv and H37Ra are point mutations or small deletions undetectable by this subtractive method. Comparisons of the hybridization patterns of restricted genomic DNA between *M. bovis* and *M. tuberculosis* revealed a subset of common restriction fragments hybridizing to the labeled subtracted probe. *M. bovis* exhibited a less complex hybridization pattern with the subtracted probe than did *M. tuberculosis*. Additional *M. tuberculosis* DNA observed to hybridize with the subtracted probe may reflect restriction fragment length polymorphisms or may indicate more gross genetic differences between *M. bovis* and *M. tuberculosis.*

Molecular cloning and analysis of *M. bovis* genomic regions absent from BCG. To isolate and clone the genetic differences between *M. bovis* and BCG, a cosmid library of *M. bovis* chromosomal DNA was constructed in sCOS and screened with a radiolabeled BCG Connaught-H37Rv subtraction product as the probe. From 600 independent cosmids, 18 reactive clones were identified and characterized by restriction endonuclease and Southern hybridization analyses. It was possible to group the 18 independent cosmid clones into three groups on the basis of the presence of coincident *Eco*RI and *Not*I restriction fragments hybridizing to the subtracted probe. These three cosmid groups presumably represented three distinct genomic regions of difference between BCG and *M. bovis*. These regions were subsequently designated *M. bovis* RD1, RD2, and RD3 and contained in their entirety in cosmids pGM540, pGM538, and pGM542, respectively. Each representative cosmid was used as a probe for Southern blot analysis of chromosomal DNA restriction digests with *Eco*RI since sCOS has two *Eco*RI sites that flank the *Bam*HI cloning site. The patterns obtained by this analysis were compared with restriction fragment patterns obtained for each cosmid to confirm that the selected cosmids were not composed of scrambled discontinuous genomic segments and to identify terminal fragments in each cosmid insert. Fragments from each of the three cosmids identified by hybridization with the subtracted probe were subcloned into pGEM7z or pGEM5z for the construction of nested deletions and DNA sequencing. Significant subclones from which all the deletion constructs for sequencing were made are described in Table 1. Table 2 summarizes the sequence analysis by identifying open reading frames (ORFs), possible ribosomal binding sites, molecular weights and pIs of potential encoded proteins, and possible homologous proteins derived from six-frame translations of the nonredundant DNA database. Physical maps of the three *M. bovis* regions, derived from DNA sequence data, are presented in Fig. 2. Homologous chromosomal fragments in BCG were identified by Southern blotting, with labeled cosmids carrying each region of difference as probes. Restriction fragments that were polymorphic in comparison to *M. bovis* were presumed to carry DNA representing the junctions of the chromosomal regions of difference. These junction fragments were subsequently cloned and sequenced. The precise limits of the regions of difference were determined by comparing the sequences from the BCG chromosome in plasmids pPS1000, pPS1007, and pGM906 with the sequences in RD1, RD2, and RD3, respectively (Fig. 3).

Analysis of *M. bovis* RD3. *M. bovis* RD3 contains at least 10 ORFs (Fig. 2) with GC-rich codon usage profiles consistent with those of expressed *M. tuberculosis* genes. Possible homology to *Streptomyces* phage phi-C31 were observed for ORFs 3D and 3E. ORFs 3I, 3J, and 3K exhibited highly significant

TABLE 2. Summary of sequence analysis of *M. bovis* RD1, RD2, and RD3[a]

ORF	ORF Size (base pairs)	Start - Stop (base pairs)	Possible Ribosome Binding Sites	Encoded Protein (~kDa - pI)	Homologies to Predicted Encoded Protein	P value	Homolog Accession #
1A	1775	889 - 2664	AGGA (10)	57 - 6.6	none		
1B	1106	3137-4243	GGA (4) GGA (9)	36 - 5.4	*M.leprae* aceA cosmid BCG uraA downstream orf	1.4e-14 3.0e-13	Z46257 U01072
1C	287	4671-4958	GGA (8)	10 - 4.3	*M.tuberculosis* esat6	2.3e-43	X79562
1D	2000	5072-7072	GAGG (5)	34 - 11.0	none		
1E	1439	7165-8604	none	59 - 9.3	none		
1F	842	8755-9597	GGA (5)	34 - 5.4	none		
1G	1661	9817-11478	AGGA (9)	48 - 4.7	none		
1H	1370	16900-15530		46 - 4.9	none		
1I	1388	15533-141145		50 - 9.9	none		
1J	1382	14038-12656	GAA (5)	46 - 4.7	*B.subtilis* subtilisin Serine proteases	3.6e-16 4.8e-14	L29506 A08331
1K	2240	11895-9655	GGGGAGGG (10)	76 - 5.4	none		
2A	560	1435 - 1995	AGGGAG (7)	25 - 12.2	none		
2B1	437	2470 - 2907	AGAA (4)	16 - 12.45	none		
2B2	590	2609 - 3199	none	34 - 12.42	none		
2C	950 761	4793 - 5743 4982 - 5743	AG? (8) GGA (8)	34 - 9.4	*E.coli* iciA lysR family	9.9e-47 <1e-5	P24194
2D	659 653	6167 - 6826 6173 - 6826	none	22 - 5.9	*Mleprae* cosmid B1620 ORF Cutinases	1.5e-7 ~4.e-5	U00015 A00975 U03393
2E	488 467	7642 - 8130 7663 - 8130	none	19 - 12.4	none		
2F	968	9546 - 10514	AGGA (11)	37 - 4.7	*S.typhimurium* RNDPR proUVWX	9.9e-146 2.7e-36	X73226 X17445
2G	668	10723 - 11391	AAGA (6)	24 - 4.4	*M.tuberculosis* mpt64	6.7e-141	A30545
2H	1445	11570 - 13015	AG (10)	51 - 10.9	*E. coli* gabP permease *S. typhimurium* asp permease	3.1e-11 1.4e-08	X65104 U04851
2I	848	13826 - 12978	GGAAGA (6)	31 - 6.9	none		
2J	1052	7866 - 6814	GAG (10)	35 - 4.0	none		
2K	668	4598 - 3930	none	25 - 12.6	none		
2L	599	4723 - 4124	AG (10)	21 - 10.45	*E. coli* hypothetical protein in fda and gapB region	7.1e-41	P11667
3A	1143	613 - 1755	none	42 - 11.48	MTB mce sau3A *M. leprae* cosmid L247	2.9e-64 5.1e-13	X70901 U00021
3B	1347	1214 - 2560		49 - 10.34	Actinophage R4 attP gene *B. subtil.* site spec. recomb. recombinases / invertases	3.0e-05 7.8e-4 8.2e-4	D90361 M29040 K00676 X01805, X07724
3C	513	2820 - 3332	GGA (6)	19 - 7.30	none		
3D	924	4007 - 4930 4070 - 4930	none GGA (7)	34 - 7.05	*S. coelicolor* phage phi-C31 early region	4.2e-26	X76288
3E	543	4795 - 5337 4915 - 5337	none GGAA (5)	21 - 7.50	*S. coelicolor* phage phi-C31 pglY pglZ genes	3.2e-11	X76288 L37531
3F	576	5639 - 6214	GA (9)	20 - 6.46	none		
3G	510	7253 - 7762 7285 - 7762	GAAGG (8)	19 - 5.19	none		
3H	1330	7868 - 9197	GA (8)	47 - 5.60	none		
3I	1665	10146 - 11810 10164 - 11810	GGAAG (6)	58 - 12.34	*M. leprae* cosmid B1170	6.2e-69	U00010
3J	918	12319 - 11402	GAA (11)	30 - 8.73	*M.leprae* bioDAYB *C. glutamicum* bioD	6.9e-53 2.6e-08	U00010 D14083
3K	702	11594 - 10893	AGG (4)	25 - 11.87	*M. leprae* cosmid B1170	1.0e-81	U00010
3L	660	10147 - 9488	AG (10)	25 - 12.30	none		

[a] Each parenthetical value in the distance (in base pairs) from the indicated putative ribosome-binding site to the start codon (e, exponent).

homologies (Table 2) to ORFs in the sequenced *M. leprae* cosmid B1170. An ORF straddling the BCG deletion junction of RD3 shows striking homology to the reported sequence of a region encoding the *M. tuberculosis mce* gene. This region was previously reported to encode genes for a putative invasin and macrophage survival determinant. However, the extent of the homology between ORF3A and the *mce* region sequence reported by Arruda et al. (2) is limited to an internal *Sau*3AI fragment in the previously reported sequence (GenBank accession number X70901), suggesting that restriction fragment scrambling occurred in the construction of the plasmid library. Since the cosmids generated in this study were also generated

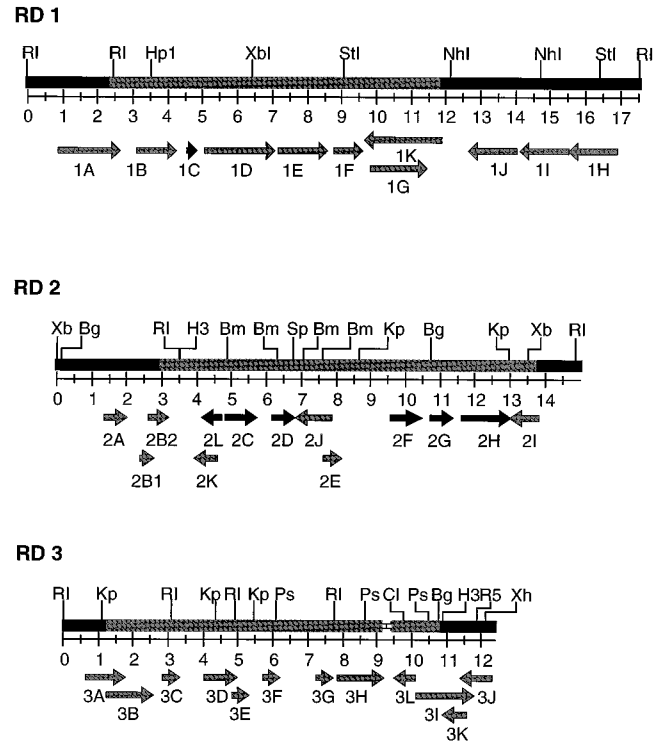

FIG. 2. Restriction and ORF maps of RD1, RD2, and RD3. These maps are abbreviated restriction maps and incorporate the predicted ORFs. Each ORF is indicated by an arrow and listed in Table 2. Common *BCG* and *M. bovis* regions are shown as black lines, and unique *M. bovis* regions (BCG deleted) are shown as gray lines. ORFs for which significant homologies have been found are shown as black arrows, while those with no significant homologies are shown as gray arrows.

by partial *Sau*3AI digests, we confirmed the fidelity of our sequence by sequencing an independently cloned restriction fragment from *M. bovis* and BCG. It has been confirmed that the previously reported sequence is derived from scrambled restriction fragments (28a). However, the sequence of one ORF (bp 181 to 807), that of the *mce* gene, is not affected by this scrambling; the downstream region potentially encoding a determinant for macrophage survival (from bp 866 to the end of the fragment) is affected. Southern blot analysis indicated the presence of a repetitive element located in the BCG homologous sequence near the 5′ end of RD3. This repeat is present in three to four copies in BCG, *M. bovis*, and *M. tuberculosis* (data not shown). The core sequence of this repeat has not been determined. By using the *Not*I fragment of pGM802 as an RD3-specific probe, RD3 was shown to be present in *M. tuberculosis* Erdman and *M. bovis* and absent from BCG (Fig. 4 and Fig. 5). Figure 4 shows that RD3 is also absent from BCG substrains Connaught, Pasteur, and Brazil. Similar Southern blots with the 1.8-kb *Eco*RI fragment of pGM821 (RD3) as the probe show identical patterns of hybridization for *M. bovis* and *M. tuberculosis* (data not shown). This indicates that few (if any) polymorphisms exist for RD3 between *M. bovis* and *M. tuberculosis*. The prevalence of RD1, RD2, and RD3 in clinical isolates of *M. tuberculosis* was analyzed by Southern blotting with region-specific radiolabeled probes (data not shown). Hybridizations with RD3-specific probes indicated that only 10 of 62 (16%) *M. tuberculosis* clinical isolates carried RD3, suggesting that RD3 is not required for virulence. It is possible that the original BCG strain was derived from an *M. bovis* strain that did not contain RD3 or that a deletion in RD3 arose during the derivation of BCG.

Analysis of *M. bovis* RD2. The *M. bovis* RD2 region extends from base 2,888 through base 13,674 of the 14,844 bp sequenced. The region contains 11 ORFs with GC-rich codon usage consistent with that of *M. tuberculosis* and *M. bovis* expressed genes (Fig. 2). Database homology searches with derived protein sequences from predicted ORFs revealed a number of interesting homologies. The derived peptide sequence of ORF2C bears striking similarity over its entire length to the IciA protein of *E. coli* (17, 20, 21, 41) and similarities of lesser degree to other members of the *lysR* family of transcriptional regulators. Each *lysR* family member possesses a helix-turn-helix motif (17), and this feature is also conserved in the putative protein encoded by ORF2C (data not shown). ORF2F potentially encodes a protein with strong homology to a *Salmonella typhimurium* ribonucleotide diphosphate reductase that catalyzes the reduction of ribonucleosides to their deoxy derivatives. The derived peptide sequence from ORF2H exhibits limited homology to the GabP aromatic amino acid permease of *E. coli*, while ORF2L has homology to a hypothetical protein encoded by a hypothetical ORF near the gene encoding GabP. ORF2G encodes the previously identified Mpt-64 protein antigen, which was investigated as an *M. tuberculosis*-specific secreted antigen not expressed in BCG (24).

Southern blot analysis with a probe from between bp 7,591 and 9,945 of RD2 identified a repetitive element present in 10 or more copies on the chromosomes of BCG, *M. tuberculosis*, and *M. bovis* (Fig. 6). The presence of this repetitive element confounded the initial characterization of RD2 by restriction mapping and Southern blotting and necessitated the use of flanking sequences as region-specific probes. RD2 contains several additional small repeated sequences; the most notable are two imperfect direct repeats (at bp 2,864 to 2,888 and 13,652 to 13,675) (Fig. 3) positioned exactly at the deletion breakpoint of the *M. bovis* sequence. These sequences have the potential for forming secondary structures, providing a possible mechanism for the evolution of this deletion (3, 10, 18, 23, 25, 26). The presence of RD2 in several BCG substrains was examined by using probes for the 3′ portion (a *Pst*I fragment containing *mpt-64* from pGM590) and the 5′ portion (an *Eco*RI-*Sph*I fragment [coordinates bp 3,484 to 6,262]) of RD2 (Fig. 4B and C, respectively). The results demonstrate that RD2 is present in *M. tuberculosis* Erdman, *M. bovis*, and BCG Brazil. RD2 was also shown to be present in BCG Japanese (data not shown). Also evident in Fig. 4 is an apparent restriction fragment length polymorphism between *M. bovis* and *M. tuberculosis*; this difference is also evident in BCG substrains carrying RD2. Southern blot analyses with two distinct probes for both the 5′ and 3′ ends of the central *Eco*RI fragment of RD2 show that the *M. bovis Eco*RI fragment is about 2 kb smaller than its *M. tuberculosis* homolog (Fig. 4B and C). With the exception of the 5′ end, RD2-specific probes failed to hybridize with *M. smegmatis* or *M. avium* (Fig. 5B). An RD2-specific probe with flanking *Eco*RI and *Sph*I sites (bp 3,484 to 6,262) that includes *iciA* and two surrounding ORFs gave a faintly discernible band that may be an *iciA* homolog in these species (Fig. 5, lanes 3 and 4). RD2 was found in 80 of 80 clinical isolates of *M. tuberculosis*, Erdman, H37Rv, and *M. bovis*. These data suggest that the deletion of RD2 is not a mutation which arose during the original derivation of BCG, but they do not rule out a possible role for RD2 in virulence.

Analysis of *M. bovis* RD1. Of the 17,499 bp sequenced in this region, *M. bovis* RD1 extends from base 2,328 through base 11,832. The presence of RD1 in virulent laboratory strains of *M. bovis* and *M. tuberculosis* and several BCG substrains was

```
Region 1

BCG     AGAAGCGGTT GCCGCCGACC GACCTGACGA CGGCGCAGCT ACGCTCGCGT TCGTGGTGGA GCGGATTTGA
bovis   AGAAGCGGTT GCCGCCGACC GACCTGACGA CGGCGCAGCT ACGCTCGCGT TCGTGGTGGA GCGGATTTGA

BCG     CGTCGTGCTT CTGGTCGACG ATTGGCACAT
bovis   CGTCGTGCTT CTGGTCGACG ATTGGCACAT GATCGTGGGT GCCGCCGGGG GGATGCCGCC GATGGCACCG
                                       (2,328)
BCG                                                                          -
bovis   CTGGCCCCGT TATTGCCGGC GGCGGCAGAT .../  \... CATCGGTGAC CCTTTGCAAA ACCTGGCTAT

BCG                                                CACCCAGCCG CCCGGATCCA GCATCTGTCT
bovis   ATTCCTGGGC CCGGTCATAG AAAGTGTCTT CATCGGCTTC CACCCAGCCG CCCGGATCCA GCATCTGTCT
                                                    (11,832)
BCG     GGCATAGCTG CCCGTCGGCC TGGTAATACT CATCCCCTAC TGCCCTCCCC AAACCGCCAG ATCGCCTCGC
bovis   GGCATAGCTG CCCGTCGGCC TGGTAATACT CATCCCCTAC TGCCCTCCCC AAACCGCCAG ATCGCCTCGC


Region 2

BCG     CAACCGGATC GAGGCGGCGT CCGCGTGCAC CACGGTAATG CCAGGGAATC GCTCGCGGAG GACACCGACT
bovis   CAACCGGATC GAGGCGGCGT CCGCGTGCAC CACGGTAATG CCAGGGAATC GCTCGCGGAG GACACCGACT

BCG     CGTCGCGGGGT GCAACTCCAC GGCGACCACC
bovis   CGTCGCGGGGT GCAACTCCAC GGCGACCACC CGCGCCCCCG CTCGCACTAG ATGCGCCGTC AGTGCCCCTT
              (2,864)                    (2,888)
BCG
bovis   CGCCGGCGCC GATGTCAAAC ACGAGCTCAC .../  \... AACAAGTTTC AGGGCTATCC GGGCGTACTT

BCG                                                CCGATGATCT TCTGTTTGAC CTCATCGGGC
bovis   CTCGTGCAGG TTGGTCAGGG TGCCCACCCG GTCGAGCACC CCGATGATCT TCTGTTTGAC CTCATCGGGC
                      (13,652)                     (13,675)
BCG     ACATCGTCGC GGTCGAGGTA CTCCAGTGCG TCGGTCTGGA ATCGACGATC CAGCCAAGAC GCGTCGGGGC
bovis   ACATCGTCGC GGTCGAGGTA CTCCAGTGCG TCGGTCTGGA ATCGACGATC CAGCCAAGAC GCGTCGGGGC


Region 3

BCG     ACCCCGACAA CCGACTCGCC GAAAAAGGCT GGACCACCCG CAAAAACACC CACGGCCACA CCGAATGGCT
bovis   ACCCCGACAA CCGACTCGCC GAAAAAGGCT GGACCACCCG CAAAAACACC CACGGCCACA C.GAATGGCT

BCG     ACCACCACCC CACCTCGACC ACGGCCAACC
bovis   ACCACCACCC CACCTCGACC ACGGCCAACC GTGGACCTGT GAGATACACT ACACCTGTGC GTGCTGCTGT
                    (1,388)              (1,400)
BCG
bovis   CTACCTCCGA ATCTCAGAAG ACCGCTCCGG .../  \... AAACGTGCTG GTGTGGTGGT CATGGTTGAT

BCG                                                CCGCACCAAC ACCTTCCACC ACCACGAGAA
bovis   CTCCTGGCGT GGAATGTTCT TCAGCAGTCC ACGGCCAACC CCGCACCAAC ACCTTCCACC ACCACGAGAA

BCG     GCTGCTACGC CACAACGACG AGGACAACCA CGACGATCCG TGAGAATCGC CGCCCGCGAA GATCTTTGGA
bovis   GCTGCTACGC CACAACGACG AGGACAACCA CGACGATCCG TGAGAATCGC CGCCCGCGAA GATCTTTGGA
```

FIG. 3. Sequence comparisons of BCG and *M. bovis* at the region of difference junctions showing breakpoints. The exact bases for the beginning and end of each region are indicated. RD2 and RD3 have direct repeated sequences at the regional breakpoints that are in bold type.

examined by Southern hybridization with the *Eco*RI-*Xba*I fragment from pPS1003 as the probe (Fig. 4A and 5A). RD1 is present in *M. tuberculosis* Erdman and *M. bovis* but absent from BCG Connaught, Pasteur, and Brazil. RD1 is also absent from BCG Tice, Japanese, and Danish (data not shown). RD1 homologous sequences were not found in *M. smegmatis* and *M. avium* when examined by Southern hybridization under moderately stringent conditions (6× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 70°C) (Fig. 5A). With the *Eco*RI-*Xba*I fragment from pPS1003 as the probe, RD1 was found to

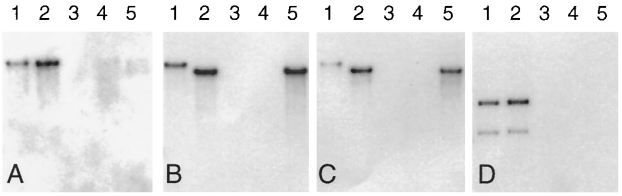

FIG. 4. Southern blots with region-specific DNA probes examining the presence or absence of each RD. All panels contained *Eco*RI-digested chromosomal DNA. Lane 1, *M. tuberculosis* Erdman; lane 2, *M. bovis*; lane 3, BCG Connaught; lane 4, BCG Pasteur; lane 5, BCG Brazil. (A) Probe was an *Eco*RI-to-*Xba*I fragment covering bases 2,523 to 6,432 of RD1. (B) Hybridization with a *Pst*I fragment covering bases 9,945 to 11,475 of RD2 and including *mpt-64*. (C) Hybridization with an *Eco*RI-*Sph*I fragment covering bases 3,484 to 6,262 of RD2 and including *iciA*. (D) Hybridization with an *Eco*RI fragment covering bases 3,075 to 4,907 of RD3.

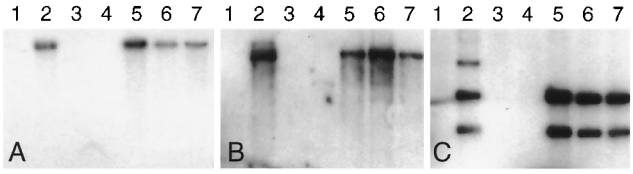

FIG. 5. Southern blots with region-specific probes examining the presence of each region of difference in diverse mycobacteria. *Eco*RI chromosomal DNAs from BCG Connaught (lane 1), *M. bovis* (lane 2), *M. smegmatis* (lane 3), *M. avium* (lane 4), *M. tuberculosis* Erdman (lane 5), *M. tuberculosis* H37Rv (lane 6), and *M. tuberculosis* H37Ra (lane 7) were probed with the *Eco*RI-*Xba*I fragment of pPS1003 (A), the *Eco*RI-*Sph*I fragment of RD2 from bp 3,484 to 6,262 (B), and the *Not*I fragment of pGM802 (as an RD3-specific probe) (C).

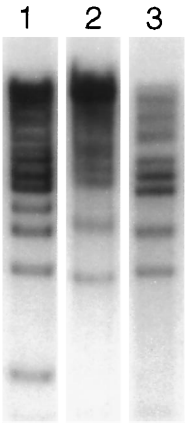

FIG. 6. Southern hybridizations of *Eco*RI-digested chromosomal DNAs of *M. tuberculosis* Erdman, *M. bovis*, and BCG Connaught (lanes 1 to 3, respectively) probed with the *Bam*HI-*Pst*I fragment of pGM895 containing the repetitive element and surrounding sequence of RD2.

be conserved in 62 of 62 clinical isolates (data not shown), the Erdman and H37Rv strains of *M. tuberculosis*, and *M. bovis.*

RD1 encodes at least eight ORFs with codon usage patterns consistent with those of expressed genes of *M. bovis* and *M. tuberculosis* (Fig. 2 and Table 2). ORF1A has a 47-bp direct repeat immediately preceding the translational start codon that is homologous to the region upstream of the start codon of *mtrB. mtrB* has been implicated as part of a two-component signal transduction system in mycobacteria whose role in an active regulatory mechanism is yet to be reported (9). ORF1B has 22% sequence homology to an ORF downstream from the *M. tuberculosis* 65-kDa antigen (32, 33). A small 287-bp ORF, designated ORF1C, was recognized as the previously identified *esat6* gene. The *esat6* gene product is a small secreted protein under investigation as an immunogen eliciting immunity to *M. tuberculosis* (1, 6, 15). An ORF, designated ORF1J, with significant homology to a *Bacillus subtilis* subtilisin was identified in DNA flanking RD1.

Analysis of the effects of regions of difference on BCG gene expression. To evaluate RD2-encoded proteins, 2-D gel electrophoretic analysis was used. A comparison of BCG Brazil (containing RD2) with BCG Connaught revealed three obvious protein differences which distinguish these two substrains (Fig. 7). The molecular weights and pIs of these proteins are consistent with the molecular weights and pIs predicted for proteins encoded by ORFs 2D, 2F, 2G, 2I, and 2J. Because BCG Brazil contains a complete RD2, it was used for subsequent comparisons with virulent tubercle bacilli. Despite the identification of sizable genomic regions absent from BCG (RD1 and RD3) but present in *M. bovis* and *M. tuberculosis*, BCG expresses at least 10 additional proteins and appears to express higher levels of many proteins.

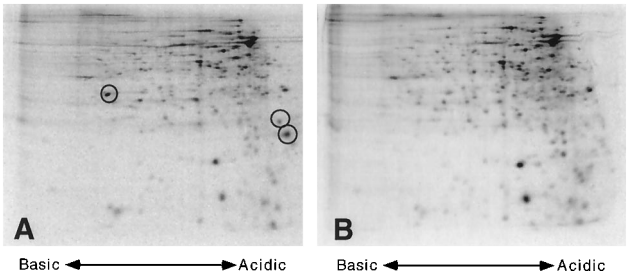

FIG. 7. Comparative 2-D protein gel electrophoresis of [^{35}S]methionine-labeled mycobacterial proteins from BCG Brazil (A) and BCG Connaught (B). BCG Brazil contains RD2, but BCG Connaught does not. Differences are highlighted by circles.

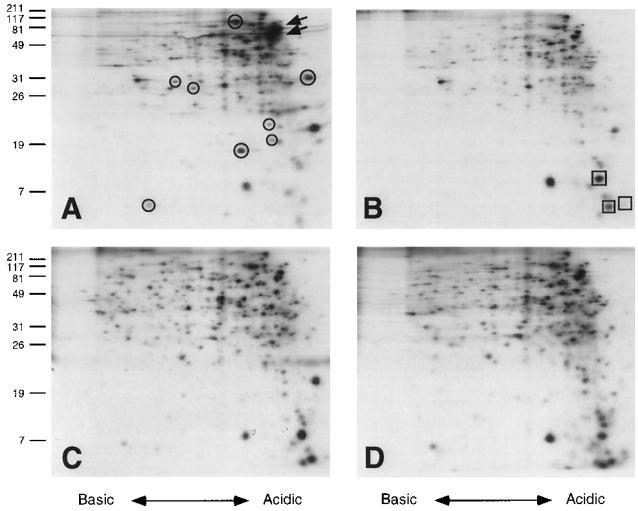

FIG. 8. Comparative 2-D protein gel electrophoresis of [^{35}S]methionine-labeled mycobacterial proteins from BCG Brazil, which contains RD2 (A); recombinant BCG Brazil carrying RD1 (B); *M. bovis*, which carries all three regions of difference (C); and *M. tuberculosis* Erdman (D). Regulated protein differences between BCG and virulent *M. bovis* are highlighted by circles. Additional proteins expressed as a result of the introduction of RD1 are highlighted by boxes. Upregulated heat shock proteins are designated by arrows. The positions of molecular mass standards (in kilodaltons) are given on the left.

In order to characterize the effect of reintroducing RD1 into BCG, the entire region was cloned into pMV306 (37), a site-specific, integrating, mycobacterial shuttle vector. The resulting clone, pGM910 (Table 1), was then introduced into the Brazil strain of BCG (35). The resulting recombinant BCG was compared with BCG, *M. tuberculosis*, and *M. bovis* by 2-D protein gel electrophoresis. As seen in Fig. 8, *M. bovis* and *M. tuberculosis* show strikingly similar protein profiles. However, *M. tuberculosis* expresses a low-molecular-weight protein not observed in *M. bovis* or BCG, while BCG and *M. bovis* express an acidic protein of approximately 20 kDa that is not observed in *M. tuberculosis*. Wild-type *M. bovis* and BCG::RD1 show identical profiles, with the exception of a single low-molecular-weight acidic protein (Fig. 8C). Two proteins which are apparently upregulated severalfold (Fig. 8A [arrows]) are consistent with the known isoelectric points, molecular weights, and expression levels of Hsp60 and Hsp70. Consistent with this finding is the result that gene fusions driven from the Hsp60 promoter with firefly luciferase are threefold higher in BCG than those in *M. tuberculosis* (data not shown). Additionally, BCG::RD1 shows the disappearance of at least 10 other proteins. BCG::RD1 also expresses several additional proteins (Fig. 8B) that may be RD1-encoded proteins; the molecular weight and pI of one of them are consistent with those of the *esat6* gene product. There are a number of very small ORFs in RD1 that could encode low-molecular-weight proteins but are not shown as ORFs in Fig. 2. An additional low-molecular-weight protein seen in *M. bovis* and *M. tuberculosis*, not in either BCG::RD1 or BCG Brazil, may be an RD3-encoded protein, since RD3 is the only region of difference between *M. bovis*, *M. tuberculosis*, and BCG::RD1.

DISCUSSION

Genomic subtractive hybridization was used to identify gross genetic differences between avirulent BCG and virulent *M. tuberculosis* and *M. bovis*. *M. tuberculosis* and *M. bovis* are closely related members of the tuberculosis complex, and because of their genetic identity, it has been suggested that they be considered strains of the same species (13, 22). The genomic subtraction between BCG Connaught and *M. tuberculosis* H37Rv performed in this study would be expected to identify gross genetic differences which might exist between BCG and *M. tuberculosis* and additional differences which might exist between *M. bovis* and *M. tuberculosis*. Since BCG is derived from *M. bovis* and the subtracted probe revealed fewer apparent differences between BCG and *M. bovis*, we first focused on the genetic differences between BCG and virulent *M. bovis*. Three large chromosomal regions present in virulent *M. bovis* but absent from BCG Connaught were identified. Southern hybridization showed all three regions to be present in virulent laboratory strains of *M. tuberculosis* (H37Rv and Erdman) but likely absent from *M. avium* and *M. smegmatis*, indicating that these genetic differences were likely unique to tuberculosis complex organisms. RD1 and RD3 were completely absent from all BCG substrains examined.

The origin of the 9.3-kb RD3 is not clear, since it appears to be present in laboratory strains of *M. bovis* and *M. tuberculosis* but absent from the great majority of clinical isolates. Limited homologies to some *Streptomyces* phage genes may suggest an origin for this polymorphic region, but a role for RD3 in virulence is doubtful. The presence of homologies to *Streptomyces* phages raises the possibility of phage-mediated genetic transfer between *Streptomyces* species and pathogenic mycobacteria.

The finding that RD2 is present in some BCG strains and absent from others indicates heterogeneity among BCG substrains and confirms previous immunological findings focusing on the MPB64 (Mpt-64) protein antigen. MPB64 was thought to be an *M. tuberculosis*-specific antigen, but subsequent analysis of BCG substrains revealed heterogeneity among them with respect to the presence and expression of the *mpt-64* gene (8, 24). BCG Brazil and the Japanese and Russian substrains, all of which express the Mpt-64 antigen encoded by ORF2G, were derived from the original BCG Pasteur strain prior to 1925. RD2 is apparently absent from BCG substrains derived from the original BCG Pasteur strain after 1925; therefore, a significant change in the BCG Pasteur strain seed lot occurred after 1925 and prior to 1934, when the next BCG substrain, Tice, was derived. A mechanism by which RD2 may have been deleted is provided by the presence of numerous small repeated sequences existing at the boundaries of RD2. The formation of hairpin secondary structures composed of indirect repeats may have resulted in the deletion of this region (3, 10, 18, 23, 25, 26). Phenotypic differences between BCG substrains which differ by the presence or absence of RD2 are not readily apparent but have not been evaluated in depth. It should also be noted that BCG substrains carrying a complete RD2 (e.g., Brazil and Japan) have been widely used as tuberculosis vaccines, as have BCG substrains lacking RD2. However, differences in vaccine efficacy which might be associated with the presence or absence of RD2 cannot be ascertained because differences in BCG vaccine efficacy associated with BCG substrains have never been consistently observed (8).

Homology searches of the nonredundant genetic database provided few clues to the function of potential coding regions. Of the putative ORFs encoded in RD2, the homology between ORF2C and IciA is most striking. In *E. coli*, IciA negatively controls the initiation of chromosomal replication and is a member of the LysR family of regulatory proteins (17, 20, 21, 30, 41). The helix-turn-helix motif of the IciA homologous protein is 100% conserved at the amino acid level, indicating that this protein probably binds DNA. Although IciA is a member of the LysR family, it has been shown to inhibit the initiation of chromosomal DNA replication by binding at the origin of replication and has yet to be implicated in transcriptional regulation (17, 20, 21, 41). RD2 also carries a novel sequence that is present in multiple copies on the chromosome and capable of differentiating between IS*6110* types when used as a probe. We are presently evaluating the usefulness of this element as an epidemiological tool.

One chromosomal region (RD1) was found to be present in 100% of *M. tuberculosis* clinical isolates and virulent laboratory strains (H37Rv and Erdman) but absent from all BCG substrains. This region encodes *esat6* and seven additional ORFs, of which one may be a regulatory gene. The putative regulatory gene encoded within RD1 is apparently deleted from or nonfunctional in all BCG substrains. As determined by 2-D gel electrophoresis, the reintroduction of RD1 appears to strongly repress the expression of at least 10 proteins and downregulate the expression of many additional cellular proteins. The possible role of this regulation in the virulence of *M. bovis* and *M. tuberculosis* is under investigation. Although a definitive link between this apparent deletion and the attenuation of BCG has not been established, several possible rationales, based on our observations, are offered for this attenuation. One or more RD1 genes are directly involved in virulence, or one or more genes directly involved in the virulence of pathogenic mycobacteria may be regulated by an RD1-encoded gene. The lack of regulation due to a RD1 deletion results in the attenuated phenotype of BCG. This hypothesis is not without precedent, as the constitutive expression of *phoP* in *S. typhimurium* also results in an attenuated phenotype (4). In BCG, the lack of tightly controlled expression of specific proteins may result in the display of a number of immunogenic proteins to the immune system, and an immune response to these proteins results in the clearance of the infection. Studies of 2-D gels and reporter genes indicate that several of these regulated proteins may be heat shock or stress proteins, and a protective immune response to heat shock proteins has been documented (42, 43). The disrupted regulation of a stress response may also affect the ability to adapt and survive in the host and to cause disease. Higher levels of expression and the expression of additional proteins could also be due to a feedback mechanism caused by the loss of an RD1 gene(s). The mechanisms by which the RD1 mutation may cause these effects are under investigation.

Although a survey of clinical isolates clearly demonstrates that the intact RD1 is conserved in virulent tubercle bacilli, the possibility that RD1-encoded gene products have no role in virulence also exists. In a preliminary animal experiment, BCG::RD1 did not show enhanced in vivo growth in mice. However, in vivo growth rates among virulent tubercle bacilli vary widely and do not determine relative virulence, making it necessary to evaluate the pathology associated with infection (12). Initial data indicate that the reintroduction of RD1 may affect BCG's ability to elicit pathology in a way similar to that of virulent *M. bovis*, but further study is necessary to more critically evaluate this effect (data not shown). The long and complex passage history of the BCG vaccine strain and subsequent substrain derivations make it likely that most BCG substrains carry multiple mutations, some of which could also affect in vivo growth and virulence. Repetitive passage under unchanging environmental conditions may have resulted in one or more mutations in regulatory function(s). The full res-

toration of virulence in BCG and the fulfillment of Koch's molecular postulates may require the repair of additional BCG mutations not detected in this study, and the restoration of RD1 may be one incremental step toward restoring virulence in BCG. Such information may be useful for the rational design of a more effective and defined live attenuated tuberculosis vaccine since it may be possible to improve the BCG vaccine incrementally by replacing defective genes encoding potentially protective antigens without restoring virulence. The identification of genetic differences between virulent tubercle bacilli and BCG could also lead to diagnostic tests which can discriminate between BCG immunization and infection with virulent tubercle bacilli. It is also hoped that the identification of critical genetic differences between virulent and avirulent mycobacteria of the tuberculosis complex will provide information about and insights into their pathogenic mechanisms, thus supplying a basis for the development of rational approaches and new treatments for tuberculosis.

ACKNOWLEDGMENTS

We thank Lynn Rose, David Sherman, Taraq Arain, and Jeff Reid for critical reviews of the manuscript; Jim Wallace for help with computer analysis of sequence data; Frank Collins for helpful discussions concerning the history of BCG; and the laboratory of Cliff Barry for help with 2-D gel electrophoresis methodology. We also thank Lee Gordon and Robert Malcolm of the PathoGenesis sequencing facility.

REFERENCES

1. **Andersen, P., A. B. Andersen, A. L. Sorenson, and S. Nagai.** 1995. Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. J. Immunol. **154:**3359–3372.
2. **Arruda, A., G. Bomfim, R. Knights, T. Huima-Byron, and L. W. Riley.** 1993. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. Science **261:**1454–1457.
3. **Balbinder, E.** 1993. Multiple pathways of deletion formation in *Escherichia coli*. Mutat. Res. **299:**193–209.
4. **Behlau, I., and S. I. Miller.** 1993. A PhoP-repressed gene promotes *Salmonella typhimurium* invasion of epithelial cells. J. Bacteriol. **175:**4475–4484.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
6. **Boesen, H., B. N. Jensen, T. Wilcke, and P. Andersen.** 1995. Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*. Infect. Immun. **63:**1491–1497.
7. **Colditz, G. A., T. F. Brewer, C. S. Berkey, M. E. Wilson, E. Burdick, H. V. Fineberg, and F. Mosteller.** 1994. Efficacy of BCG vaccine in the prevention of tuberculosis. JAMA **271:**698–702.
8. **Collins, F.** 1995. Personal communication.
9. **Curcic, R., S. Dhandayuthapani, and V. Deretic.** 1994. Gene expression in mycobacteria: transcriptional fusions based on *xylE* and analysis of the promoter region of the response regulator *mtrA* from *Mycobacterium tuberculosis*. Mol. Microbiol. **13:**1057–1064.
10. **Deng, G., and M. A. McIntosh.** 1994. An amplifiable DNA region from the *Mycoplasma hyorhinis* genome. J. Bacteriol. **176:**5929–5937.
11. **Dunbar, B. S., H. Kimura, and T. M. Timmons.** 1990. Protein analysis using high-resolution two-dimensional polyacrylamide gel electrophoresis. Methods Enzymol. **182:**441–459.
12. **Dunn, P. L., and R. J. North.** 1995. Virulence ranking of some *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice. Infect. Immun. **63:**3428–3437.
13. **Frothingham, R., H. G. Hills, and K. H. Wilson.** 1994. Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **32:**1639–1643.
14. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:**3961–3965.
15. **Hasløv, K., Å. Andersen, S. Nagai, A. Gottschau, T. Sørensen, and P. Andersen.** 1995. Guinea pig cellular immune responses to proteins secreted by *Mycobacterium tuberculosis*. Infect. Immun. **63:**804–810.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
17. **Henikoff, S., G. W. Haugh, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85:**6602–6606.
18. **Hill, S. A., S. G. Morrison, and J. Swanson.** 1990. The role of oligonucleotide repeats in gonococcal pilin gene variation. Mol. Microbiol. **4:**1341–1352.
19. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:**933–938.
20. **Hwang, D. S., and A. Kornberg.** 1992. Opposed actions of regulatory proteins, DnaA and IciA, in opening the replication origin of *Escherichia coli*. J. Biol. Chem. **267:**23087–23091.
21. **Hwang, D. S., S. B. Thony, and A. Kornberg.** 1992. IciA protein, a specific inhibitor of initiation of *Escherichia coli* chromosomal replication. J. Biol. Chem. **267:**2209–2213.
22. **Imaeda, T.** 1985. Deoxyribonucleic acid relatedness among selected strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*. Int. J. Syst. Bacteriol. **35:**147–150.
23. **Kirillov, M. I., I. L. Shumakov, E. V. Nechaeva, L. N. Siniashina, and G. I. Karataev.** 1993. Nucleotide sequence and properties of an inverted repeating element of a *Bordetella pertussis* chromosome. Genetika **29:**1267–1277.
24. **Li, H., J. C. Ulstrup, T. Ø. Jonassen, K. Melby, S. Nagai, and M. Harboe.** 1993. Evidence for absence of the MPB64 gene in some substrains of *Mycobacterium bovis* BCG. Infect. Immun. **61:**1730–1734.
25. **Lovett, S. T., T. J. Gluckman, P. J. Simon, V. A. Sutera, Jr., and P. T. Drapkin.** 1994. Recombination between repeats in *Escherichia coli* by a *recA*-independent proximity-sensitive mechanism. Mol. Gen. Genet. **245:**294–300.
26. **Novak, E. M., M. P. de Mello, H. B. Gomes, I. Galindo, P. Guevara, J. L. Ramirez, and J. F. da Silveira.** 1993. Repetitive sequences in the ribosomal intergenic spacer of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. **60:**273–280.
27. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:**4007–4021.
28. **Raviglione, M. C., D. E. J. Snider, and A. Kochi.** 1995. Global epidemiology of tuberculosis—morbidity and mortality of a worldwide epidemic. JAMA **273:**220–226.

28a. **Riley, L.** Personal communication.

29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47:**597–626.
31. **Schliesinger, D.** 1995. Personal communication.
32. **Shinnick, T. M.** 1987. The 65-kilodalton antigen of *Mycobacterium tuberculosis*. J. Bacteriol. **169:**1080–1088.
33. **Shinnick, T. M., M. H. Vodkin, and J. C. Williams.** 1988. The *Mycobacterium tuberculosis* 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the *Escherichia coli* GroEL protein. Infect. Immun. **56:**446–451.
34. **Snapper, S. B., L. Lugosi, A. Jekkel, R. E. Melton, T. Kieser, B. R. Bloom, and W. R. Jacobs, Jr.** 1988. Lysogeny and transformation in mycobacteria: stable expression of foreign genes. Proc. Natl. Acad. Sci. USA **85:**6987–6991.
35. **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.
36. **Stover, C. K., G. P. Bansal, M. S. Hanson, J. E. Burlein, S. R. Palaszynski, J. F. Young, S. Koenig, D. B. Young, A. Sadziene, and A. G. Barbour.** 1992. Protective immunity elicited by recombinant BCG expressing OspA lipoprotein: a candidate lyme disease vaccine. J. Exp. Med. **178:**197–209.
37. **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, and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. Nature (London) **351:**456–460.
38. **Strain, S. M., M. R. Tuobiana, E. Ribi, and R. Parker.** 1977. Separation of the mixture of trehalose 6,6′-dimycolates comprising the mycobacterial lipid fraction, "P3". Biochem. Biophys. Res. Commun. **77:**449–456.
39. **Straus, D., and F. M. Ausubel.** 1990. Genomic subtraction for cloning DNA corresponding to deletion mutations. Proc. Natl. Acad. Sci. USA **87:**1889–1893.
40. **Takayama, K., L. Wang, and H. L. David.** 1972. Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **2:**29–35.
41. **Thony, B., D. S. Hwang, L. Fradkin, and A. Kornberg.** 1991. *iciA*, and *Escherichia coli* gene encoding a specific inhibitor of chromosomal initiation of replication *in vitro*. Proc. Natl. Acad. Sci. USA **88:**4066–4070.
42. **Young, D. B., and T. R. Garbe.** 1991. Heat shock proteins and antigens of *Mycobacterium tuberculosis*. Infect. Immun. **59:**3086–3093.
43. **Young, R. A., and T. J. Elliott.** 1989. Stress proteins, infection, and immune surveillance. Cell **59:**5–8.