

Characterization of PE-PGRS Proteins of
Mycobacterium tuberculosis and Molecular Studies
of Strains from Bangladesh

Ph. D. Thesis

Sayera Bannu
Reg. No. 27/98-99

572.6
BAC
P. 1

DEPARTMENT OF MICROBIOLOGY
FACULTY OF BIOLOGICAL SCIENCES
UNIVERSITY OF DHAKA
BANGLADESH

NOVEMBER 2002

SL-NO6

**Characterization of PE-PGRS Proteins of
Mycobacterium tuberculosis and Molecular Studies of
Strains from Bangladesh**

**A dissertation submitted to the University of Dhaka in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in Microbiology**



400598

**Department of Microbiology
Faculty of Biological Sciences
University of Dhaka
Bangladesh**



**(in collaboration with Institut
Pasteur, Paris and ICDDR,B)**

Submitted by

**Sayera Banu
Reg. No. 27/98-99**



INSTITUT PASTEUR



Table of Contents

ABSTRACT	1
ABBREVIATIONS	2
1.0 INTRODUCTION AND REVIEW OF LITERATURE	4
1.1 Epidemiology of TB	4
1.2 Pathogenesis	5
1.3 HIV and TB	7
1.4 Anti-TB agents	8
1.5 Multidrug-resistant TB	8
1.6 Mechanisms of drug resistance	9
1.7 Multidrug-resistant TB in Bangladesh	11
1.8 Directly observed therapy, short course (DOTS) in control of TB and drug resistance	12
1.9 Vaccines for TB	13
1.10 DNA vaccination	16
1.11 <i>Mycobacterium tuberculosis</i> complex	18
1.12 Genomics of <i>M. tuberculosis</i>	19
1.13 Strain identification of <i>M. tuberculosis</i> by DNA fingerprinting	21
1.14 The PE and PPE multigene families	25
1.15 Antigenic variation in other pathogenic bacteria	28
1.16 Antigenic variation in <i>Mycobacterium tuberculosis</i>	32
AIMS OF THE STUDY	35
2.0. MATERIALS AND METHODS	37
2.1 Bacterial strains	37
2.2 Selection of PE-PGRS gene	37
2.3 RNA preparation	37

400598



2.4 RT-PCR	42
2.5 DNA preparation	42
2.6 Polymerase chain reaction conditions	43
2.7.1 Plasmid construction	44
2.7.2 DNA recovery from agarose gel	46
2.7.3 Restriction enzyme digestion, dephosphorylation and ligation	46
2.7.4 Electrocompetent cell preparation	47
2.7.5 Transformation by electroporation	47
2.7.6 Mini-preparation of plasmid DNA for sizing and direct sequencing	48
2.7.7 Sizing of inserts	48
2.8 Direct sequencing	49
2.9 Large scale DNA preparation	49
2.10 DNA vaccination	50
2.11 Expression in HeLa and HEK 293 cells	50
2.12 Expression in reticulocyte extracts	52
2.13 Western blotting	52
2.14 Sub-cellular fractionation	53
2.15 Immunoelectron microscopy	54
2.16 Region of difference (RD)-PCR analysis	55
2.17 Spoligotyping	55
400598	
3.0 RESULTS	59
3.1.1 Expression of PE-PGRS genes by RT-PCR	59
3.1.2 Cloning of PGRS genes into DNA vaccination vector	59
3.1.3 Antibody production using DNA vaccination	61
3.1.4 Expression in HeLa cells and immune detection	62
3.1.5 Expression in reticulocytes extracts and immune detection	65
3.1.6 Detection of PE-PGRS proteins in <i>M. tuberculosis</i> H37Rv, clinical isolates, subcellular fractions	65
3.1.7 Detection of PE-PGRS proteins in other mycobacteria, subcellular fractionation of <i>M. smegmatis</i>	72



3.1.8 Analysis of <i>M. smegmatis</i> proteome in silico	75
3.2.1 Description of strains from Bangladesh	78
3.2.2 Principles of Deletion analysis	78
3.2.3 Results obtained by Deletion analysis	79
3.2.4 Results obtained by spoligotyping	81
 4.0 DISCUSSION	 84
 5.0 CONCLUDING REMARKS	 92
 ACKNOWLEDGEMENTS	 94
 6.0 REFERENCES	 96
 PUBLICATIONS	

Abstract

This study on *Mycobacterium tuberculosis* has two aspects which are different but related. These consist of (a) immunological analyses of the PE-PGRS (Pro-Glu-polymorphic GC-rich sequences) proteins of *M. tuberculosis*, and (b) molecular typing of clinical strains isolated from Bangladeshi patients.

About 8% of the coding region of the genome of *Mycobacterium tuberculosis* H37Rv is comprised of genes for two large unrelated protein families, one of which is the PE family. The largest member of this family is the PE-PGRS class, containing 67 genes. The functions of these proteins were not well known. It has been speculated that the PGRS proteins may be a source of antigenic variation in what is otherwise a genetically and antigenically homogeneous bacterium. Ten of these PE-PGRS genes were selected for investigation in this study. All of them were shown to be expressed *in vitro* by reverse transcription-polymerase chain reaction (RT-PCR). DNA vaccination was used to produce antibodies against five PE-PGRS proteins in mice. The PE-PGRS construct, along with Green Fluorescence Protein DNA, was transfected into HeLa cells and extracts from these cells were probed with the homologous antibody raised in mice and immune complexes were visualized by fluorescence microscopy. Antibodies detected single proteins when the same plasmid constructs used for immunization were expressed in epithelial cells or in reticulocyte extracts, confirming that the PE-PGRS proteins are antigenic. As expected from the conserved repetitive structure, the antibodies cross-reacted with more than one PE-PGRS protein, suggesting that the different proteins share common epitopes. PE-PGRS proteins were detected by Western blotting in five different mycobacterial species (*M. tuberculosis*, *M. bovis* BCG, *M. smegmatis*, *M. marinum* and *M. goodii*) and 11 clinical isolates of *M. tuberculosis*. Immunoblot studies using clinical isolates showed size variation among the proteins, confirming the predicted allelic diversity in the PE-PGRS family in *M. tuberculosis*. Subcellular fractionation studies and immunoelectron microscopy localized many PE-PGRS proteins in the cell wall and cell membrane of *M. tuberculosis* and *M. smegmatis*. The data suggest that some PE-PGRS proteins are variable surface antigens and that they may be an important source of antigenic variation in *M. tuberculosis*.

Tuberculosis has been a major public-health problem in Bangladesh for many decades. A total of 48 sputum samples from adult pulmonary tuberculosis patients were investigated. Deletion analysis, a recently developed PCR-based technique was used to identify the members of *M. tuberculosis* complex with special emphasis on *M. bovis* and *M. tuberculosis*. The results showed that the isolates tested were all *M. tuberculosis* strains, no *M. bovis* was isolated. An interesting finding was that out of 48 strains tested 13 were found to be 'ancestral' type strains while 35 were 'modern' strains. These results indicate that a number of 'ancestral' strains (mostly endemic strains) are also taking part in the transmission of tuberculosis. Spoligotyping, based on the DNA polymorphism in the DR region, was also used to type the strains. A total of 33 (69%) strains could be grouped into 8 different clusters. The largest cluster comprised 15 isolates, whereas the remaining seven clusters consisted of 2 to 5 isolates. Unique (nonclustered) spoligotype patterns were seen in 15 (31%) isolates. The largest cluster comprised 15 (31%) Beijing strains identified both by deletion analysis and spoligotyping. Our results suggest that Beijing *M. tuberculosis* clusters may have a significant role in causing tuberculosis epidemics in Bangladesh. This is a matter of concern since Beijing strains are often associated with drug resistance, display greater virulence and have been detected in many areas of high tuberculosis prevalence in Asia.

Abbreviations

A	adenine or adenosine; one-letter code for alanine
Ag85	antigen 85
AIDS	acquired immunodeficiency syndrome
AMV	Avian Myeloblastosis Virus
BAC	bacterial artificial chromosome
bp	base pair
BCG	bacilli de Calmette et Guerin
BGH	bovine growth hormone
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumin
C	cytosine or cytidine; one-letter code for cystine
cDNA	complementary deoxyribonucleic acid
CIAP	calf intestine alkaline phosphatase
DNA	deoxyribonucleic acid
DNase	deoxyribinuclease
dNTP	deoxynucleoside 5'-triphosphate
DOTS	directly observed therapy, short course
DR	direct repeat
EBNA	Epstein-Barr virus nuclear antigen
EDTA	ethylene-diamine-tetra-acetic acid
EMB	ethambutol
G	guanine or guanosine; one-letter code for glycine
γ -IFN	γ -interferon
HIV	human immunodeficiency virus
IDCH	Institute of Diseases of Chest and Hospital
i.m.	intramuscular
INH	isoniazid
<i>inhA</i>	NADH-dependent enoyl ACP reductase gene
IPTG	isopropyl- β -D-thiogalactopyranoside
IS	insertion sequence
<i>kasA</i>	gene encoding β -ketoacyl carrier protein synthase
<i>katG</i>	catalase-peroxidase gene
kb	kilo base
kDa	kiloDalton
LB	Luria-Bertani (medium)
MPTR	major polymorphic tandem repeat
MDR-TB	multidrug-resistant tuberculosis
MEM	cytoplasmic membrane fraction
MHC I	major histocompatibility complex class I
MIRU	mycobacterial interspersed repetitive unit
NGO	non governmental organization
OD ₆₀₀	optical density at 600 nm

ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PE	proline-glutamine
PEG	polythelene glycol
PGRS	polymorphic GC-rich sequence
<i>pncA</i>	gene encoding pyrazinamidase
PPE	proline-proline-glutamine
PZA	pyrazinamide
R	resistant
RD	region of difference
RE	restriction enzyme
RFLP	restriction fragment length polymorphism
RMP	rifampicin
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
<i>rpoB</i>	gene encoding RNA polymerase subunit β
<i>rpsL</i>	gene coding for S12 ribosomal protein
<i>rrs</i>	gene coding for 16S rRNA
RT-PCR	reverse transcription-polymerase chain reaction
S	sensitive
SDS	sodium dodecyl sulfate
SM	streptomycin
T	thymine or thymidine
<i>Taq</i>	<i>Thermus aquaticus</i> (DNA polymerase)
TB	tuberculosis
TCH	thiophen-2-carboxylic acid hydrazide
TbD1	<i>M. tuberculosis</i> specific deletion 1
tPA	tissue-specific plasminogen activator
U	unit
UV	ultraviolet
VMP	variable major protein
VNTRs	variable number tandem repeats
VSG	variant-specific glycoprotein
WCL	whole cell lysate
WHO	World Health Organization

1. Introduction and review of literature

Tuberculosis, mainly caused by *Mycobacterium tuberculosis*, is by far the most frequently encountered mycobacterial disease in the world. Tuberculosis (TB) remains one of the 10 most important causes of premature mortality worldwide (132). Among the infectious diseases, it is the commonest cause of death in adults and, after acute respiratory and diarrhoeal diseases, in children. In John Bunyan's words of 1680, "TB is the captain of all these men of death". *M. tuberculosis* is the world's most successful pathogen, having infected one in three people on the planet. It is an ancient disease which was present in Egypt from early dynastic times, perhaps as early as 3700 BC (130). There is evidence that suggests that TB evolved during the sixth and seventh millennia BC at which time population increases and cattle domestication occurred in Europe and the eastern Mediterranean (121). It is not known how and when tuberculosis reached the New World. The study of TB did not really begin before 1550 when Girolamo Fracastoro, regarded by some as the first epidemiologist, recognized the contagious nature of the disease (168). In 1882, Robert Koch reported the isolation of *M. tuberculosis* in a study that convinced the medical community and the world of the communicable nature of tuberculosis and brought about concerted efforts to combat the disease. A large number of countries in Asia experienced the height of the epidemic in the late 19th or early 20th century.

1.1 Epidemiology of TB

Each year there are more than 6.5 million new cases of TB and more than 2 million deaths world wide (2). According to WHO predictions, without radical changes in our approach to TB, 200 million people alive today will eventually develop this disease(3). While in developed communities TB occurs mainly in older adults, a high disease incidence is also encountered in young children in developing countries. A large proportion of the population is aged less than 15 years and as many as 40% of tuberculosis notifications may be in children (20). TB in children is generally underdiagnosed and underreported because of difficulties in diagnosis, especially if

children are malnourished. It may be responsible for 10% or more of childhood hospital admissions and 10% or more of hospital deaths (162). The number of infected children aged less than 2 years is relatively small, but disease incidence is very high and frequently takes the form of tuberculosis meningitis or miliary tuberculosis (62). There is no doubt that the best way of preventing childhood disease is to successfully control adult tuberculosis. It is believed that childhood TB will fall automatically once the spread of the disease in adults is brought under control. In developing countries, tuberculosis is the third most common cause of morbidity and mortality combined in women (133). Tuberculosis kills more women than any other infectious disease including malaria and AIDS. Every year one million women die of tuberculosis and it is estimated that 646 million women and girls are already infected (60). The ratio of female to male tuberculosis cases notified is 1:1.5-2.1. More smear-positive male than female tuberculosis patients are diagnosed every year and notified to the WHO. The reason behind this is not very clear but it possible that both biological and social factors are responsible for these differences. In sub-Saharan Africa the seroprevalence of HIV is higher in female tuberculosis patients than in male, whereas in Asia, HIV prevalence among tuberculosis cases is higher in males (60). Tuberculosis causes more maternal deaths than all other causes of maternal mortality put together. The situation in Russia is also alarming where doubling of case rates, from 38 per 100,000 in 1991 to 74 per 100,000 in 1997 has been reported (11). Rates have been recorded that are 100 times higher in prisons than in general society.

1.2 Pathogenesis

Human tuberculosis is characterized by two features. One is that infection with *M. tuberculosis* does not necessarily lead to overt TB. Only a small percentage of individuals infected with *M. tuberculosis* develop disease while most of the others remain disease free. Another unique feature of human TB is its pathogenesis. The pathogenesis of tuberculosis involves the cell-mediated immune responses against *M. tuberculosis*. TB is an infectious disease but it is also an immunological disease which makes its pathogenesis very complicated and human TB difficult to control. *M. tuberculosis* is a

facultative intracellular parasite which requires aerobic conditions for growth, and has a predilection for invading phagocytic monocytes (182). This invasion triggers a complex series of events involving many aspects of the immune system. The characteristic feature of *M. tuberculosis* lies in its cell wall which contains abundant lipids and waxes and these are responsible for strong adjuvanticity and intracellular persistence. Infection generally occurs as small numbers of bacilli reach terminal alveoli of the lung, and then they are ingested by alveolar macrophages within which they are carried to the local lymph nodes. Some replication of the bacilli may occur in the macrophages, but some of the bacilli are destroyed and broken up within phago-lysosomes. Mycobacterial organisms, their proteins or the lipid fractions of the cell wall, directly induce the activation of macrophages to synthesize various cytokines, called monokines. This direct effect of mycobacteria on macrophages does not require activation or involvement of *M. tuberculosis*-specific T cells. Bacteria-specific T cells are activated in the later phase of tuberculosis infection. Interleukin, IL-1 is released by macrophages and IL-2 by T cells. These in turn induce T cells to proliferate into a clone of cells with its particular specificity. TH1-CD4 type cells secrete cytokines, most importantly IL-2 and gamma interferon (γ -IFN), which in turn activate macrophages to ingest and digest mycobacteria and to secrete other cytokines such as IL-8, IL-6 and TNF- α which activate the acute phase inflammatory responses (179). In such infections, the T cell-mediated delayed-type hypersensitivity (DTH) is dominantly induced while the B cell-mediated humoral immune response has a lower contribution. Cytokines produced by macrophages stimulated directly with mycobacterial organisms seem to favour cell-mediated immunity. The immune response to *M. tuberculosis* varies within the individual as a function of age and between individuals both on a genetic basis and in association with acquired conditions including immunosuppression and malnutrition (68). Infants show increased vulnerability to the progression of primary infection to miliary TB and TB meningitis. Late adolescents and young adults are at increased risk of the disease presumably due to reactivation or progression of a primary infection to cavitary disease.

1.3 HIV and TB

The pandemic of acquired immunodeficiency syndrome (AIDS) has caused marked increases in TB notifications in some countries. By virtue of its ability to destroy the immune system, HIV has emerged as the most important risk factor for progression of dormant tuberculosis infection to clinical disease (149). In 1990, 4.6% of all TB deaths were due to HIV infection (61). The Global Programme on AIDS of the WHO estimated that by mid 1994 the prevalence of HIV infection among adults worldwide was between 13 and 14 million (148). Nearly 90% of HIV infections have occurred in developing countries, and the majority has occurred in the age group 15 to 49 years. Because almost two thirds of the world's tuberculosis-infected population is in Asia, the recent increase in HIV infection in Asian communities may result in dreadful increases in HIV-associated TB. The HIV/AIDS epidemic is now a big problem in sub-Saharan African countries, affecting every section of the community (38). It has brought a devastating epidemic which has overwhelmed the health services of all sub-Saharan African countries. Over 60% of children and 70% of adults are co-infected with HIV. Mortality in co-infected tuberculosis patients is as high as 50% within 18 months in some settings. In Central Africa, over a quarter of pregnant women with TB are infected with the virus so that the mother-to-child transmission rates are high.

The natural history of HIV infection is one of progressive decline in immune function, best characterized by loss of CD4⁺ T lymphocytes and TB infection is controlled by cell-mediated immunity, but is generally not eliminated (99). Three mechanisms appear relevant for the development of HIV-associated tuberculosis: reactivation, progression from recent infection, and reinfection. Tuberculosis results in activation of T cells and macrophages, witnessed by increased serum levels of different cytokines (192). Although firm evidence is limited, tuberculosis has been suggested to lead to immune activation, increased virus expression, and more rapid progression to AIDS (41).

1.4 Anti-TB agents

Although chemotherapy has been used for over fifty years to treat tuberculosis, it is only in the last decade that our understanding of the mode and site of action of antituberculosis drugs has improved significantly. Within a decade of development of the first effective agents against tuberculosis, drug resistance had been described. Streptomycin (SM) was the first drug used in the treatment of tuberculosis in the early 1940s. However, it was soon noticed that patients were not responding to the drug resulting in an advanced form of disease and the early response to treatment in others was rapidly followed by deterioration and the emergence of drug-resistant strains. Isonicotinic acid hydrazide or isoniazid (INH), which is more potent against *M. tuberculosis*, was introduced shortly afterwards. Problems of resistance, leading to therapeutic failure, were often encountered when these compounds were employed singly but regimes combining these agents were found to be highly effective both in treatment outcome as well as in preventing the emergence of drug resistance. Thus after a decade of the development of the first effective agents against tuberculosis, the principle of combination drug therapy was established to combat the emergence of resistance (51). The biological basis for the need for combination therapy is believed to be due to the heavy pulmonary bacillary burden that exists prior to therapy and this is sufficiently large to contain spontaneous mutants resistant to a single anti-tuberculosis drug, which will be rapidly selected if treatment commences with a single drug. There are five main drugs which are currently available for the treatment of tuberculosis and these are SM, INH, rifampicin (RMP), pyrazinamide (PZA) and ethambutol (EMB). Standard short course chemotherapy (SCC) comprises two phases: the intensive lasting two months in which INH, RMP, PZA and EMB are taken, followed by the continuation phase in which INH and RMP are administered for four months in order to kill any persisting organisms.

1.5 Multidrug-resistant TB

The emergence of multidrug-resistant *M. tuberculosis* is a worldwide phenomenon that threatens to make TB an incurable disease. Multidrug-resistant TB (MDR-TB) refers to

resistance to two or more first-line antituberculous medications, or to resistance at least to INH and RMP. These two drugs are very important in the treatment of TB, as they are used throughout SCC, and the inability to use them leads to increased morbidity and mortality. Drug-resistant strains are as contagious as the susceptible bacillus. Cure rates up to 95% can be achieved for drug-susceptible tuberculosis but decrease to 56% or less with isoniazid and rifampicin resistance (80). MDR-TB was not a threat to tuberculosis control until the 1980s, and studies done during this period suggested that MDR strains were rare (108), and tended to occur due to inadequate therapy or poor compliance. But soon after that a micro-epidemic of MDR-TB appeared which was first seen in the USA but was gradually detected in different parts of the world. In New York State, the resurgence of TB has been more serious with a doubling of cases since 1978. The case rates exceeded 50 per 100,000 and were approximately five times those for the rest of the United States. This increase was associated primarily with the HIV epidemic, but was also linked to homelessness, immigration, intravenous drug use, poverty and poor access to primary health care. A report from New York revealed that 19% of all isolates were resistant to both RMP and INH and 33% were resistant to at least one drug (76).

1.6 Mechanisms of drug resistance

Streptomycin is a broad spectrum antibiotic that acts on ribosomes and provokes translational defects (128). It binds to the 30S ribosomal subunit leading to inhibition of translational initiation and misreading of messenger RNA. It has been established that mutations in *rpsL*, the gene coding for the S12 ribosomal protein, are associated with resistance (98). These mutations were found to occur in isolates exhibiting high level resistance to SM. Sequence analysis of the *rrs* gene, which codes for the 16S rRNA, from SM-resistant clinical isolates also identified a limited series of single nucleotide substitutions that were, in general, associated with an intermediate level of resistance. Mutations in *rpsL* and *rrs* are the principal SM-resistance conferring mechanisms. Certain species of mycobacteria, such as *Mycobacterium avium*, are naturally resistant to SM, despite harbouring drug-susceptible *rrs* and *rpsL* alleles (98).

Rifampicin is the key component of anti-tuberculous therapeutic regimens and was found to be more potent. Clinical introduction of RMP enabled the duration of chemotherapy for tuberculosis to be reduced to 6 months. RMP is the backbone of the multidrug therapy currently recommended by the World Health Organization for the treatment of leprosy (97). The resistance mechanism of RMP has been well characterized in *Escherichia coli*, making elucidation of the mechanism of RMP resistance in mycobacteria relatively straightforward. RMP blocks transcription by inactivating RNA polymerase (91, 92). It inhibits RNA-polymerase by covalently binding to the β -subunit, encoded by the *rpoB* gene, which is involved in chain initiation and elongation. Many mycobacteria (*M. smegmatis*, *M. avium* and *M. intracellulare*) are naturally resistant to rifampicin as a result of permeability barriers (45).

Unlike SM and RMP, INH is a highly specific antimycobacterial agent, being exquisitely potent against *M. tuberculosis* (and the other members of the *M. tuberculosis* complex: *M. bovis*, *M. microti* and *M. africanum*). It is a bactericidal agent that is used as a first-line drug against TB worldwide. It possesses little to no activity against atypical mycobacteria and *M. leprae* (66). In contrast to RMP, resistance to INH is associated with a variety of mutations affecting one or more genes, such as the one coding for catalase-peroxidase (*katG*), an NADH-dependent enoyl ACP reductase (*inhA*) and a β -ketoacyl carrier protein synthase (*kasA*), both of which are involved in mycolic acid biosynthesis (82, 89, 147, 201).

Pyrazinamide used during the initial intensive phase of short course chemotherapy, is more potent at acidic pH and is believed to be particularly active on intracellular *M. tuberculosis*. PZA resistance has long been associated with the loss of activity of PZase, a cytosolic enzyme of 186 amino acids that deaminates pyrazinamide to pyrazinoic acid (92). A gene, *pncA*, encoding both the pyrazinamidase and nicotinamidase activities of *M. tuberculosis* has been cloned and characterized (160). Mutations in *pncA* confer resistance to pyrazinamide in *M. tuberculosis*. These mutations were found in PZA-resistant isolates of *M. tuberculosis*, and also in *M. bovis* and BCG, thereby explaining the natural resistance to pyrazinamide of these species (160).

Ethambutol is a specific drug that is bacteriostatic and has effects on nucleic acid metabolism, mycolic acid synthesis, phospholipid metabolism, and arabinogalactan synthesis (106, 176). The cell wall is the primary target of EMB (92). Three genes have been found to be associated with resistance to EMB although details of the resistance mechanism are yet to be elucidated. Multidrug resistance appears to result from the stepwise acquisition of independent mutations in the genes encoding drug targets and is not due to a single event (90). A number of operational difficulties are known to be involved with drug resistance, including inadequate prescription and delivery of chemotherapy, poor compliance, or an insufficient number of active drugs in the treatment regimen. In a study done in 1997, it was shown that overall primary resistance (new cases with no previous treatment or no report of previous treatment) rates were 0-16.9% for INH, 0-3% for RMP, 0-4.2% for EMB, and 0.1-23.5% for streptomycin ; acquired resistance rates (resistance in patients described as treatment failures – i.e., those patients who are re-treatment cases or who have had treatment previously) are significantly higher (42).

1.7 Multidrug-resistant TB in Bangladesh

About 0.5% of the population of Bangladesh suffers from smear-positive pulmonary tuberculosis, the estimated average incidence of all forms of the disease being twice that of smear-positive pulmonary cases (1). Each year nearly 80,000 deaths occur due to tuberculosis, and 150,000 additional cases are registered throughout Bangladesh (39). Failure to control TB at the national level is believed to be due to poor case detection, inadequate chemotherapy, and emergence of multidrug resistant strains of *M. tuberculosis*. Inappropriate combination and dosage of active drugs, poor compliance, and failure to continue treatment for the optimum period of time are the causes of development of acquired or secondary drug resistance. Until now, very little was known about the extent of MDR-TB in Bangladesh. A study done on the sputum samples of 1281 cases in urban Dhaka showed that 30% of the *M. tuberculosis* strains were resistant to at least one of the antitubercular drugs and the prevalence of MDR-TB was 5% (5). It

is not possible to take appropriate measures in the TB control program without having clear knowledge about the pattern of resistance in Bangladesh.

1.8 Directly observed therapy, short course (DOTS) in control of TB and drug resistance

Recognizing that tuberculosis is one of the most neglected global health problems and that the TB epidemic is out of control in many parts of the world, the WHO declared TB to be a global health emergency in April 1993. The current strategy is to have all countries provide standardized, effective short-course chemotherapy through a program called DOTS to all TB patients, with emphasis on cure of sputum-positive cases. The key components of DOTS strategy are standardized detection of smear-positive patients by means of sputum smear microscopy, short-course chemotherapy regimens directly observed by trained health workers during the initial two months of intensive phase of treatment, recording and reporting of treatment outcomes, together with mechanisms to ensure a regular drug supply (65). Short-course chemotherapy is currently the most effective treatment for most patients with TB and supervision of treatment is the most important health strategy of the decade by ensuring that most patients complete their regimen (17).

Many developing and developed countries have adopted the DOTS strategy for the treatment of tuberculosis and DOTS has been very effective in reducing the incidence of the disease by two-thirds and in preventing the emergence of resistance in an epidemic of HIV-associated TB in the United States (34). The potential effects of DOTS on the treatment of TB in developing countries is high and sometimes even greater than the results achieved in developed countries when drugs became available in the 1940s (65). Although DOTS has been found to be a very effective treatment strategy, there are reports where self-supervision has achieved equivalent treatment outcomes at lower cost and better outcomes for retreatment patients (202). In November 1993, the DOTS strategy was implemented in four *thanas* (sub-districts) of Bangladesh and the treatment outcomes in the first cohort of sputum smear-positive patients were successful (111).

Based on the above success, the programme was expanded to more thanas. Implementation of DOTS in 40 *thanas* of Bangladesh resulted in a relatively higher cure rate (111), although the results are still far from satisfactory. The DOTS strategy was first adopted by the Government of Bangladesh at thana health complexes but later non governmental organizations (NGOs) began implementing the same project and thus helping the government to run the TB control programme. Bangladesh Rural Advancement Committee (BRAC) is the first NGO to implement DOTS strategy with success (134). The DOTS programme has been successful in various parts of the world but it has been shown, from projected data on global deaths due to TB, that even with a highly effective DOTS programme, some 2.5 million deaths will occur in the years to come (132). However, the death rate will come down gradually to about 1.5 million per year in the next 25 years. This projection clearly indicates that DOTS alone cannot effectively reduce the burden of the disease.

1.9 Vaccines for TB

Although tuberculosis remains one of the leading causes of death worldwide, the only vaccine currently available is BCG, the bacille de Calmette et Guérin, which has been found to impart highly variable levels of efficacy, particularly against pulmonary disease (70). In 1907, a live attenuated *M. bovis* BCG vaccine was developed and it has been used extensively to control human tuberculosis for the past 50 years (85). It was derived from a virulent bovine strain by subculturing the strain 231 times in medium containing ox bile (40). In 1921, after BCG had been shown to be avirulent with no apparent reversion frequency, Calmette, Guérin and Weill-Halle gave the first vaccine with the attenuated bacillus to an infant in a family in which individuals were already infected with tuberculosis, and the child did not contract this disease. The BCG vaccine has numerous advantages; it can be given safely soon after birth, it generates long-term protection, it consistently protects against tubercular meningitis, has few side effects and most importantly it is cheap to produce (136). Studies of BCG have found widely differing protection in different populations (43). Thus, while BCG vaccination has been

highly effective in some trials, in others, such as the widely cited trial in south India, the calculated efficacy has been close to zero (Fig. 1)(71).

Methodological biases, factors interfering with immune responses, such as contamination with environmental mycobacteria, genetic diversity of vaccinated populations, follow-up of tuberculosis cases in young children, and the quality of BCG vaccine strains, may interfere with the efficacy of BCG vaccination (25). Differential immunogenicity of various BCG strains also contributes to the efficiency of BCG vaccination (112). Glaxo, Pasteur and Russian strains of BCG have been found to be superior to Prague and Japanese strains in terms of immunogenicity in mice (multiplication and persistence in target organs) (112). The variable efficacy of the BCG vaccine in different parts of the world, coupled with other serious factors such as the HIV epidemic and the increasing incidence of outbreaks of drug-resistant tuberculosis, has underlined the need to identify protective antigens of the tubercle bacillus so that an effective vaccine can be developed to control this devastating epidemic and also to understand the reasons for the failure of BCG to confer uniform protection.

Vaccines currently used for the prevention of infectious diseases are based on whole pathogens, either killed (eg. polio, rabies, cholera vaccine) or live attenuated (eg. oral polio, measles, mumps, tuberculosis). Sub-unit vaccines also exist, composed of purified proteins (eg. tetanus, diphtheria, pertussis, hepatitis B, influenza vaccine) or polysaccharide antigens (eg. meningococcus, pneumococcus, influenza B vaccine). DNA vaccines (HIV) or polynucleotide vaccines are relatively new (100). The major disadvantage of the live attenuated vaccine is the inability to use it safely in an immunocompromised individual such as in cancer patients or HIV-infected people. Killed whole cell pathogens do not replicate but their efficacy is lower than the live attenuated ones. Immunization with DNA is thus an effective means of eliciting both antibody as well as cell-mediated immune responses. The major advantages of DNA vaccine are that they can be administered in immunocompromised people and can induce a cellular immune response that is dominated by antigen-specific T lymphocytes that

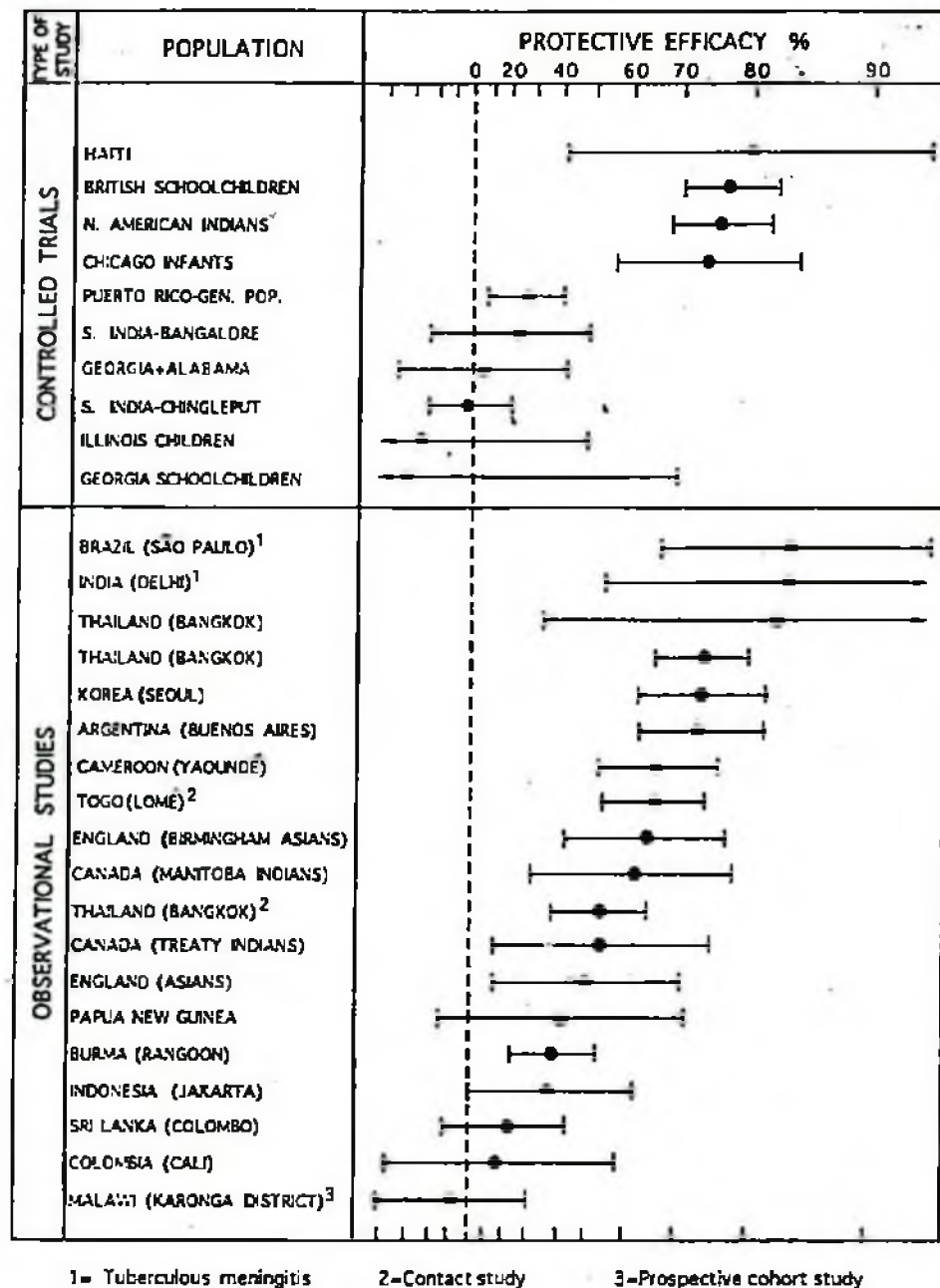


Fig. 1. Efficacy of BCG as determined in controlled trials and observational studies in various populations. Lines reflect 95% confidence intervals (71).

produce both γ -interferon (γ -IFN) and are also cytotoxic towards infected cells (Th1 type of cellular response). Both these functions have been shown to be required for effective antimycobacterial immunity in mice (25). DNA vaccination has been shown to have an immunotherapeutic role in boosting the efficiency of the immune system in *M. tuberculosis*-infected patients when administered with antibacterial chemotherapy (117).

1.10 DNA vaccination

The general principle of DNA vaccination is simple (Fig. 2)(100). A gene encoding a microbial antigen is inserted into a DNA vaccination vector that is an expression plasmid. The plasmid DNA is then transformed into a bacterium and amplified. The purified DNA encoding the immunogen is injected into an immunocompetent host. The plasmid directly transfects a living cell, the gene is transcribed to RNA in the nucleus and translated to protein by the ribosomes. Finally, the host becomes immunized against a heterologous protein produced by its own cells. Instead of administering a protein, the genetic information for the protein is injected and the host becomes a factory for the production of the gene products. Intramuscular vaccination leads to the generation of strong cellular immune responses with a Th1 cytokine profile. It is precisely this type of immunity, in which infected macrophages are activated in their bactericidal activity by Th1 cytokines such as γ -IFN or lysed by specific cytotoxic T-cells, that is needed in a good TB vaccine. So far, three types of antigens have been tested as DNA vaccines: Antigen 85 (118) (184), 38 kD phosphate binding PstS antigen (74) and 65 kD heat shock protein (23).

In recent years, considerable efforts have been made towards the development of new vaccines against tuberculosis (136) including plasmid DNA vector-based vaccines, recombinant and mutant BCG vaccines, and sub-unit vaccines. DNA vaccination thus plays a crucial role in the development of a better tuberculosis vaccine, as this technique offers a simple method to define the actual protective antigens of *M. tuberculosis*. Administration of plasmid DNA expression vectors has been shown to result in protein expression *in vivo*, generation of humoral and cell-mediated immune responses, and protective efficacy in animal models of infectious disease (63). It has been shown to be

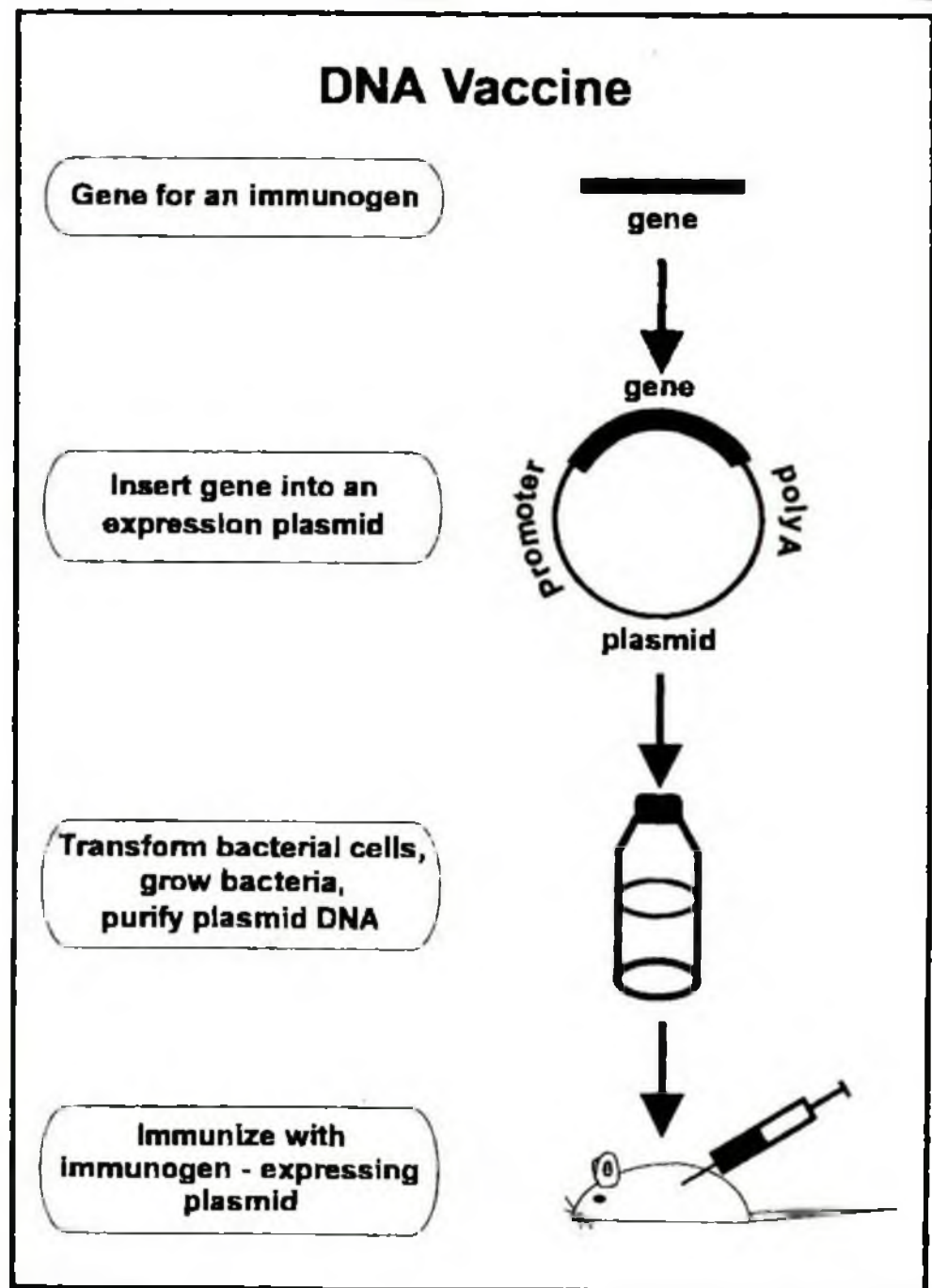


Fig. 2. Principle of DNA vaccination

effective for inducing immune responses and/or protection in other animal models, including HIV in mice, influenza in chickens and ferrets, bovine herpes virus in mice and cattle, rabies virus, malaria in mice, leishmaniasis in mice, and herpes simplex virus in mice (63). DNA vaccines induce such responses without the need for live organisms, replicating vectors or adjuvants and, therefore, may offer significant practical and safety advantages over certain other approaches. Silva and Lowrie did the first study on DNA vaccination in tuberculosis (164). They immunized mice with a tumor-derived macrophage cell line transfected with the *hsp60* gene from *M. leprae* and these mice were shown to be protected from an intravenous inoculum (5×10^6) of *M. tuberculosis* H37Rv or BCG. This protection was manifested as a relatively slow decline over a number of weeks in the bacterial load in the liver and spleen of vaccinated mice. A BCG control was not included in this study but other studies have shown that in BCG-infected mice, the kinetics of protection are very rapid which suggest a different mechanism involved in the cell line immunized mice. In another study, Huygen and her colleagues made a plasmid DNA construct encoding the secreted form of the mature culture filtrate protein of *M. tuberculosis*, antigen 85 (Ag85), using the signal sequence from tissue plasminogen activator, and a mature form of Ag85 which did not contain a signal sequence (101). Immunization with this plasmid induced substantial humoral and cell-mediated immune responses and conferred significant protection against challenge with live *M. tuberculosis* and *M. bovis* BCG.

1.11 *Mycobacterium tuberculosis* complex

The *M. tuberculosis* complex comprises five members, namely *M. tuberculosis*, *Mycobacterium africanum*, *Mycobacterium canetti*, *Mycobacterium microti*, *Mycobacterium bovis* including *M. bovis* BCG. The term ‘tubercle bacillus’ is also used to denote the members of the *M. tuberculosis* complex. They are 99.9% similar at the nucleotide level and have identical 16S rRNA sequences but the disease spectrum they cause varies widely. They mostly differ in their host range and molecular basis of pathogenicity: *M. tuberculosis*, *M. africanum*, *M. canettii* cause disease exclusively in humans, *M. bovis* infects a wide range of mammals, including humans (157) while *M.*

microti does not cause disease in humans, but only in voles. Mycobacterial infections due to nontuberculous mycobacteria (NTM), such as the *Mycobacterium avium* complex (MAC), *M. fortuitum*, and *M. chelonae*, are also on the increase (137).

It is not very easy to identify the members of the *M. tuberculosis* complex by examination of the cultures, confirmatory tests are needed. There are two types of *M. tuberculosis*, classical and Asian variants. On subculture, colonies on egg media are 1-2 mm in diameter, rough, heaped and resemble breadcrumbs or cauliflower. Growth is restricted to 34-38°C. Both variants are nitratase positive, aerobic and are susceptible to pyrazinamide and cycloserine. The classical variant is resistant to thiophen-2-carboxylic acid hydrazide (TCH) but the Asian variant is susceptible. Both variants are usually niacin positive. Colonies of *M. bovis* on egg media are about 1 mm in diameter, flatter and smoother than those of the human variant. Growth on pyruvate media is usually much better than that on glycerol media. Strains are nitratase negative, microaerophilic, susceptible to TCH and cycloserine but resistant to pyrazinamide. They are niacin negative. *Mycobacterium africanum* was first found in west Africa and two variants occur (49). The African I variant, principally of west African origin, is nitratase negative and, on subculture, colonies on egg media resemble those of *M. bovis*. The African II variant, of east African origin, is nitratase positive and colonies resemble those of *M. tuberculosis*. Even though the BCG vaccine strain was derived from a strain of *M. bovis*, the cultural and morphological features of this vaccine strain are similar to those of the classical human variant of *M. tuberculosis* (198). There is no preference for pyruvate and, some strains grow on glycerol-based media. *M. microti* also known as 'vole or murine tubercle bacillus' is not very well characterized but has been used as a live vaccine.

1.12 Genomics of *M. tuberculosis*

Completion of the *Mycobacterium tuberculosis* genome sequence has provided tools for the rapid development of mycobacterial genomics which would greatly enhance the understanding of these organisms (Fig. 3)(46). The complete genome sequence of the

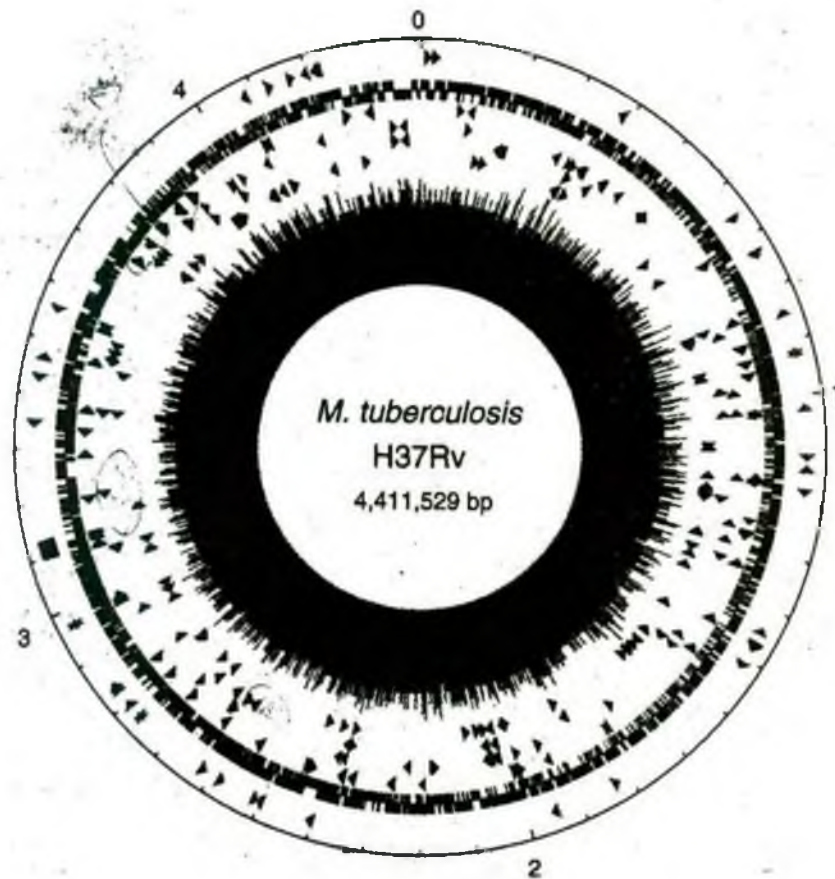


Fig. 3. Circular map of the chromosome of *M. tuberculosis* H37Rv. The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes; the second ring inwards shows the coding sequence by strand; the third ring depicts repetitive DNA; the fourth ring shows the positions of the PPE family members; the fifth ring shows the PE family members; and the sixth ring shows the positions of the PGRS sequences (46).

widely used reference strain H37Rv comprises 4,411,532 bp and contains around 4000 genes. The guanine (G) and cytosine (C) content of the DNA is 65.6% although some areas with an exceptionally high G+C content (>80%) were found to be a novel gene family (28). Over 51% of the genes have arisen as a result of gene duplication or domain shuffling events, and 3.4% of the genome is composed of insertion sequences (IS) and phages. There are 56 copies of well known IS elements including a new IS family, IS/535. On homology searching using bioinformatics, 50 genes encoding stable RNA species and 3924 genes encoding proteins were identified accounting for >91% of the

potential coding capacity (44). Various data base comparisons attributed precise functions to ~40% of the predicted proteins, 44% showed some similarity to known proteins while the functions of the remaining 16% were completely unknown. It was assumed that the latter proteins might be responsible for the specific mycobacterial functions (178).

The genome sequence encodes multiple copies of many of the enzymes involved in either the synthesis of mycolic acids and other complex lipids or polyketides (46). At least 8% of the genome is dedicated to lipid metabolism and, many of these encode enzymes that appear to be involved in fatty acid degradation. Database searches revealed two classes of potential new virulence factors which bear similarity to proteins involved in pathogenicity elsewhere, and systems required for growth inside phagocytic cells. A series of phospholipases C, proteases, lipases and esterases have been identified from the genome sequence which could act as virulence factors. These might attack cellular or vacuolar membranes. Four copies of *mce* genes encoding proteins thought to have role in invasion of host cells were present in the genome (8). The tubercle bacillus possesses several other genes that could enhance its growth in the macrophage. These include two hemoglobin-like proteins, which may protect against oxidative stress or be involved in oxygen capture. Comparisons of the genomes of the various tubercle bacilli could provide us sources of informations about the mycobacterial virulence determinants specific for humans and other host-range determining factors. Comparative genomics has uncovered a large variety of polymorphic regions in the otherwise highly conserved genomes of the tubercle bacilli. Functional analysis of all these polymorphic regions will certainly help to discover the role of these genomic differences in pathogenesis and biology.

1.13 Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting

Outbreaks of infectious disease often result from exposure to a common source of the etiologic agent (75). Generally, the etiologic agent causing an outbreak of infection is derived from a single cell whose progeny are genetically identical or closely related to the

source organism. In epidemiological terms, the organisms involved in the outbreak are clonally related. DNA fingerprinting techniques now exist which identify specific strains of *M. tuberculosis* (135). These techniques along with conventional approaches have become powerful tools in tuberculosis epidemiology (22, 78, 86, 87, 142, 154, 167, 185, 187). A number of genetic elements have been found to be associated with rearrangements in *M. tuberculosis* complex strains and are being used as the genetic marker for strain differentiation.

DNA fingerprinting with the widely reported insertion sequence IS6110 has been considered as an excellent tool for investigating the spread of tuberculosis because of the high degree of discrimination obtained by this element (37, 59, 73, 155, 166, 175, 186, 190, 197, 199). The technique involves extracting the DNA from a *M. tuberculosis* culture, digesting this DNA with a restriction endonuclease, then separating the resulting fragments according to size using agarose gel electrophoresis (84, 105) followed by “Southern” blotting these fragments to a membrane and probing this to identify which fragments contain the IS6110 element. For each strain analysed this process yields a unique “fingerprint” pattern with 2 to >20 bands. The RFLP fingerprint has been shown to be stable by examining *M. tuberculosis* isolated from chronic secretors and in outbreak settings (110). Repeat isolates from individuals who persistently excrete *M. tuberculosis* have been shown to have identical DNA fingerprints. Thus there is increasing consensus that RFLP patterns are a reliable indicator of strain identity (196). Most *M. tuberculosis* strains can be differentiated by IS6110 RFLP fingerprinting, except those harbouring 1 or 2 copies of IS6110, and the reliability of RFLP for strain identification increases with the number of copies of IS6110. The use of RFLP technique is limited to culture-positive cases as it can only be performed on microgram quantities of DNA.

A recently developed PCR based method allows simultaneous detection and strain differentiation of *M. tuberculosis* present in clinical specimens without the need to culture this slowly organism. The method is referred to as spacer oligotyping or “spoligotyping” because it is based on strain-dependent hybridization patterns of in vitro-amplified DNA with multiple spacer oligonucleotides (16, 52, 143). It is a simple method

which allows simultaneous detection and typing of *M. tuberculosis* in clinical specimens and reduces the time between suspicion of the disease and typing from 1 or several months to 1 or 2 days (103). Spoligotyping is particularly useful for analysis of isolates with low IS6110 copy numbers. It is more discriminative than RFLP analysis for strains with one or two copies of IS6110 (81). The method relies on the in vitro amplification of DNA across the unique, highly polymorphic direct repeat (DR) locus present in the *M. tuberculosis* complex chromosome. This region contains multiple short 36-bp DRs, and nonrepetitive spacers, which are 35 to 41 bp in length, and interspersed between the DRs. Spoligotyping is a rapid method that allows large numbers of isolates to be handled in a short time. The direct repeats are extremely well conserved among *M. tuberculosis* complex strains and because they are present as a multiple target for DNA amplification, spoligotyping is a specific method for the detection of *M. tuberculosis* complex (103). Spoligotyping has been used successfully to detect and type *M. tuberculosis* in clinical materials such as sputum, bronchial lavages, and tissues, though another recently described PCR-based typing method called ligation-mediated PCR (LMPCR) has shown to be more discriminatory than spoligotyping (24).

DNA typing with other repetitive elements such as the GC-rich repetitive sequence (PGRS) is also used in the characterization of *M. tuberculosis* strains (64, 154, 187). PGRS have been used as a genetic markers of polymorphism for epidemiological studies on *M. tuberculosis* and *M. bovis* (36, 151, 180). Fingerprinting using PGRS has shown to be a useful tool for characterization of the *M. tuberculosis* strains with a single copy of IS6110. Several studies indicate that outbreak strains can be well recognized by the PGRS DNA probe and that PGRS DNA restriction fragment patterns have enough stability to be used as effective tools in the epidemiology of tuberculosis (187). RFLP analysis with the *M. tuberculosis* polymorphic GC-rich repeat sequence contained in the recombinant plasmid pTBN12 was found to have similar discriminatory efficiency as IS6110-based RFLP analysis (13).

Recently, a typing method based on variable number tandem repeats (VNTRs) of genetic elements named mycobacterial interspersed repetitive units (MIRUs), in 12 human

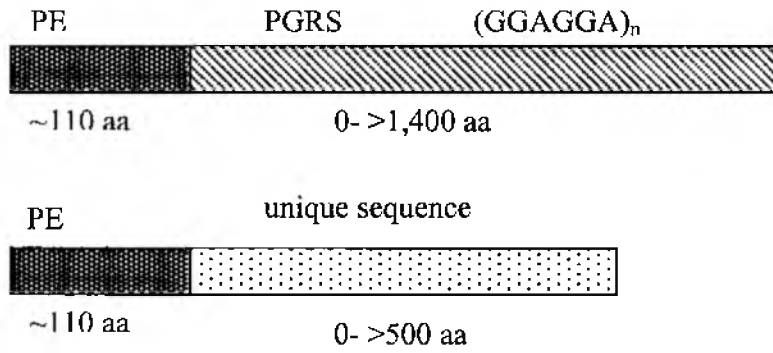
minisatellite-like regions of the *M. tuberculosis* genome, has been developed (122, 172). MIRUs are composed of 40-100 bp repetitive DNA sequences, dispersed in 41 intergenic regions of the *M. tuberculosis* complex genome (120, 173, 174). Twelve of these locations display polymorphisms in MIRU copy numbers among non-related *M. tuberculosis* isolates. From 2 to 8 MIRU-VNTR alleles were identified in the 12 regions in these strains, which corresponds to a potential of over 16 million different combinations, yielding a resolution power close to that of IS6110-RFLP. Large-scale genotyping of *M. tuberculosis* using IS6110 is labor-intensive, requires culturing of the slow growing mycobacteria and the results are sometimes difficult to compare among laboratories. Typing using MIRUs is a PCR-based method where strains can be typed by a numerical code corresponding to the numbers of MIRUs in the 12 different loci. MIRUs have been used to differentiate between *M. bovis* BCG and virulent strains of *M. tuberculosis* complex (119). These genetic markers provide nonambiguous data which are portable between different laboratories and which are therefore highly suitable for global epidemiological studies (77, 122).

In the highly conserved *M. tuberculosis* complex, where single nucleotide polymorphism is rare, insertion and deletion events are the principal source of genome diversity. Completion of the genome sequence of *M. tuberculosis* H37Rv along with comparative genomics approaches has greatly enhanced the understanding of the mechanisms of insertion and deletion resulting distribution of variable regions around the genomes of tubercle bacilli (29, 32, 46). Based on this knowledge a deletion analysis system has been developed to differentiate the members of *M. tuberculosis* (29). There are 20 variable regions, of which 14 regions of difference (RD1-14) were found to be absent from bacillus Calmette-Guerin Pasteur relative to *M. tuberculosis* H37Rv (18, 83). Six regions, H37Rv related deletions (RvD)1-5, and *M. tuberculosis* specific deletion 1 (TbD1), are absent from the *M. tuberculosis* H37Rv genome relative to other members of the *M. tuberculosis* complex. Based on the presence or absence of TbD1 region, *M. tuberculosis* strains can be divided into “ancestral” and “modern” type. The strains of major epidemics like Beijing, Haarlem, and African *M. tuberculosis* are of modern types (29).

1.14 The PE and PPE multigene families

Deciphering the entire genome sequence of *Mycobacterium tuberculosis* H37Rv has opened a new era in mycobacteriology, and revealed many unsuspected aspects of the biology of this important pathogen (Fig. 3)(46). Among the major surprises, was the finding that 8% of the potential coding capacity was accounted for by two large unrelated gene families encoding the PE and PPE proteins (Fig. 4) (44), and preliminary analyses indicate a similar level of abundance in the genomes of *M. bovis*, *M. bovis* BCG Pasteur, and the CDC1551 strain of *M. tuberculosis* (30). The names PE and PPE are derived from the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) found in most cases near the N terminus of these glycine-alanine-rich proteins in conserved domains of ~110 and 180 amino acid residues, respectively (Fig. 4)(48). The PE family contains ~100 members while the PPE protein family comprises 68 members. The PE family can be subdivided into three classes on the basis of their domain structure (Fig. 5)(30)). The first class contains 29 species which consists of the conserved ~100 amino acid long N-terminal segment only. The largest class of the PE family, with 67 members in *M. tuberculosis* H37Rv, is referred to as the PE-PGRS subfamily since these proteins consist of the PE domain followed by a C-terminal extension with multiple tandem repetitions of Gly-Gly-Ala or Gly-Gly-Asn encoded by the PGRS motif (polymorphic Guanine+Cytosine rich repetitive sequences) (Fig. 4). PE-PGRS proteins are predicted to contain as many as 1900 amino acids of which up to 50% can be glycine. There is some sequence similarity to EBNA, the Epstein-Barr virus nuclear antigens, as these also contain Ala-Gly repeats that have been shown to act as proteasome inhibitors and block antigen presentation by the major histocompatibility complex (MHC) class I pathway (114, 115). MHC class I knockout mice are very susceptible to *M. tuberculosis*, underlining the importance of a cytotoxic T-cell response in protection against disease. CD8 T cells that respond to MHC class I antigen presentation are known to be important mediators of resistance to mycobacterial disease (72). The third class of PE proteins contains the conserved N-terminal segment followed by domains that are unique in sequence, comprising between 100 and 400 amino acid residues.

Representatives of the PE family



Representatives of the PPE family

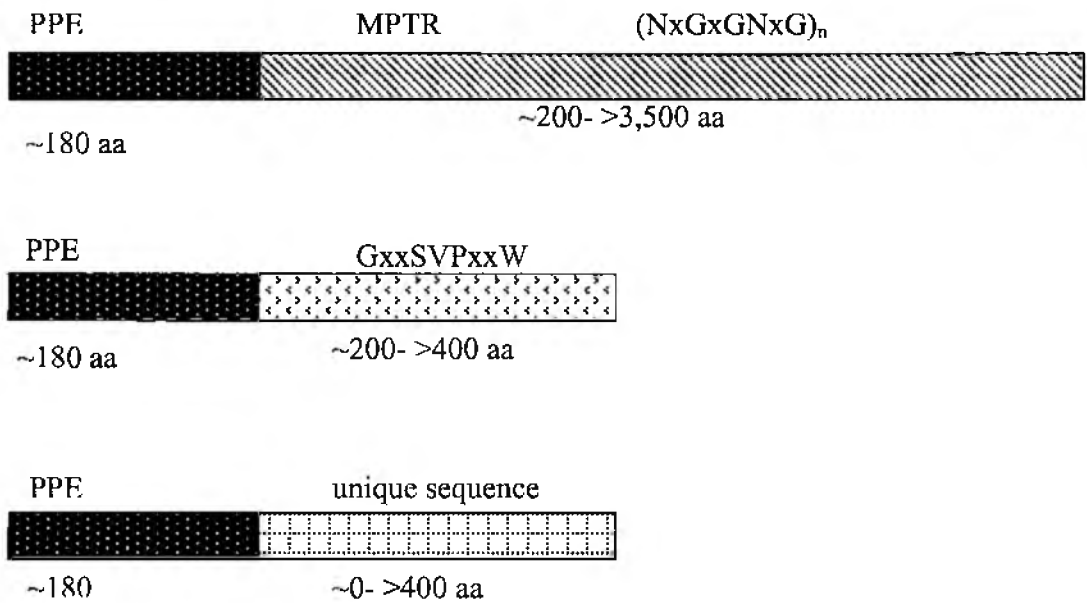


Fig. 4. Classification of the PE and PPE protein families

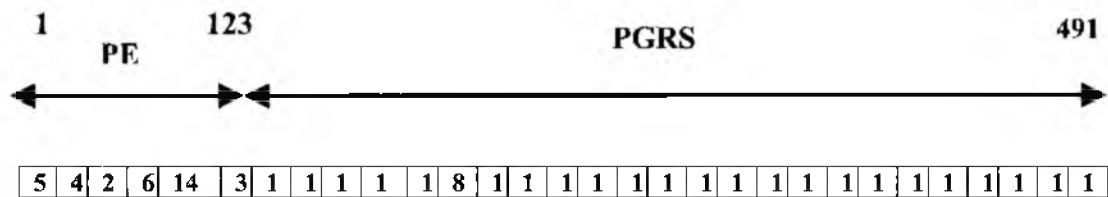


Fig. 5. Domain organization of PE-PGRS protein Rv1441. The figure is not to scale, and the most significant amino acid motifs, detected using the MEME and MAST algorithms (10, 9, 178), have the following consensus sequence: 1, GNGGAGG; 2, YQALSAQAAAFH; 3, INAPTQALLGRPLI; DEVSAAIAALF; 5, MSFVVAAPEMLAAAAADL; 6, QFVQALTAAGAYAAAE; 8, GGWLYGNG; 14, SPLQTMQQ.

The PPE family which comprises 68 members also has a well conserved N-terminal domain of about 180 amino acids that is followed by a C-terminal segment that varies markedly in sequence and length (46, 48). These proteins fall into three groups, one of which constitutes the MPTR (major polymorphic tandem repeat) class characterized by the presence of multiple, tandem copies of the motif Asn-X-Gly-X-Gly-Asn-X-Gly. There is little evidence available concerning the possible function of the PPE proteins, but roles in cell structure or storage cannot be excluded. The largest PPE-MPTR protein that has been described contains 3,300 amino acid residues. MPTR has showed restriction fragment length polymorphism on Southern blotting experiments when different isolates were compared (88). The second group comprises 10 proteins characterized by the 180-residue N-terminal domain and the third subfamily consists of a group of proteins with a C-terminal extension of 200 to 400 amino acids that shows a reasonable level of sequence conservation (30). The highly repetitive nature of the PGRS and MPTR genes suggests that they should be prone to insertion and deletion events, resulting from replication errors (127). Twenty one insertion/deletion events have been found to be involved in PE and PPE gene families when the *M. tuberculosis* strains H37Rv and CDC 1551 were compared (32).

Despite their numerical abundance, no PE proteins have been identified yet by biochemical means including extensive analysis of the *M. tuberculosis* proteome by means of 2-dimensional electrophoresis and mass spectrometry (102, 153). However, as

pointed out by Betts et al. (19), this may be due to the relative paucity of trypsin cleavage sites in these acidic proteins. Other explanations such as localization in cell membranes are possible since proteins from this cellular compartment are notoriously refractory to analysis by 2-dimensional gel electrophoresis. The remarkably low level of nucleotide sequence variation in structural genes, and hence the high degree of conservation contrasts with the situation seen in the PE-PGRS and PPE-MPTR genes that display extensive restriction fragment length polymorphisms reflecting physical differences such as repeat length or copy number. This could have resulted from inter- or intragenic recombinational events or, more probably, from strand slippage during replication (140). It is probably significant in this respect that the genes encoding PE and PPE proteins often occur in clusters on the chromosome of H37Rv. These mechanisms have been found to generate antigenic variability in other bacterial pathogens (150).

1.15 Antigenic variation in other pathogenic bacteria

Antigenic variation among the surface proteins of many intra- and extracellular pathogens is commonly observed, presumably reflecting the selective pressure of the host's immune system. This has been well documented in both eukaryotes and prokaryotes. There are a number of mechanisms used to generate antigenic variation in pathogenic bacteria (Table 1). Homologous recombination is widely used to generate molecular variation in bacteria and depends on the RecA function. In one of the best studied cases, the pili of *Neisseria gonorrhoeae*, antigenic variation often involves recombinational events in which previously silent coding sequences are inserted into expression sites (35, 116, 124). The multiple silent pilin loci are transcriptionally inactive. A silent locus carries one or more partial pilin gene copies located in a tandem array and connected by repetitive sequences. While these partial gene copies show the same arrangement of variable minicassettes found in variant expression genes, they lack the conserved amino-terminal coding sequences.

Silent loci carry several repetitive sequence motifs with counterparts in the expression locus. Sequence homologies and similarities in the organization of silent and expressed

Table 1. Summary of mechanisms used to generate genetic variation in pathogenic bacteria

Mechanism of variation	Organism	Component	Function or element involved
General homologous recombination	<i>N. gonorrhoeae</i>	<i>pilE</i>	RecA
	<i>B. hermsii</i>	<i>vmp</i>	?
Variation via repetitive domains	<i>A. marginale</i>	<i>msp1</i>	?
	<i>M. hyorhinis</i>	<i>vlp</i>	?
	Streptococci	<i>emm</i>	?
Site-specific recombination (inversion)	<i>S. typhimurium</i>	<i>h2</i>	Hin, Fis, HU
	<i>E. coli</i>	<i>fimA</i>	FimB, FimE
	<i>M. bovis</i>	<i>tfp</i>	Piv
RecA-independent recombination via short repeats	<i>N. gonorrhoeae</i>	<i>opa</i>	CTCTT
	<i>N. meningitidis</i>	<i>opa</i>	CTCTT
	<i>H. influenzae</i>	<i>lic</i> locus	CAAT
	<i>H. influenzae</i>	<i>hifA</i> , <i>hifB</i>	TA
	<i>N. meningitidis</i>	<i>opc</i>	poly C
	<i>N. gonorrhoeae</i>	<i>pilC</i>	poly G
Phase variation by DNA modification	<i>E. coli</i>	<i>papa</i>	Lrp, Dam

genes are also evident between gonococci and meningococci. Variation via repetitive domains, site-specific recombination, RecA-independent recombination via short repeats and phase variation by DNA modification are other mechanisms involved in generating antigenic variation in pathogenic bacteria (150).

Perhaps the best-studied example of antigenic variation and its role in infection is that of pathogenic African trypanosomes, the causative organisms of sleeping sickness (57). The antigenic variation of African trypanosomes is mediated by major variant-specific glycoproteins (VSGs) that coat the organism and are the major antigens recognized by

the immune system. Individual trypanosomes generally express a single VSG. *Trypanosoma brucei* produces rising and falling waves of parasitemia by generating subpopulations that have antigenically different forms of VSGs. The waves of parasitemia are a principal feature of a chronic infection that persists until the individual is either treated or dies of the disease. In the course of infection, expanding numbers of parasites give rise to subpopulations that have switched to different VSGs and therefore are able to escape the immediate impact of the antibody response to the parent population (138). These subpopulations are subsequently recognized by the immune system, but continued generation of novel VSGs prevents clearance of the infection. This capacity for switching among different VSGs is thought to lead to eventual exhaustion of host defenses in the terminal phases of the disease (55).

Individual trypanosome parasites are capable of expressing more than 100 different VSGs during the course of an infection. The genes encoding these proteins are members of a diverse multicopy family and are thought to be transcribed exclusively from one of a limited number of telomeric expression sites in the genome. VSG sequences typically comprise 400 to 500 amino acids and show dramatic levels of polymorphism in primary structure. The primary sequences show greater conservation at the carboxy terminus, where the VSGs are anchored at the parasite membrane. VSGs in solution are present as dimers, possibly attached through disulfide bonds. The VSG coat on the trypanosome surface is a closely packed monolayer, so that only a small part of the VSG molecule is accessible to antibodies on the cell surface. Malaria infections caused by *Plasmodium falciparum* are persistent, recurrent and complex (57). Studies have showed that the incidence of childhood malaria in regions of endemic infection is highly variable and the incidence of severe disease is not evenly distributed but, rather, shows periodicity and space-time clustering (50).

The cause of this variability is not very well determined but an explanation for this is that they arise from changes in the cytoadherence and antigenic properties of the parasitized erythrocytes (57). The adherence properties of parasites are implicated in a number of the severe manifestations of malaria, including cerebral malaria, while changes in

antigenicity allow the parasites to persist and produce repeated infections in individuals. The antigenic variation that enables the parasites to avoid elimination by the host immune response leads to the generation of highly diverse parasite populations in regions of high malaria transmission. Variable forms of proteins at the surface of infected erythrocytes mediate both the adherence and antigenic properties of parasite-infected erythrocytes (14, 96). These proteins can change at rates reported to be as high as 2% per generation. Molecules responsible for the adherence and antigenic variation of *P. falciparum*-infected erythrocytes are termed PfEMP1 and are encoded by members of the *var* family of genes (15). A limited number of proteins of 200 to 350 kDa encoded by *var* genes are expressed by individual parasites from a diverse repertoire of 50 to 150 copies accounting for about 2 to 6% of the genomic DNA. Switches in *var* gene expression are thus capable of generating a large variety of antigenic forms. The proteins from *var* genes are transported to the surface of infected erythrocytes and anchored in membrane structures. The predicted amino acid sequences of these proteins show a large, variable extracellular segment, a transmembrane sequence, and a terminal segment that serves as a submembrane anchor.

Major outer membrane proteins, termed variable major proteins (VMPs) play a principal role in the antigenic variation of *Borrelia* spp. VMP is the primary variant antigen recognized in *Borrelia* infection (12). Switches among the expressed members of this family of genes allow small subpopulations to continually arise during infection and to stay ahead of the immune response. A single *B. hermsii* bacterium has been shown to produce up to 40 different serotypes, and switching rates among different VMP genes have been estimated to range from $\sim 10^{-4}$ to 10^{-3} (171). Proteins similar to VMPs are found on the surface of *Borrelia burgdorferi*, the causative agent of Lyme disease (200).

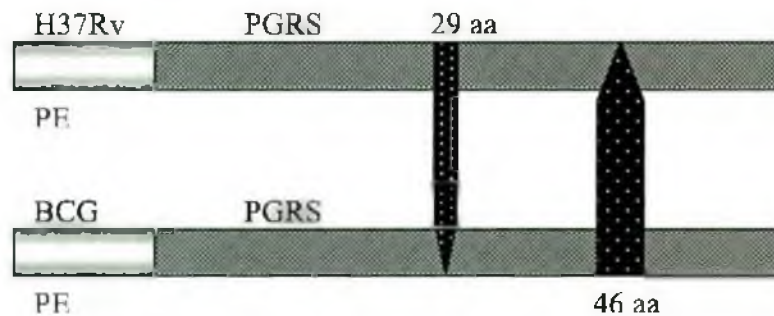
Haemophilus influenzae has evolved a highly adaptive means by antigenic variation to survive in an environment (the human nasopharynx) that is constantly bathed in antibodies (79, 123). A number of genetic mechanisms are involved in antigenic variation of *H. influenzae* to assure rapid variation in its surface antigens. Spontaneous point mutations in the genes encoding protein antigens which alter the physical topography of

the molecule resulting in the antibodies directed against the original antigen being unable to recognize the altered form, is the most common phenomenon in antigenic diversity of proteins. Nucleotide sequence analysis of the genes encoding several *H. influenzae* surface proteins reveals many single base changes distributed throughout the variable regions. Co-colonization by more than one *H. influenzae* strain facilitates horizontal transfer of genes by providing a supply of potentially diverse DNA that can be taken up by the colonizing strains and incorporated into the bacterial genome. In addition, *H. influenzae* use phase variation to turn surface antigens on and off through regulatory mechanisms that involve nucleotide repeats. *H. influenzae* uses two or three genetic mechanisms to alter a single antigen, such as pili. Group A streptococcal M protein shows extraordinarily both size and antigenic variation among strains (95). The M protein is a major surface protein and virulence factor for the group A streptococcus. The size variation observed among M proteins is extensive and occurs both among strains of the same M type and of different M types. Sequence analysis of spontaneous M protein size variants shows that homologous recombination events between intragenic tandem repeats can lead to the generation of antigenic variation.

1.16 Antigenic variation in *Mycobacterium tuberculosis*

The PGRS genes are highly Guanine+Cytosine rich (about 80%) and a major source of polymorphism in the *M. tuberculosis* complex, which otherwise displays remarkable genetic homogeneity and very few single nucleotide polymorphisms (169). A number of observations related to the PE and PPE families inspires us to speculate that they might be of immunological importance and could even represent the principal source of antigenic variation (Fig. 6). It is known that a PE-PGRS member, Rv1759c, a fibronectin-binding protein of 55 kDa produced during infection and disease, elicits a variable antibody response (4). When sera from 14 different tuberculosis patients were screened for reactivity to a recombinant form of 22T, strongly positive responses were observed in eight cases, suggesting either that individuals mount different immune responses or that this PE-PGRS protein may not be produced by all strains of *M. tuberculosis* (4, 69).

PE-PGRS sequence variation in ORF Rv0746



H37Rv	... GSGAPGGAGGAAGLWGTGGAGGAGGSSAGGGGAGGAGGAGGWLLGDGGAGGIGGAST	
BCG	...GSGAPGGAGGAAGLWGT-----GCAGGIGGAST	
H37Rv	VLGGTGGGGGVGGLWGAGGAGGAGGTGLVGGDGGAGGAGGTGGLLAGLIGAGGGHGGTGG	
BCG	VLGGTGGGGGVGGLWGAGGAGGAGGTGLVGGDGGAGGAGGTGGLLAGLIGAGGGHGGTGG	
H37Rv	LSTNGDGGVGGAGGNAGMLAGPGGAGGAGGDGENLDTGGDGGAGGSAGLLFGSGGAGGAG	
BCG	LSTNGDGGVGGAGGNAGMLAGPGGAGGAGGDGENLDTGGDGGAGGSAGLLFGSGGAGGAG	
H37Rv	GFGFLGGDGGAGGNAGLLSSGGAGGFGGFGTAGGVGGAGGNAGWLGF-----	
	GFGFLGGDGGAGGNAGLLSSGGAGGFGGFGTAGGVGGAGGNAGWLGF GGAGGGIGGIGGN	
H37Rv	-----GGAGGVGG SAGLIGTGNGGNGGTGA	
BCG	ANGGAGGNGGTGGQLWGSGGAGGEGGAALS VGDTGGAGGVGGSAGLIGTGNGGNGGTGA	
H37Rv	NAGSPGTGGAGGLLLGQNGLNGLP	
BCG	NAGSPGTGGAGGLLLGQNGLNGLP	

Fig. 6. Sequence variation between *M. tuberculosis* H37Rv and *M. bovis* BCG-Pasteur in the PE-PGRS encoded by open reading frame (ORF) Rv0746. In H37Rv, 29 codons are deleted which are present in BCG and there is insertion of 46 codons that is absent in BCG.

Indirect support in favour of the latter is provided by restriction fragment length polymorphism for various PE-PGRS genes in clinical isolates (48), whereas direct evidence of allelic variation of the PE-PGRS protein encoded by Rv0746 has also been reported between *M. tuberculosis* H37Rv and *M. bovis* BCG Pasteur (Fig. 6)(46). The

gene for the PE-PGRS protein Rv0746 BCG differs from that in H37Rv by the deletion of 29 codons and the insertion of 46 codons.

Genes homologous to PGRS and MPTR-coding sequences have been detected in *Mycobacterium gordonae*, *Mycobacterium kansasii* and *Mycobacterium marinum* (140, 141, 154), and in *Mycobacterium asiaticum*, *Mycobacterium gastri* and *Mycobacterium szulgai*, respectively (88). In *Mycobacterium leprae*, a close relative of the tubercle bacillus, only 9 intact PE and PPE genes were found although 30 pseudogenes were present (47). No intact members of the PE-PGRS subfamily were found. No PE-PGRS was found in any other bacteria. The abundance of these genes in mycobacteria implies that they must have an important biological role restricted to mycobacteria.

Aims of the study

A) The completion of the genome sequence of *Mycoabacterium tuberculosis* has led to the identification of two large gene families, PE and PPE. Of them PE-PGRS which is the largest class of the PE family is of particular importance. However our understanding about their biological functions and host response to these proteins is very limited. The repetitive C-terminal part of these proteins is reminiscent of protein families known to generate antigenic diversity in other organisms. It is tempting to speculate that, if these proteins are indeed antigenic, their polymorphic nature could result in differences in immune response that would account for the apparent variability of disease amongst patients with *M. tuberculosis*.

The overall aims of this study were to better understand the biological and immunological functions of the PE-PGRS proteins of *M. tuberculosis*. The following were included:

- Monitor expression of some of the PE-PGRS genes at the transcriptional level in *in vitro*-cultured *M. tuberculosis* cells.
- Produce antibodies against selected PE-PGRS proteins in mice by DNA vaccination.
- Express PE-PGRS proteins in eukaryotic cells in order to test the specificity of the antibodies produced by DNA vaccination.
- Detect PE-PGRS proteins in *M. tuberculosis*, other mycobacteria and clinical isolates of *M. tuberculosis*.
- Determine subcellular localization of PE-PGRS proteins of *M. tuberculosis*.

B) Tuberculosis is one of the leading causes of morbidity and mortality in Bangladesh. Bangladesh ranked fourth of the 22 highest incidence countries in the world, with a yearly incidence of 246 per hundred thousand and annual death rate of 55 per hundred thousand. Emergence of multidrug-resistant strains and further spread of these resistant as well as sensitive organisms to others might be one of the precipitating factors. Unfortunately, there is virtually no information about the characters and types of mycobacterial strains causing this disease in Bangladesh.

The objectives of this part of the study was to characterize the clinical isolates from TB patients in Bangladesh using molecular methods. Specific aims were:

- Identification of members of the *M. tuberculosis* complex.
- Investigation of *M. bovis* tuberculosis in humans.
- Molecular typing of *M. tuberculosis* strains of Bangladesh.

2. Materials and methods

2.1 Bacterial strains

Bacterial strains used in this study are listed in Table 2. *M. tuberculosis* H37Rv, *M. bovis* BCG, *Mycobacterium smegmatis*, *M. marinum* and *M. goodii* strains were obtained from laboratory stocks. Clinical isolates of *M. tuberculosis* (strains 1, 9, 23, 26, 33, 34, 35) have been described previously (90). Strain 1 is from Indonesia, strain 9 is from Russia, strain 23 is from Vietnam, strain 26 is from Africa, strains 33 and 35 are from Netherlands and strain 34 is of unknown origin. Strains 357 and 1,627 are of Spanish origin and strains 616 and 1,349 were from San Francisco. The vectors and derived recombinants were maintained in *Escherichia coli* DH5 α .

Bacterial strains for deletion analysis and spoligotyping were collected from the Institute of Diseases of Chest and Hospital (IDCH) in Bangladesh (Table 3). They were the isolates from the patients of pulmonary tuberculosis who came to the hospital for treatment during different time periods. The total number of samples was 48.

2.2 Selection of PE-PGRS gene

Ten PE-PGRS genes were randomly selected for investigation (Table 4). Smaller genes were selected to make PCR amplification easier. The size of the selected PE-PGRS genes varied from 1 to 2 kb. The genes having *Bgl*II and *Bam*HI sites inside the gene sequence were excluded.

2.3 RNA preparation

RNA for RT-PCR reactions was prepared from a one week old 10 ml culture of *M. tuberculosis* H37Rv. The bacteria were resuspended in 1 ml of TRIzol reagent (Gibco) after harvesting by low speed centrifugation, and broken using 500 μ l of 600 μ m

Table 2. Mycobacterial strains used in PE-PGRS study

Species	Strain	Origin	Source
<i>M. tuberculosis</i>	H37Rv		Institut Pasteur, Paris
	I	Indonesia	Institut Pasteur, Paris
	9	Russia	Institut Pasteur, Paris
	23	Vietnam	Institut Pasteur, Paris
	26	Africa	Institut Pasteur, Paris
	33	Netherlands	Institut Pasteur, Paris
	34	?	Institut Pasteur, Paris
	35	Netherlands	Institut Pasteur, Paris
	357	Spain	Institut Pasteur, Paris
	616	San Francisco	Institut Pasteur, Paris
	1349	San Francisco	Institut Pasteur, Paris
	1627	Spain	Institut Pasteur, Paris
<i>M. bovis</i> BCG			Institut Pasteur, Paris
<i>M. smegmaris</i>	MC ² 155		Institut Pasteur, Paris
<i>M. marinum</i>	930732		Institut Pasteur, Paris
<i>M. goodii</i>	CIPT0310002		Institut Pasteur, Paris

Table 3. Bangladeshi patient characteristics

Strain	Age (yr)	Sex	Treatment history	BCG scar	Sensitivity				Origin
					INH	RMP	EMB	SM	
TB1	28	M	previously treated	present		not done			Dhaka
TB2	30	M	previously treated	present	Sensitive (S)	S	S	S	Dhaka
TB6	18	M	new case	absent	S	S	S	S	Dhaka
TB8	25	M	new case	present		not done			Dhaka
TB9	60	M	new case	absent	S	S	S	S	Dhaka
TB10	50	M	previously treated	present		not done			Patuakhali
TB11	30	M	new case	present		not done			Dhaka
TB12	56	M	new case	present	S	S	S	S	Natore
TB14	52	M	previously treated	present	Resistant (R)	R	R	R	Rangpur
TB16	20	F	new case	present		not done			Gazipur
TB19	50	F	new case	present	S	S	S	S	Dhaka
TB20	16	M	new case	absent		not done			Narayanganj
TB21	22	M	new case	absent	S	S	S	S	Narayanganj
TB23	35	M	new case	present	S	S	S	S	Kishoreganj
TB24	20	F	new case	absent		not done			Dhaka
TB27	26	M	new case	present	S	S	S	S	Dhaka
TB29	22	F	new case	present	S	S	S	S	Dhaka
TB30	20	F	new case	absent	S	S	S	S	Dhaka
TB32	35	F	new case	present		not done			Dhaka
MA1	30	M	previously treated	not known	S	S	S	S	Dhaka
MA2	32	F	previously treated	absent	S	S	S	S	Dhaka
MA5	42	M	previously treated	absent	S	S	S	S	Dhaka
MA6	32	M	previously treated	absent	R	R	S	S	Shariatpur
MA7	50	M	previously treated	absent	R	R	S	S	Noakhali
MA9	33	M	previously treated	absent	R	R	S	S	Munshiganj
MA11	27	M	previously treated	present	S	S	S	S	Comilla
MA12	38	F	previously treated	absent	R	S	S	R	Dhaka
MA13	33	F	previously treated	absent	S	S	S	S	Dhaka
MA16	35	M	previously treated	absent	R	R	R	R	Dhaka
MA17	46	M	previously treated	not known	R	S	R	S	Narayanganj
MA19	25	M	previously treated	present	R	S	S	S	Narayanganj
MA20	40	M	previously treated	absent	R	R	R	R	Kishoreganj
MA25	55	M	previously treated	present	S	R	S	R	Dhaka
MA27	50	M	previously treated	present	S	R	S	S	Shariatpur
MA28	40	M	previously treated	absent	R	R	S	R	Narsingdi
MA29	30	M	previously treated	present	R	S	S	S	Kishoreganj
MA31	31	M	previously treated	absent	R	R	R	R	Narayanganj
MA32	28	M	previously treated	absent	S	S	S	S	Dhaka
MA35	23	M	previously treated	present	S	S	S	R	Dhaka
MA36	50	M	previously treated	not known	S	S	S	S	Dhaka
MA37	45	M	previously treated	present	S	S	S	S	Habiganj
MA38	39	M	previously treated	absent		not done			Comilla
MA39	22	F	previously treated	present	R	R	S	R	Gazipur
MA41	45	M	previously treated	absent	R	S	S	S	Dhaka
MA42	22	M	previously treated	present	R	S	R	R	Shariatpur
MA45	17	F	previously treated	not known	S	S	S	S	Dhaka
MA46	40	M	previously treated	not known	S	S	S	S	Dhaka
MA48	20	F	previously treated	not known	R	S	S	S	Dhaka

Table 4. Selected PE-PGRS genes and primers used in RT-PCR

Gene ^a	Size (kb)	Mol. Wt (Da)	Primers used for RT-PCR ^b	
			Name of primer	Sequence
Rv0980c	1.3	42 059	RT-Rv0980cL	5'-GGCGACCACCCAGGTGTTG-3'
			RT-Rv0980cR	5'-GGTCCCGTCGATCCCGTTC-3'
Rv1067c	2.0	54 683	RT-Rv1067cL	5'-TGTTGGTGTCTCCGTCGCAG-3'
			RT-Rv1067cR	5'-TGGTCGACATTGGTCGCCTC-3'
Rv1068c	1.3	39 294	RT-Rv1068cL	5'-GTGCCAGACATGCTCAGCTC-3'
			RT-Rv1068cR	5'-GCACCTGCTGTTTCGACGTTG-3'
Rv1243c	1.6	47 270	RT-Rv1243L	5'-GGTGCCGACGAGGTATCAG-3'
			RT-Rv1243R	5'-TGCCGTACAACAAGCCACCG-3'
Rv1325c	1.8	49 565	RT-Rv1325L	5'-CGGATTTGGCGAACATCGGG-3'
			RT-Rv1325R	5'-GGGTGGGCGTGTTCACAATG-3'
Rv1396c	1.7	47 890	RT-Rv1396L	5'-CGGATCTGGCAAGCATCGG-3'
			RT-Rv1396R	5'-AGCTGCTGCTCGACGTTGG-3'
Rv1441c	1.4	40797	RT-Rv1441cL	5'-GGAATGTTGTCAGCGGCGG-3'
			RT-Rv1441cR	5'-GCGAGCACCTGTTGTTTCGAG-3'
Rv1818c	1.4	40743	RT-Rv1818L	5'-GCGATGGCGGCATTGTTCTC-3'
			RT-Rv1818R	5'-GGCTCCGTTACCGATCAGTG-3'
Rv1983	1.6	53692	RT-Rv1983L	5'-CCGGCATGAGTGGTGGAG-3'
			RT-Rv1983R	5'-GCGAAGATGTAGTACAGCCC-3'
Rv2162c	1.5	44 464	RT-Rv2162L	5'-TCTCGGCGTTGTTGGGGTC-3'
			RT-Rv2162R	5'-CGTTGGGATTGGCGGTGTC-3'

Genes^a are designated and primers^b were designed according to the H37Rv genome sequence.

diameter beads and a Mickle apparatus. The homogenized samples were incubated for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes. After centrifugation at 12,000 g for 10 min the supernatant was chloroform extracted. Per 1 ml of TRIzol reagent 0.2 ml of chloroform was added. The tubes were shaken vigorously by hand for 15 seconds and incubated at room temperature for 2 to 3 min. The samples were centrifuged at $12,000 \times g$ for 15 min at 4°C. Following centrifugation, the mixture separated into a lower red phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The volume of the aqueous phase is about 60% of the volume of TRIzol reagent used for homogenization. The aqueous phase was transferred to a fresh tube and RNA was precipitated from the aqueous phase by mixing with isopropyl alcohol. Five hundred microlitres of isopropyl alcohol was used per 1 ml of TRIzol reagent used for the initial homogenization. The samples were incubated at room temperature for 10 min and centrifuged at $12,000 \times g$ for 10 min at 4°C. The precipitated RNA, invisible before centrifugation, formed a gel-like pellet on the side and bottom of the tube. The supernatant was removed and the RNA pellet was washed once with 75% ethanol, adding at least 1 ml of 75% ethanol per 1 ml of TRIzol reagent used for the initial homogenization. The sample was mixed by vortexing and centrifuged at $7,500 \times g$ for 5 min at 4°C. At the end of the procedure, the pellet was dried briefly (air-dry or vacuum-dry for 5-10 min) and resuspended in 100 µl of DNase-digestion buffer and the nucleic acid was allowed to dissolve for 10 min, before treatment with DNase I (Boehringer Mannheim). Ten units of RNase-free DNase I was added and incubated at 37°C for 1 hr to enable digestion of DNA. In order to remove the DNase, 100 µl of phenol:chloroform:isoamyl alcohol was added and extracted. The tubes were then centrifuged. The aqueous phase was extracted with one volume of chloroform to remove traces of phenol and the RNA collected by ethanol precipitation. The RNA pellet was dissolved in DEPC-treated deionized water and the sample was placed on ice. The concentration of RNA was determined by measuring the absorbance at 260 nm of a small aliquot. The remaining RNA was stored at -20°C and was subsequently used.

2.4 RT-PCR

Primers for RT-PCR were designed using computer assisted programme (OSP). The Promega Access RT-PCR System was used to perform RT-PCR. The system is designed for the reverse transcription (RT) and polymerase chain reaction (PCR) amplification of a specific target RNA from either total RNA or mRNA. This is a one tube, two-enzyme system which is quick and sensitive. The system uses Avian Myeloblastosis Virus Reverse Transcriptase (AMV RT) and the reaction mixture contained 10 µl of AMV/*Tfl* 5 × reaction buffer, 1 µl dNTP mix (10 mM each dNTP), 1 µM concentration of each primer, 2 ml of 25 µM MgSO₄, 0.1 µl of AMV reverse transcriptase, 0.1 µl of *Tfl* DNA polymerase, 30 to 50 ng of RNA sample and nuclease-free water up to 50 µl. The reagents were taken in a 0.2 ml microcentrifuge tube on ice. The components were mixed by pipetting. AMV reverse transcriptase and *Tfl* DNA polymerase were added to the reaction at the end. The tube was gently vortexed for 10 seconds to mix the components. Thermal cycling was performed on a PTC-100 thermal cycler (MJ Research, Watertown, Mass., USA) with one cycle of 45 min at 48°C and one cycle of 2 min at 94°C for the first strand cDNA synthesis (reverse transcription) followed by 40 cycles of 30 sec at 94°C, 1 min at 60°C, 2 min at 68°C and 1 cycle for 7 min at 68°C. Duplex RT-PCR was performed. Reverse and forward primers from the 16S rRNA gene were also added to the same reaction mixture. As a negative control, blank reagent that contained H₂O instead of nucleic acid was used. The PCR products were analyzed by agarose gel electrophoresis, stained with ethidium bromide and visualised under a UV transilluminator.

2.5 DNA preparation

Mycobacterium tuberculosis H37Rv was grown in Middlebrook 7H9 medium (Difco) supplemented with albumin-dextrose (ADC, Difco) for 10-12 days. After overnight incubation with glycine (0.2 M final concentration) cells were harvested by low speed centrifugation and resuspended in lysozyme solution (lysozyme 500 µg ml⁻¹, sucrose 25%, Tris 50 mM [pH 8.0], EDTA 50 mM). Following incubation for 4 hours at 37 °C an equal volume of Proteinase K solution (Tris 100 mM pH 8.0, Proteinase K 400 µg ml⁻¹,

SDS 1%) was added and reincubated for a further 4 hours at 55 °C. RNase (10 µg per ml) was added and incubated at 37°C for 30 min before phenol-chloroform extraction. For phenol-chloroform extraction, an equal volume of phenol:chloroform (1:1) was added to the DNA solution and vortexed. It was then centrifuged at 15,000 rpm for 5 min, the supernatant was transferred to new tubes. An equal volume of chloroform:isoamyl alcohol was added to it and vortexed again, the mixture was then centrifuged at 15,000 rpm for 3 min and the supernatant was taken and precipitated with 0.6 volume of isopropanol. The DNA pellet was washed with 70% ethanol, centrifuged, dried in a vacuum dessicator and dissolved in distilled water to obtain chromosomal DNA. DNA from the bacterial artificial chromosome (BAC) clones was prepared as described previously (27). Each BAC clone was inoculated into a 50 ml Falcon polypropylene tube containing 40 ml of 2 YT medium with 12.5 µg of chloramphenicol per ml and grown overnight at 37°C with shaking. Cells were harvested by centrifugation and stored at -20°C. The frozen pellet was resuspended in 4 ml of solution A (50 mM glucose, 10 mM EDTA, 25 mM Tris [pH 8.0]), and 4 ml of freshly prepared solution B (0.2 M NaOH, 0.2% sodium dodecyl sulfate) was then added. The solution was gently mixed and kept at room temperature for 5 min before addition of 4 ml of ice-cold solution C (3 M sodium acetate, pH 4.7). Tubes were kept on ice for 15 min and centrifuged at 10,000 rpm for 15 min. After isopropanol precipitation, the DNA pellet was dissolved in 600 µl of RNase solution (15 mM Tris HCl [pH 8.0], 10 µg of RNase A per ml). After 30 min at 37°C, the DNA solution was extracted with chloroform-isoamyl alcohol (24:1) and precipitated from the aqueous phase with isopropanol. The DNA pellet was then rinsed with 70% ethanol, air dried, and dissolved in 30 µl of distilled water. Cosmid DNA from the pYUB18 shuttle library of *M. tuberculosis* H37Rv was prepared as described (67).

2.6 Polymerase chain reaction (PCR) conditions

Chromosomal DNA was used to amplify the selected PE-PGRS genes by PCR using the suitable oligonucleotide primers in a MJ Research PTC-100 thermal cycler. The name and sequence of the primers are listed in Table 4. A reaction mixture of 50 µl containing

5 μ l of 10 $\times$ PCR buffer (600 mM Tris-HCl, pH 8.8/20 mM $MgCl_2$ /170 mM $(NH_4)_2SO_4$ /100 mM β -mercaptoethanol), 5 μ l of 20 mM nucleotide mix, 50 nM of each primer, 5-50 ng of template DNA, 10% DMSO, 0.25 μ l of *Taq* polymerase (Gibco-BRL), and sterile distilled water to 50 μ l was first denatured at 95°C for 2 min and then subjected to amplification for 35 cycles with a cycling profile of denaturation at 95°C for 30 s, 2 min at annealing temperature (varied from 58° to 70°C) and extension at 72°C for 1 min with a final extension of 10 min at 72°C.

2.7.1 Plasmid construction

PE-PGRS genes from *M. tuberculosis* H37Rv strain were amplified by PCR from H37Rv chromosomal DNA, cosmids (139) or BAC clones (27) using customized primers with *Bgl*II or *Bam*HI sites, and cloned in VIIns.tPA vector (163). V1J was derived from V1 (129, 183) and pUC19 (Fig. 7). V1 contains the CMVintA enhancer, promoter, and intron A for transcription initiation and BGH (bovine growth hormone) terminator for transcription termination in a partially deleted pBR322 backbone. To construct a vector that is well defined, more compact, and would improve plasmid yields, the CMVintA/BGH fragment from V1 was transferred to a pUC19 backbone. The entire *lac* operon was removed from pUC19 because it was unnecessary, and a functional *lac* operon was found to be detrimental to plasmid stability in *E. coli* (129). The tissue-specific plasminogen activator (tPA) leader peptide was ligated into the optimized vaccination vector (163). Suitable primers were designed from the flanking regions of the selected PE-PGRS genes. Amplified DNA was digested with *Bgl*II or *Bam*HI, isolated on agarose gels (1%) containing TBE buffer, detected with ethidium bromide and extracted using a GeneClean II kit (Bio 101). Fragments were cloned into the *Bgl*II or *Bam*HI site of the VIIns.tPA vector, as previously described (129) and transformed into *E. coli*. The methods of DNA recovery from agarose gel, restriction enzyme digestion, cloning and transformation are described below.

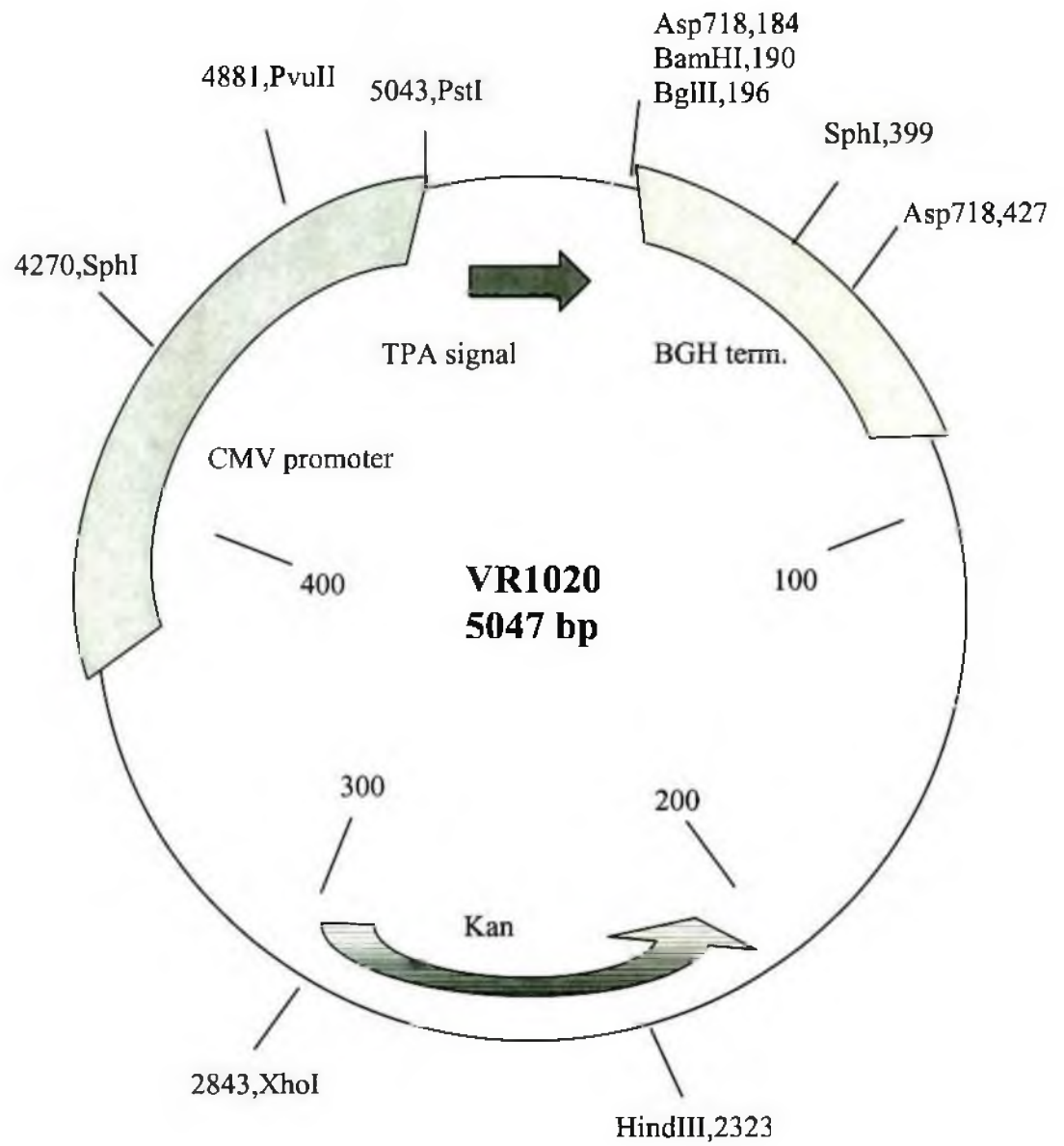


Fig. 7. Physical map of DNA vaccination vector

2.7.2 DNA recovery from agarose gel (using Gene Clean II kit)

The gel slice containing the desired fragment was cut from the agarose gel by a sharp scalpel. The gel fragment was cut into small pieces and put into a microcentrifuge tube, and, weighed properly. An appropriate volume (for 100 µg of gel 300 µl of NaI) of NaI was added to dissolve the gel by up and down movements or keeping it at 45°-55°C for 5 min. Five microlitre (for 5 µg of DNA) of glass milk was added and mixed by gentle vortexing. The mixture was kept on ice for 10 min. The tube was then centrifuged at 15000 rpm for 10 min. Supernatant was discarded and the pellet of glass milk was washed with NEW wash solution by pipetting. The washing was repeated 3 times and sample was kept on ice until the washing was complete. The tube was centrifuged at 15000 rpm for 10 min and the supernatant was removed completely. The pellet was resuspended in 10 µl of distilled water.

2.7.3 Restriction enzyme (RE) digestion, dephosphorylation and ligation

A total volume of 50 µl containing 5-10 µl of DNA, 5 µl of 10 × buffer, 5 µl of RE (3-5 U of RE for 1 µg of DNA) and 30 µl of distilled water was taken in a tube. The mixture was then incubated at 37°C for 1 hour. Both the vector and insert were RE digested. The vector was then dephosphorylated using calf intestine alkaline phosphatase (CIAP). To 50 µl of digested mixture, 10 µl of 10 × CIAP buffer, 1 µl of CIAP (0.01 U of CIAP/pmol ends) were added. Sterile distilled water was added to a final volume of 100 µl. The reaction mixture was then incubated at 37°C for 1 hour. To stop the reaction, 2 µl of 0.5 M EDTA was added and the dephosphorylated DNA was then phenol/chloroform extracted. After ethanol precipitation DNA was resuspended with 10 µl of distilled water. The gene cleaned fragments were ligated to *Bgl*II or *Bam*HI-digested, dephosphorylated vector (V1Jns.tPA) in a 1:10 molar ratio, using 10 U of T4 DNA ligase (New England Biolabs, Beverly, Mass.) at 16°C for 20 h.

2.7.4 Electrocompetent cell preparation

A single colony was picked from a freshly streaked LB plate and inoculated into 5 ml of LB broth and grown overnight. In a 500 ml conical flask 200 ml LB broth was inoculated with 1/100 volume the overnight culture. Cells were grown at 37°C with vigorous shaking to an OD₆₀₀ of 0.5 to 1. The flask was put on ice for 15 to 30 min and centrifuged at 7,000 g for 15 min at 4°C. The supernatant was discarded and the pellet was resuspended in 200 ml of cold sterile distilled water and centrifuged at 7,000 g for 15 min at 4°C. The pellet was resuspended in 100 ml of cold water and centrifuged as above. The pellet was then resuspended in 20 ml of 10% glycerol and was centrifuged. The pellet was resuspended in a final volume of 1 ml 10% glycerol. The cell concentration should be about 3×10^{10} cells/ml. The cell suspension was aliquoted on ice and stored at -70°C and was used for electro-transformation.

2.7.5 Transformation by electroporation

One microliter of the ligation mix was used for electroporation of 30 µl of electrocompetent *E. coli* DH5α in an Easyject Plus electroporator (Eurogentec, Seraing, Belgium), with settings of 2.5 kV, 25 µF, and 99 Ω, in 2-mm-wide electroporation cuvettes. After electroporation, cells were resuspended in 600 µl of LB medium, allowed to recover for 45 min at 37°C with gentle shaking, and then plated on LB agar containing chloramphenicol (12.5 µg/ml), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; 50 µg/ml), and isopropyl-β-D-thiogalactopyranoside (IPTG; 25 µg/ml). The plates were incubated overnight, and recombinants (white colonies) were picked manually. As a control, the V1Jns.tPA vector not encoding a gene insert was used. PCR products were also cloned in pUC19 and pET vectors and transformed into *E. coli*.

2.7.6 Mini-preparation of plasmid DNA for sizing and directing sequencing

Each clone was inoculated into a glass tube containing 2 ml of LB broth with 12.5 µg/ml of chloramphenicol and grown overnight (~12-14 h) at 37°C with gentle shaking. Cells were harvested by centrifugation at 300 rpm for 8 min in 2 ml microcentrifuge (Eppendorf) tube. The cells were resuspended in 100 µl of solution S1 (25 mM Tris, 10 mM EDTA, [pH: 7.6]), vortexed and placed on ice. Two hundred microliter of freshly prepared solution S2 (0.2 M NaOH, 1% SDS) was added. The solution was gently mixed and kept on ice for 5 min before adding 150 µl of cold solution S3 (3 M Na acetate, pH 4.8). The solution was mixed by inverting the tubes for 2 to 3 times. The tubes were kept on ice for 45 min and then centrifuged at 3000 rpm for 40 min at 20°C. The supernatant (400 µl) was taken in a fresh tube and 800 µl of ethanol was added and kept at room temperature for 30 min. The tubes were centrifuged at 3000 rpm for 40 min at 12°C. The supernatant was discarded and the tubes were kept open for 5 min. The DNA pellet was dissolved in 200 µl of solution S4 (10 mM Tris, 1 mM EDTA) containing 10 µg/ml of RNase. The tubes were vortexed and incubated at 65°C for 10 min. They were vortexed again and 400 µl of ethanol was added and kept at room temperature for 30 min. The tubes were centrifuged at 3000 rpm for 40 min. The supernatant was discarded, the pellet was air dried and dissolved in 100 µl of distilled water. For sequencing DNA, an additional polythelene glycol (PEG) precipitation step was necessary, which was performed as follows. The 30 µl of DNA solution was diluted to 64 µl, mixed gently, and precipitated with 16 µl of 4 M NaCL and 80 µl of 13% PEG 8000. After 30 min on ice, the tubes were centrifuged at 4°C, and the pellet was carefully rinsed with 70% ethanol, air dried, and diluted in 20 µl of distilled water.

2.7.7 Sizing of inserts

Insert sizes were determined by agarose gel electrophoresis after cleavage with *Bgl*III or *Bam*HI. DNA was cleaved using restriction enzymes in a 20 µl total reaction volume as recommended by the manufacturer and then loaded onto a 1% agarose gel and migrated.

Five microliter of plasmid DNA was run in the 1% agarose gel, stained with ethidium bromide and seen under UV light to confirm the expected size of the plasmid constructs.

2.8 Direct sequencing

For each sequencing reaction 7 µl of plasmid DNA (300 to 500 ng), 2 µl of primer (2 µM), 8 µl of reaction mix of a *Taq* DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems), and 3 µl of distilled water were used. After 26 cycles (96°C for 30s; 56°C for 15s; 60°C for 4 min) in a thermal cycler (MJ Research Inc., Watertown, Mass.), DNA was precipitated with 70 µl of 70% ethanol-0.5 mM MgCl₂, centrifuged, rinsed with 70% ethanol, dried, and dissolved in 2 µl of formamide-EDTA buffer. Samples of plasmid clones were loaded onto 64-lane, 6% polyacrylamide gels, and electrophoresis was performed on a model 373A automatic DNA sequencer (Applied Biosystems) for 12 to 16 h.

2.9 Large scale DNA preparation

Large amounts of plasmid DNA for vaccination were prepared using both QIAfilter protocol (Qiagen) and the method as described by Sambrook (158). All expression vectors were grown with vigorous aeration in *E. coli* strains. A single colony was inoculated in 30-ml culture and grown to late log phase (i.e., to an OD₆₀₀ of ~ 0.6). Twenty five milliliter of the late-log-phase culture was inoculated in 500 ml LB broth containing antibiotic in a 2-liter flask and incubated at 37°C overnight. Cells were harvested and lysed by a modification of the alkaline, sodium dodecyl sulphate procedure (158). The modification consisted of increasing the volumes three-fold for cell lysis and DNA extraction. The resulting DNA was extracted with an equal volume of chloroform/isoamyl alcohol, centrifuged at 6000 × g for 10 min at room temperature. The aqueous phase was transferred to a fresh tube and was phenol/chloroform extracted. The aqueous phase was transferred to a fresh tube and precipitated with 0.6 vol of isopropanol. The precipitate was washed with 70% ethanol. The supernatant was removed and the DNA was resuspended in 0.9% NaCl for injection into mice. The

concentration and purity of each DNA preparation was determined by measuring the OD_{260/280}. The 260/280 ratios were ≥ 1.8 . DNA was also prepared using the QIAfilter Giga protocol (Qiagen). Using this kit 5-10 mg of purified DNA was prepared at a time. DNA was resuspended in 0.9% NaCl and stored at -70°C for further use.

2.10 DNA Vaccination

BALB/c female mice of 6 to 8 weeks were purchased from IFFA-CREDO, France. Five PE-PGRS plasmid constructs (Rv0980c, Rv1068c, Rv1067c, Rv1441c and Rv1983) based on the DNA vaccination vector, V1Jns.tPA were used for DNA vaccination in mice. The mice were anesthetized by intraperitoneal injection of ketamine. Ketamine was used at the dose of 100 mg/kg. Five percent glucose was added to concentrated ketamine to make a solution of 4 mg/ml and from that 0.5 ml was injected into the peritoneum of each mouse. Ten to 15 min after ketamine injection, plasmid DNA was injected intramuscularly. Five to ten mice were in each immunization group. PE-PGRS DNA was administered intramuscularly in each quadriceps muscle in 0.05 ml of 2 mg/ml of DNA in 0.9% NaCl using a 0.3 ml insulin syringe (101). DNA was injected 3 times at 3 week intervals. As negative controls, mice were uninjected, injected with saline or with control plasmid DNA (not containing a gene insert) in saline. Pooled sera from blood samples were taken 3 weeks after the third immunization.

2.11 Expression in HeLa and HEK 293 cells

To verify the expression of PE-PGRS proteins and the specificity of the corresponding antibodies, HeLa and HEK 293 epithelial cells were transfected with the same constructs used for DNA vaccination and expression of the five PE-PGRS proteins (Rv0980c, Rv1068c, Rv1067c, Rv1441c and Rv1983) was observed by both immuno-fluorescence and Western blot analysis. HeLa cells were grown routinely in minimal essential medium (MEM, Gibco Life Technologies, Paisley, Scotland) with 10% fetal calf serum (FCS, Gibco) and supplemented with penicillin and streptomycin (Gibco). The human embryonic kidney cell line 293 was cultured in DMEM (Gibco) with 10% FCS and

antibiotics. For routine cell culture, cells were trypsinized (0.5% trypsin, Gibco, in PBS with 1 mM EDTA) and passaged for normal maintenance, one 10 cm dish into 5 dishes. For transfection experiments, cells were trypsinized and plated onto glass coverslips to give approximately 50% confluency the following day. Cells were then transfected with each plasmid construct using FuGene transfection reagent (Roche Biochemicals, Indianapolis, IN). A ratio of FuGene reagent:DNA of 3:1 (μl and μg respectively) was used in each well of a 6-well plate. In a small sterile tube, sufficient serum-free medium as diluent for FuGene transfection reagent is added to a volume of 100 μl . Three to six μl of FuGene reagent was added directly into that medium. The order of addition is critical. The serum-free medium must be pipetted into the tube first. The tube was tapped gently to mix. DNA solution of 1 to 2 μg (0.02-2.0 $\mu\text{g}/\mu\text{l}$) was added to the prediluted Fugene reagent. A total volume of DNA solution between 0.5-50 μl was used. The contents of the tube were mixed by gently tapping the tube. The mixture was incubated for a minimum of 15 min at room temperature. The complex mixture was dropwise added to the cells, distributing it around the well. The wells were swirled to ensure even dispersal. The cells were then put back in the incubator. Immunofluorescence staining was performed the following day HeLa and HEK 293 cells were plated onto coverslips or in 6-well plates at a density of 1×10^5 cells/ml and transfected the following day using Fugene reagent with 500 ng of plasmid PE-PGRS DNA or vector DNA alone, according to the manufacturer's directions. Analyses were performed 48 hours post transfection. For immunofluorescence, cells were fixed with 4% paraformaldehyde in PBS for 15 min, permeabilized with 0.2% Triton-X100 in PBS and subsequently stained with the corresponding antibody. Following labelling with Cy3-linked anti-rabbit antibodies (Jackson ImmunoResearch Laboratory, West Grove, PA), coverslips were mounted and viewed with a conventional immunofluorescence microscope (BX50, Olympus Optical). For Western blot analysis, HeLa and HEK 293 cells grown in 6-well plates were rinsed once with PBS and lysed with lysis buffer (20 mM HEPES pH 7.8, 10 mM KCl, 1 mM DTT, 0.5 mM EDTA, 0.5 mM EGTA, 1% Nonidet P-40 and a protease inhibitor tablet (mini-Complete, Roche). Samples were vortexed and centrifuged at 14,000 rpm for 5 min. The resulting supernatant was then run on a 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), transferred to nitrocellulose and

specific antibody binding was revealed using an enhanced chemiluminescence detection system (Amersham, Aylesbury, UK).

2.12 Expression in reticulocyte extracts

Two of the PE-PGRS genes were expressed in reticulocyte extracts using TNT^R Coupled Transcription/Translation Systems (Promega). The TNT^R Quick Coupled Transcription/Translation Systems are convenient single-tube, coupled transcription/translation reactions for eukaryotic in vitro translation. Standard rabbit reticulocyte systems commonly use RNA synthesized in vitro from SP6, T3 or T7 RNA polymerase. The TNT^R Quick Coupled Transcription/Translation System further simplifies the process by combining the RNA Polymerase, nucleotides, salts and Recombinant RNasin^R Ribonuclease Inhibitor with the reticulocyte lysate solution to form a single TNT^R Quick Master Mix. For most gene constructs, the TNT^R Quick reaction produces significantly more protein (2 to 6-fold) in a 60-90 minute reaction than do standard in vitro rabbit reticulocyte lysate reactions using RNA templates. Rv1068c and Rv1441c genes were cloned in the pET22 vector. Forty µl of TNT^R T7 Quick Master Mix, 2 µl of ³⁵S methionine (1,000 Ci/mmol), one µg plasmid DNA, nuclease-free water to a final volume of 50 µl are taken in a 1.5 ml microcentrifuge tube, gently mixed by pipetting. The reaction mixture was incubated at 30°C for 60-90 min. The synthesized proteins were analysed by SDS-PAGE and autoradiography.

2.13 Western blotting

Different mycobacteria and clinical isolates of *M. tuberculosis* were grown in Middlebrook 7H9 medium. Harvested bacteria were resuspended in 1/20 of the original culture volume in phosphate-buffered saline, lysed, and stored at -20°C until use. Extracted proteins were separated by SDS-PAGE in phosphate-buffered saline on slab gels of 12% polyacrylamide. Ten µg of total protein was loaded in each lane of the gel. A mixture of standard protein markers was used for the determination of molecular mass. For immunoblotting analysis, proteins were transferred from polyacrylamide gels onto nitrocellulose membranes (181). Eight to ten pieces of Whatman 3MM filter paper were

cut to the same size as the gel, prewetted with electroblotting buffer (20 mM Tris/150 mM glycine, pH 8.0) and placed in the tank. The filter paper was placed toward the negative electrode. A piece of cut, marked, and wetted nitrocellulose membrane filter was placed on the filter paper and all air bubbles were removed. The gel was placed on the nitrocellulose membrane and the surface of the gel was moistened with electroblotting buffer. Any air bubble was removed between gel and filter paper by gently pushing and lifting the gel as needed. Gloves were used when manipulating filter papers, gels, and nitrocellulose membranes. Pieces (8-10) of wetted Whatman filter paper were placed on the cathode side of the gel and all air bubbles were removed. This Western blot sandwich was placed into a plastic support. The sandwich was placed into the electroblotting apparatus in the correct orientation so that the nitrocellulose faces the anodal side of the tank. The tank was filled with electroblotting buffer. Electrophoretic transfer of the proteins from the gel to the nitrocellulose was done at 14 V constant voltage for 2 hrs. The nitrocellulose filter was placed in Ponceau S solution for one min to stain proteins. The filter was completely destained in water for 10 min. Nonspecific antibody sites on the nitrocellulose were blocked with blocking solution (1 g of instant nonfat dry milk in 100 ml PBS) for 1 hr at room temperature on a rocking platform. To test the immune responses of murine antibodies to different mycobacterial proteins (primary antibodies), blots were hybridized with serum samples diluted 1:100. The blot with diluted primary antibody was incubated at room temperature for 1 hr on a rocking platform. Nonspecifically bound primary antibody was removed by washing the filter 4 times by agitating with 200 ml of PBS each time. Bound antibody was revealed by incubation with peroxidase-conjugated anti-mouse immunoglobulin.

2.14 Sub-cellular fractionation

Sub-cellular fractions of *M. tuberculosis* (produced through funds from the National Institutes of Health, National Institute of Allergy and Infectious Diseases, Contract NO1-AI-75320) were obtained from the Department of Microbiology, Colorado State University, USA. Cell wall, cell membrane, cytosol and whole cell lysate were separated using the procedure described previously (94), (113). In brief, each strain was grown to

late-log-phase (day 14) in glycerol-alanine-salts (GAS) medium, washed with PBS pH 7.4 and inactivated by gamma-irradiation. The bacilli are suspended (2 g/ml) in PBS containing DNase, RNase and a proteinase inhibitor cocktail (PMSF, pepstatin A, and leupeptin) and broken in a French Press pressure cell at 4°C. Unbroken cells were removed by low speed ($3,000 \times g$) centrifugation. The supernatant isolated is a crude or whole cell isolate containing the cell wall, cytosol, and membrane. The cell wall was isolated by centrifugation at $27,000 \times g$ for one hour and washed 2 times in PBS. The final cell wall pellet is suspended in PBS. The supernatant was subjected to a $100,000 \times g$ centrifugation for four hours. The resulting membrane pellet was washed with PBS and suspended in 0.01 M ammonium bicarbonate. In order to obtain cytosol, the supernatant was collected, dialyzed against 0.01M ammonium bicarbonate. Protein contents of all fractions were quantified using the BCA protein assay, and aliquots were stored at -80°C.

Subcellular fractions of *M. smegmatis* were prepared by methods described previously (6, 21). Two hundred fifty millilitre of cells were grown in Middlebrook 7H9 broth to an OD₆₀₀ of 0.5 to 0.8. The cells are taken in 50 ml polypropylene tubes and are centrifuged at $10,000 g$ for 20 min. Proteins in culture supernatant were precipitated with 10% (v/v) trichloroacetic acid. The precipitated material was resuspended in 1/20 of the original volume in PBS. Cell extracts were obtained by sonication in an ice bath for 10 min. The lysate was cleared by centrifugation at $20,000 g$ for 15 min. The supernatant was centrifuged at $100,000 g$ for 2.5 h. The pellet (membranes) was washed with PBS and resuspended in 2 ml distilled water. All fractions were resuspended in lysis buffer and an amount of protein equivalent to 300 µl culture was electrophoresed per lane of an SDS-polyacrylamide gel (10%, w/v, polyacrylamide) followed by Western blotting to sera raised in mice.

2.15 Immunoelectron microscopy

Immunoelectron microscopy was performed as follows. Nickel grids previously coated with carbon on a formvar film were laid on a bacterial suspension for 3 min then fixed with 4% para-formaldehyde in PBS for 20 min at room temperature. Grids were

incubated for 20 min in PBS containing 0.25% NH₄Cl then blocked by Aurion Blocking solution (Aurion 905002) for 20 min. This was followed by incubation for one hour at room temperature with antibody 1441 (1/5 dilution) in PBS BSA-C TM Aurion. After 5 washes in PBG (PBS, 0.5% BSA, 0.1% Gelatin [Sigma G-7765]), the grids were incubated with anti-mouse IgG antibody (1/25 dilution) conjugated with 10 nm gold particles in PBG (Amersham RPN 431), washed and negatively stained. Grids were examined in a J  ol 1200 EX II electronic microscope at an accelerating voltage of 80 Kv.

2.16 RD (region of difference)-PCR analysis

RD-PCR analysis was done using the methods described by Brosch et al. (29). Sequences inside or flanking RD and RvD regions were obtained from the web sites <http://genolist.pasteur.fr/TubercuList/> and http://www.sanger.ac.uk/Projects/M_bovis/. Primers were designed by using the Primer3 web site <http://www-genome.wi.mit.edu/cgi-bin/primer3-www.cgi> that would amplify ~500 bp fragments (Table 5). Polymerase chain reactions containing per reaction 1.25 µl of 10 × PCR buffer (600 mM Tris HCL pH 8.8, 20 mM MgCl₂, 170 mM (NH₄)₂SO₄, 100 mM β-mercaptoethanol), 1.25 µl 20 mM nucleotide mix, 50 nM of each primer, 1-10 ng of template DNA, 10% DMSO, 0.2 units *Taq* polymerase (Gibco-BRL) and sterile distilled water to 12.5 µl were performed on a PTC-100 amplifier (MJ Inc.) with an initial denaturation step of 90 seconds at 95°C, followed by 35 cycles of 30 seconds at 95°C, 1 min at 58°C, and 4 min at 72°C.

2.18 Spoligotyping

Spoligotyping was performed as previously described by Kamerbeek et al. (103). The DR region was amplified by PCR with oligonucleotide primers derived from the DR sequence. Mycobacterial genomic DNA was extracted from cultured cells as described previously (107, 109) Approximately 10 ng of purified DNA was used as a target for spoligotyping. Fifty microlitres of the following reaction mixture were used for the PCR: *Tth* buffer (5 mM TrisHCl, 5 mM KCl, 0.7 mM MgCl₂, pH 9.0), each deoxynucleoside triphosphate at 200 mM, 20 pmol each of forward and reverse primers DR, 10 ng of

Table 5. Primers used in Deletion analysis

Region absent from BCG	Gene	Size kb	Internal primer pair	Flanking primers or 2 nd internal* primer pair
RD1	Rv3871-Rv3879c	9.5	RD1in-Rv3878F GTC AGC CAA GTC AGG CTA CC	RD1-flank left GAA ACA GTC CCC AGC AGG T
			RD1in-Rv3878R CAA CGT TGT GGT TGT TGA GG	RD1-flank.right TTC AAC GGG TTA CTG CGA AT
RD2	Rv1978-Rv1988	10.8	RD2-Rv1979.int.F TAT AGC TCT CGG CAG GTT CC	RD2-flank.F CTC GAC CGC GAC GAT GTG C
			RD2-Rv1979.int.R ATC GGC ATC TAT GTC GGT GT	RD2-flank.R CCT CGT TGT CAC CGC GTA TG
RD3*	Rv1573-Rv1586c	9.2	RD3-Rv1586.int.F TTA TCT TGG CGT TGA CGA TG	RD3-int-REP.F CTG ACG TCG TTG TCG AGG TA*
			RD3-Rv1586.int.R CAT ATA AGG GTG CCC GCT AC	RD3-int-REP.R GTA CCC CCA GGC GAT CTT*
RD4	Rv1505c-Rv1516c	12.7	RD4-Rv1516.int.F CAA GGG GTA TGA GGT TCA CG	RD4-flank.F CTC GTC GAA GGC CAC TAA AG
			RD4-Rv1516.int.R CGG TGA TTC GTG ATT GAA CA	RD4-flank R AAG GCG AAC AGA TTC AGC AT
RD5*	Rv2346c-Rv2353c	9.0	RD5A-Rv2348.int.F AAT CAC GCT GCT GCT ACT CC	RD5B-pleA.int.F CAA GTT GGG TCT GGT CGA AT
			RD5A-Rv2348.int.R GTG CTT TTG CCT CTT GGT C	RD5B-pleA.int.R GCT ACC CAA GGT CTC CTG GT
RD6*	Rv3425-Rv3428c	4.9	RD6-IS1532F CAG CTG GTG AGT TCA AAT GC	ND
			RD6-IS1532R CTC CCG ACA CCT GTT CGT	ND
RD7	Rv1964-Rv1977	12.7	RD7-Rv1976.int.F TGG ATT GTC GAC GGT ATG AA	RD7-flank.F GGT AAT CGT GGC CGA CAA G
			RD7-Rv1976.int.R GGT CGA TAA GGT CAC GGA AC	RD7-flank.R CAG CTC TTC CCC TCT CGA C
RD8	<i>ephA-lpqG</i>	5.9	RD8-ephA.F GGT GTG ATT TGG TGA GAC GAT G	RD8-flank.F CAA TCA GGG CTG TGC TAA CC
			RD8-ephA R AGT TCC TCC TGA CTA ATC CAG GC	RD8-flank.R CGA CAG TTG TGC GTA CTG GT
RD9	<i>cobL</i> -Rv2075	2.0	RD9-intF CGA TGG TCA ACA CCA CTA CG	RD9-flankF GTG TAG GTC AGC CCC ATC C
			RD9-intR CTG GAC CTC GAT GAC CAC TC	RD9-flankR GCC CAA CAG CTC GAC ATC
RD10	Rv0221-Rv0223	1.9	RD10-intF GTA ACC GCT TCA CCG GAA T	RD10-flankF CTG CAA CCA TCC GGT ACA C
			RD10-intR GTC AAC TCC ACG GAA AGA CC	RD10-flankR GTC ATG AAC GCC GGA CAG
RD11	Rv2645-Rv2659c	11.0	RD11-Rv2646F CGG CAG CTA GAC GAC CTC	RD11-fla-F TCA CAT AGG GGC TGC GAT AG

			RD11-Rv2646R AAC GTG CTG CGA TAG GTT TT	RD11-fla-R AGA GGA ACC TTT CGG TGG TT
RD12	sseC-Rv3121	2.8	RD12-Rv3120.int.F GAA ATA CGA GTG CGC TGA CC	RD12-flank.F GCC ATC AAC GTC AAG AAC CT
			RD12-Rv3120.int.R CTC TGA ACC ATC GGT GTC G	RD12-flank.R CGG CCA GGT AAC AAG GAG T
RD13	Rv1255c-Rv1257c	3.0	RD13intF GGA TGT CAC TCG GAA CGG CA	RD13-flank.F CGA TGG TGT TTC TTG GTG AG
			RD13intR CAC CGG GCT GAT CGA GCG A	RD13-flank.R GGA TCG GCT CAG TGA ATA CC
RD14	Rv1765c-Rv1773c	9.0	RD14-Rv1769.int.F GTG GAG CAC CTT GAC CTG AT	RD14-flankF TTG ATT CGC CAA CAA CTG AA
			RD14-Rv1769.int.R CGT CGA ATA CGA GTC GAA CA	RD14-flankR GGG CTG GTT AGT GTC GAT TC
Region missing from <i>M. tuberculosis</i> H37Rv				
RvD1*		5.0	RvD1-int1F AGC GCG TCG AAC ACC GGC	RvD1-int2.F GAG CCA CTC CGA TGT TGA CT
			RvD1-int1R CCT GAA TCC GCG CAA TTC CAT	RvD1-int2.R CAC GCG AAC CCT ACC TAC AT
RvD2*	plcD	5.1	RvD2-int1F GTT CTC CTG TCG AAC CTC CA	RvD2-int2F GGA CGG TGA CGG TAT TTG TC
			RvD2-int1R ACT TCA CCG GTT TCA TCT CG	RvD2-int2R TCG CCA ACT TCT ATG GAC CT
RvD3		1.0	RvD3-intF ATC GAT CAG GTC GTC AAT GC	RvD3-flank.F AAA CCA TGC AGC GTC TGC CA
			RvD3-intR ACG CCA CCA TCA AGA TCC	RvD3-flankR GCG TTT CTG CGT CTG GTT GA
RvD4*	PPE gene	0.8	RvD4-intF-PPE GGT TGC CAA CGT TAC CGA TGC	ND
			RvD4-intR-PPE CCG GTG GTG GTG GCG GCT	ND
RvD5	moa	4.0	RvD5intF GGG TTC ACG TTC ATT ACT GTT C	RvD5-flankF CCC ATC GTG GTC GTT CAC C
			RvD5intR CCT GCG CTT ATC TCT AGC GG	RvD5-flankR GTA CCC GCA CCA CCT GCT G
TbD1	mmpS6 mmpL6	2.1	TbD1intS.F CGT TCA ACC CCA AAC AGG TA	TbD1fla1-F CTA CCT CAT CTT CCG GTC CA
			TbD1intS.R AAT CGA ACT CGT GGA ACA CC	TbD1fla1-R CAT AGA TCC CGG ACA TGG TG

DNA, and 0.5 U of *Tth* polymerase (HT Biotechnology Ltd.). The mixture was heated for 3 min at 96°C and subjected to 20 cycles of 1 min at 96°C, 1 min at 55°C, and 30 s at 72°C. The amplified product was hybridized to a set of 43 immobilized oligonucleotides, each corresponding to one of the unique spacer DNA sequences within the DR locus. These oligonucleotides were covalently bound to a membrane. In short, a membrane (Biodyne C; Pall Biosupport, Portsmouth, United Kingdom) was activated by using 16% (wt/vol) 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (Sigma). The oligonucleotides were applied to the membrane in parallel by using a miniblotted system. After a short incubation, the membrane was inactivated by using 100 mM NaOH and washed in 2× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH₂PO₄, and 1 mM EDTA, pH 7.7; Gibco BRL Life Technologies, Inc.) supplemented with 0.1% sodium dodecyl sulfate (SDS; Sigma). For hybridization, 20 µl of the amplified PCR product was diluted in 150 µl of 2× SSPE-0.1% SDS and heat denatured. The diluted samples (130 µl) were pipetted into the parallel channels in such a way that the channels of the miniblotted apparatus were perpendicular to the rows of oligonucleotides deposited previously. Hybridization was done for 60 min at 60°C. After hybridization, the membrane was washed twice in 250 ml of 2× SSPE-0.5% SDS for 10 min each time at 60°C and then incubated in 1: 4000-diluted streptavidin-peroxidase conjugate (Boehringer) for 45 to 60 min at 42°C. The membrane was washed twice, for 10 min each time, in 250 ml of 2× SSPE-0.5% SDS at 42°C and rinsed with 250 ml of 2× SSPE for 5 min at room temperature. Detection of hybridizing DNA was done by chemiluminescent ECL (Amersham) detection (187, 188), followed by exposure to X-ray film (Hyperfilm ECL, Amersham) in accordance with the instructions of the manufacturer. For repeated use of membranes, the membranes were stripped by washing two times for 30 min, in 1% SDS at 80°C, and then incubated for 15 min in 20 nM EDTA (pH 8) at room temperature. Membranes were sealed in plastic and stored at 4°C until further use.

3. Results

3.1 Expression of PE-PGRS genes by RT-PCR

Studies were carried out to investigate the biological and immunological role of some of the PE-PGRS proteins. Ten PE-PGRS genes were selected (Table 4) with the aim of monitoring expression at the transcriptional level in *in vitro*-cultured *M. tuberculosis* cells by RT-PCR (Fig. 8). Smaller genes were picked in order to make PCR amplification easier with the size of the selected PE-PGRS genes varying from 1 to 2 kb. Suitable primers were designed to amplify regions of between 200 to 300 bp of each PE-PGRS gene following cDNA production (Table 4). Duplex RT-PCR results for the 10 selected PE-PGRS genes are shown in the figure 8. A 224 bp segment of 16S rRNA was amplified in the same PCR reaction as the internal control and this was generally seen below the products derived from the PE-PGRS mRNA on agarose gel electrophoresis (Fig. 8). The size of the PCR product of Rv2162 is 219 bp and the band appears just below the PCR product of 16S rRNA. As the negative control same reaction mixture was used without reverse transcription. No amplicons were obtained in the absence of reverse transcription. These experiments demonstrated that all 10 PE-PGRS genes examined were expressed at the transcription level in cultured mycobacteria (Fig. 8).

3.2 Cloning of PGRS genes into DNA vaccination vector

The five PGRS genes (Rv0980c, Rv1067c, Rv1068c, Rv1441c, Rv1983) were taken for cloning into VIJns.tPA vector (129)(Fig. 7). VIJns.tPA is a expression vector derived from the earlier vector VI (183). The PCR amplified products of the PE-PGRS genes were restriction enzyme digested and ligated into the corresponding enzyme sites of the DNA vaccination vector. The PGRS genes were cloned downstream of the cytomegalovirus promoter then transformed into *E. coli*. The recombinant plasmids were prepared after overnight growth of *E. coli* and digested with appropriate restriction enzymes, electrophoresed on agarose gels, and checked for the appropriate size. All of the clones (Rv0980c, Rv1067c, Rv1068c,

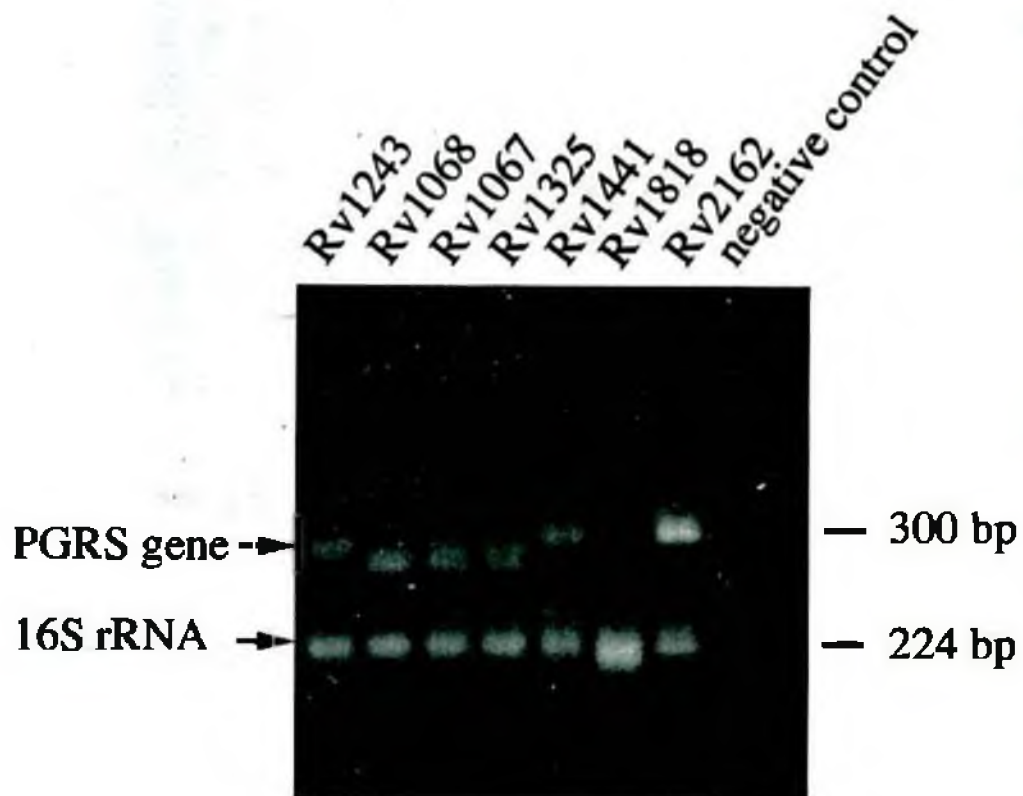


Fig. 8. Duplex RT-PCR analysis of total RNA extracted from *Mycobacterium tuberculosis*. Amplicons of seven PE-PGRS genes are shown. The size of amplicons varies between 200 to 300 bp. A 224 bp segment of 16S rRNA of *M. tuberculosis* was amplified in the same PCR as the internal control.

Rv1441c, Rv1983) were of the correct sizes. For each PE-PGRS gene five clones were picked for sequence analysis. About 400-500 bp segment could be read in each sequencing reaction. Primers from the internal sequences of the gene were designed for sequencing the whole gene avoiding the highly repetitive part. In most cases, it was very difficult to sequence the whole gene. Sometimes primers had to be re-designed. One of the Rv0980c clones had correct sequences for the first 200-400 bases and then the sequences were completely different from that of the original Rv0980c. Four others were found to have correct sequences except for 1-2 mismatched bases. One of them had one mismatched base without any amino acid change and no reading-frame shift, and, was chosen for DNA vaccination experiments. Rv1068c had a similar one base mismatching in one of the clones which was picked for further investigation. Rv1067c, Rv1441c and Rv1983 clones had the correct sequence. Thus the five selected PE-PGRS genes were successfully cloned into the DNA vaccination vector though PCR amplification of PE-PGRS genes was sometimes difficult due to the presence of highly repetitive sequences.

3.3 Antibody production using DNA vaccination

After confirming that the selected PE-PGRS genes were expressed in *M. tuberculosis* H37Rv as shown by RT-PCR, DNA vaccination was used to produce antibodies against the selected PE-PGRS proteins in mice. DNA vaccination has been demonstrated to be effective in inducing both antibody and cell-mediated immune responses (56, 63, 101, 125, 177, 183, 193). Wolf et al. demonstrated that DNA expression vectors can be taken up by muscle when injected into mice intramuscularly (i.m.) without the use of a delivery vehicle and can persistently express genes encoded by the plasmid (194, 195). The genes for five PE-PGRS proteins (Rv0980c, Rv1067c, Rv1068c, Rv1441c, Rv1983) were cloned downstream of the cytomegalovirus promoter in the V1Jns.tPA vector, and the recombinant plasmids were injected i.m. three times at 3 to 4 week intervals to immunize mice and induce antibody responses. As negative controls mouse were kept uninjected or injected with plasmid DNA without inserted mycobacterial sequences or only with saline. Sera from the mice were collected 3 weeks after the 3rd DNA injection.

3.4 Expression in HeLa cells and immune detection

The same plasmids used for DNA vaccination were used to express proteins in eukaryotic cells in order to determine the specificity of the resultant murine antibodies against the PE-PGRS proteins. The PE-PGRS constructs and the DNA vaccination vector without any insert were transfected into HeLa and HEK 293 epithelial cells. The cell extracts were then probed with the corresponding antibody and immune complexes visualized by fluorescence microscopy. Typical results for the construct producing Rv1441c are presented (Fig. 9) and similar results were obtained with Rv0980c, Rv1067c, Rv1068c, and Rv1983. No immune complexes were visualized when the sera from the mice immunized with the vector alone were tested. These experiments showed that the PE-PGRS genes were strongly expressed in both HeLa and HEK 293 (data not shown) cell lines.

After confirming that the PE-PGRS proteins were well-expressed in both HeLa and HEK 293 cells by fluorescence microscopy experiments, proteins were extracted from HeLa and HEK 293 cells, transfected with the above five plasmids expressing the PE-PGRS proteins, or the vector alone. To test the specificity of the antibodies, extracted proteins were examined by Western blotting using antibodies raised against the corresponding clone. The predicted sizes of the Rv0980c and Rv1441c proteins are about 42 and 40.7 kDa, respectively (Table 4). The expression of both PE-PGRS proteins was greater in HeLa cells than in HEK 293 cells. Antibody 0980c and antibody 1441c specifically recognized single proteins in both transfected cell lines but not in control extracts (Fig. 10). However, in contrast to Rv0980c, the Rv1441c protein displayed an aberrantly high apparent molecular weight (~60 kDa).

To determine whether the five PE-PGRS proteins tested share similar epitopes, dot blot experiments were carried out with transfected HeLa cell extracts using all five antibodies.

A



B

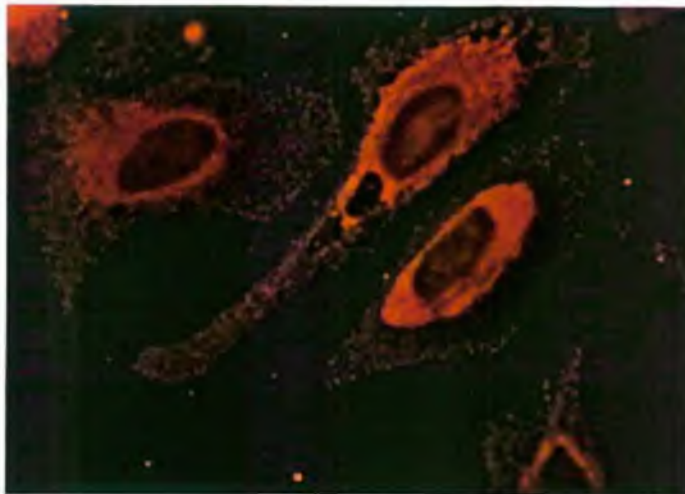


Fig. 9. Expression of Rv1441 in HeLa epithelial cells. HeLa cells were transfected with the vector alone (A) or the Rv1441 construct (B). Cells were stained with the Rv1441-specific antibody and visualized by fluorescence microscopy. Strong perinuclear staining of Rv1441 is seen in (B).

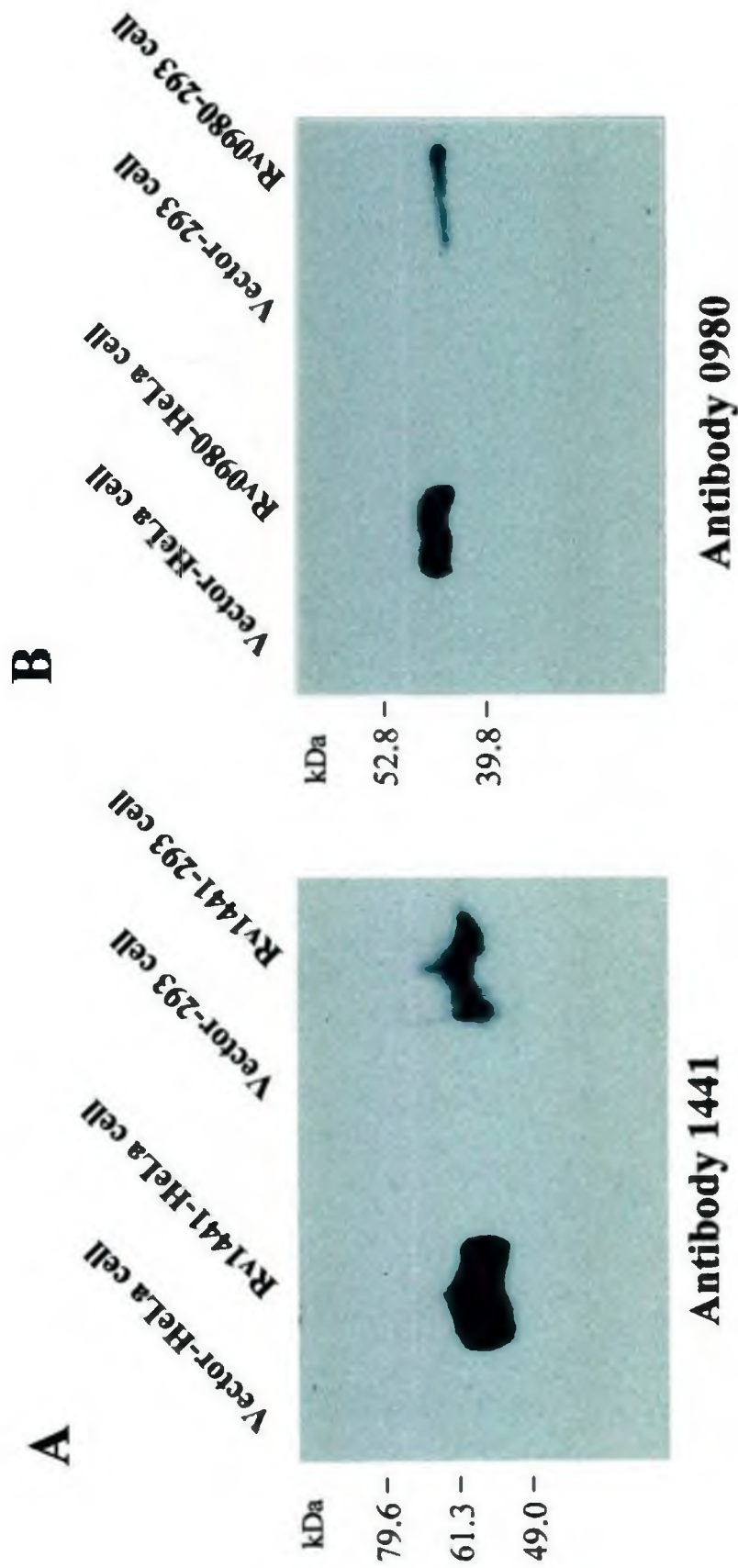


Fig. 10. Western blot analysis of Rv1441 and Rv0980 proteins expressed in eukaryotic cells. The same plasmid constructs used for DNA vaccination were transfected into HeLa and HEK 293 epithelial cells. Vector without insert was transfected to use as a negative control. Protein extracts from HeLa and 293 cells transfected with plasmid Rv1441 immunoblotted with antibody 1441 (A). Similar experiments were done using plasmid Rv0980 and antibody 0980 (B).

Different extracts containing recombinant Rv0980c, Rv1067c, Rv1068c, Rv1441c and Rv1983 proteins were dot blotted onto nylon membranes (Table 6) and incubated with the five different murine antibodies, separately. All antibodies gave signals with their corresponding protein. All five antibodies reacted with Rv0980c and Rv1441c proteins while antibody to Rv1441c also cross-reacted with the Rv1983 protein (Fig. 11). As expected from the extensive amino acid sequence conservation among the PE family members (27, 48) these results clearly indicate that the different antibodies tested cross-react implying that some PE-PGRS proteins share common epitopes.

3.5 Expression in reticulocyte extracts and immune detection

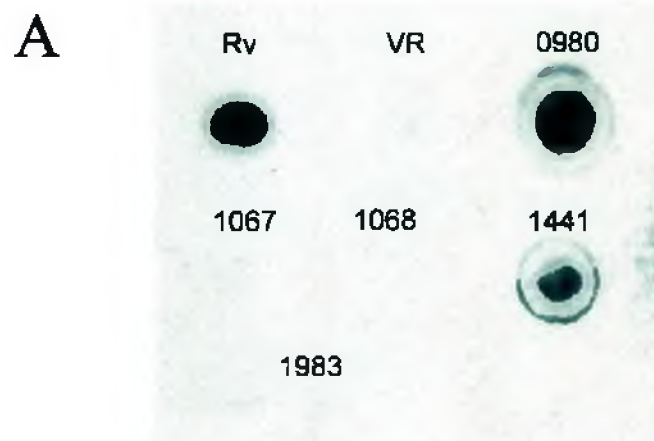
To examine whether the size discrepancy of Rv1441c protein expressed in HeLa cells was also seen in other expression systems or if it resulted from a post-translational modification, Rv1441c was produced from pET vectors in both reticulocyte extracts and in *E. coli*. In both systems the protein appeared as a ~60 kDa band suggesting that its biased amino acid composition might influence its electrophoretic mobility (Fig. 12). The size of the Rv0980c protein produced in reticulocyte extracts was identical to that seen in HeLa cells. The three other antibodies (1067c, 1068c and 1983) also detected single bands of the expected size when HeLa cell extracts were analysed on Western blots.

3.6 Detection of PE-PGRS proteins in *M. tuberculosis* H37Rv, clinical isolates and subcellular fractions

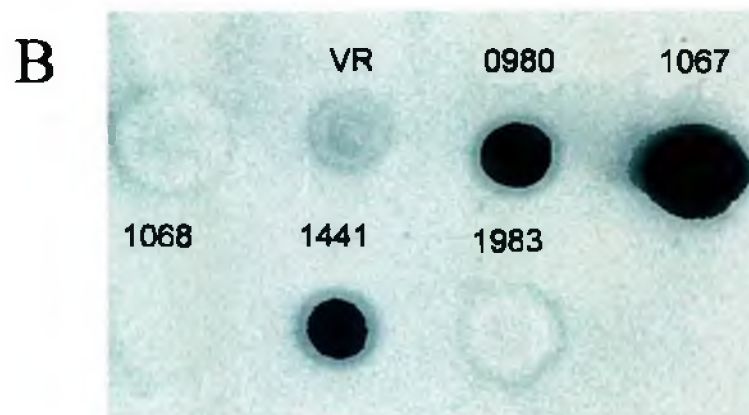
Expression of PE-PGRS proteins in eukaryotic cells and reticulocytes and detection using the corresponding antibodies confirmed the specificity of the murine antibodies against different PE-PGRS proteins. To assess the possibility of genetic variation leading to alterations in size or antigenicity among the different clinical isolates and *M. bovis* BCG Pasteur, all five antibodies were tested against the protein extracts of 11 clinical isolates by immunoblot experiments. Figure 13 (A) shows Western blot results comparing proteins of the H37Rv strain of *M. tuberculosis* and six clinical isolates (strains 1, 9, 23,

Table 6. Cross-reactivity of PE-PGRS proteins

Antibody	Proteins from transfected HeLa cells				
	Rv0980c	Rv1067c	Rv1068c	Rv1441c	Rv1983
Rv0980c	+	–	–	+	–
Rv1067c	+	+	–	+	–
Rv1068c	+	–	+	+	–
Rv1441c	+	–	–	+	+
Rv1983	+	–	–	+	+



Antibody 0980



Antibody 1067

Fig. 11. Dot-blot analysis of Rv0980, Rv1067, Rv1068, Rv1441 and Rv1983 proteins expressed in HeLa epithelial cells. Proteins extracted from vector (VR) transfected cells were used as negative control. Whole cell lysate of *M. tuberculosis* H37Rv (Rv) was also blotted on the membrane (A). Immunoblot analysis was done using antibody 0980 (A) and antibody 1067 (B).

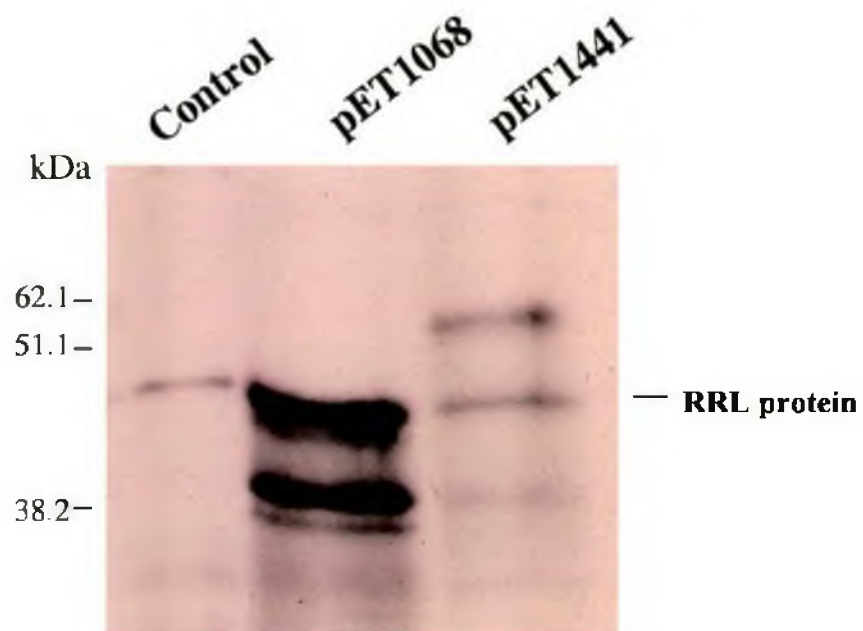


Fig. 12. In vitro expression of Rv1068 and Rv1441 proteins in reticulocytes. Rv1068 and Rv1441 constructs in pET vectors were used for eukaryotic in vitro translation. Vector without insert was used as negative control. On the background a band of about 42 kDa rabbit reticulocyte lysate (RRL) protein was seen in all lanes.

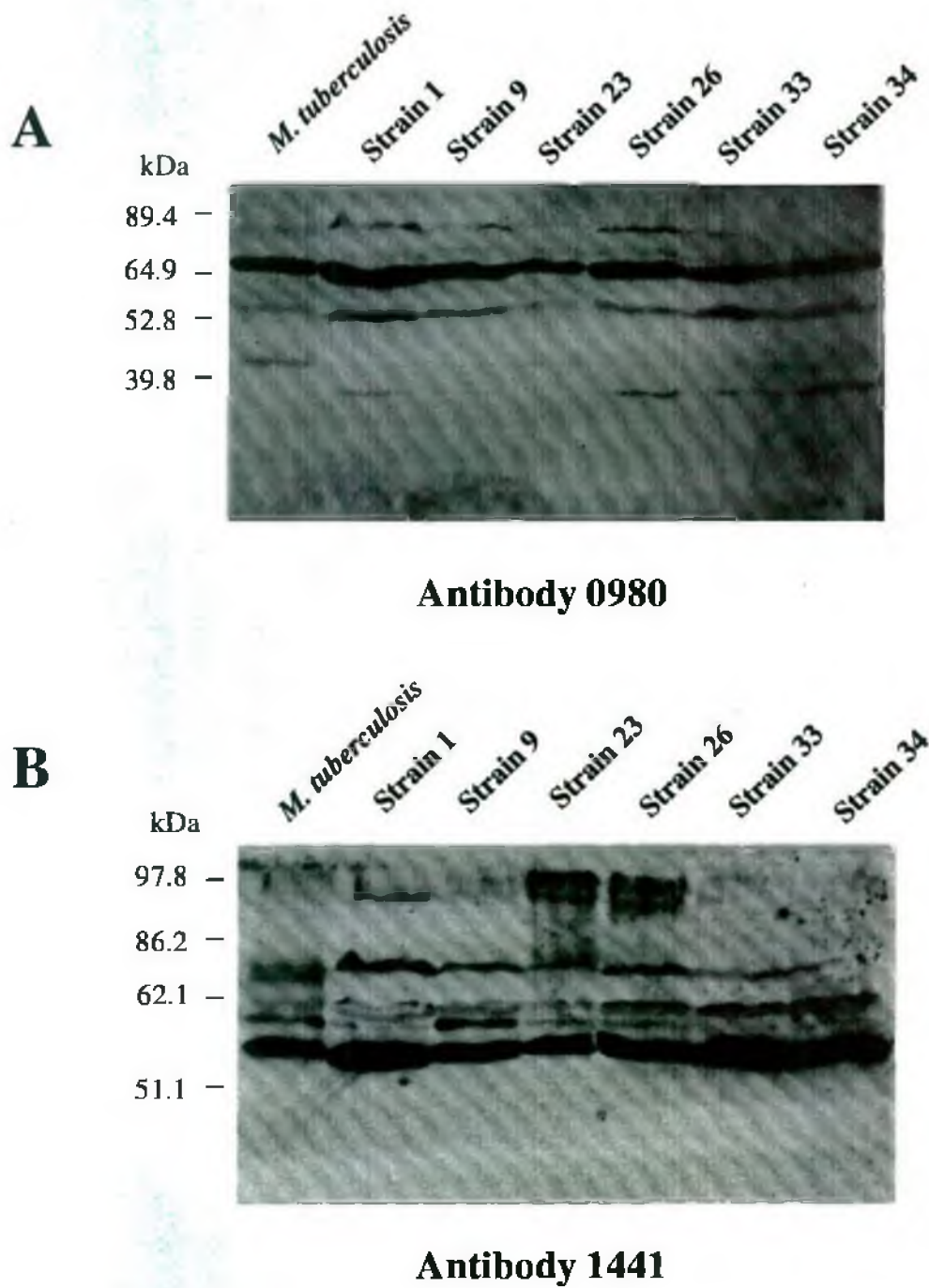


Fig. 13. Immunoblot analysis of protein extracts from different clinical isolates of *M. tuberculosis* using antibody 0980 (A) and antibody 1441 (B). Several proteins of various sizes from each strain were identified by immunoblotting using antibody 0980 (A) and antibody 1441 (B).

26, 33 and 34) using antibody Rv0980c. Similar results were obtained with the other clinical isolates (data not shown) and the various antibodies. Reproducible patterns were obtained, although some size differences were seen for the proteins in some clinical isolates. In all cases antibody Rv0980c recognized several proteins (>6) of different sizes. Figure 13 (B) shows Western blot analysis of protein extracts from clinical isolates of *M. tuberculosis* using antibody Rv1441c. Fewer protein species were detected when antibodies raised against Rv1067c, Rv1068c, and Rv1983 were used.

The predicted molecular weight of the Rv0980c protein is ~42 kDa (Table 4) and, in *M. tuberculosis* H37Rv a band of this size was detected, in addition to several larger proteins. In most of the clinical isolates, the 42 kDa protein was absent but a 36 kDa band was seen, along with other larger sized proteins. A strong band of about 65 kDa was found in *M. tuberculosis* and all clinical isolates but a larger protein of >80 kDa present in *M. tuberculosis* H37Rv and some clinical isolates was absent from patient strains 23 and 34.

The sub-cellular location of the PE-PGRS proteins is unknown and to understand their functions clearly it is essential to know where they act in the cell. Consequently, cell fractionation studies and immunoblotting were carried out. For comparison, cell wall, cell membrane and cytosol fractions were prepared from the H37Rv strain of *M. tuberculosis* and from CDC1551, a clinical isolate of *M. tuberculosis*, since the genome sequences of both organisms are available, and subjected to Western blot analysis using the antibody to Rv1441c, to the membrane protein OmpA (161), and to catalase-peroxidase, KatG (156). In both strains, several putative PGRS proteins were recognized with differing intensities by the Rv1441c antibody (Fig. 14) and the majority of them were present in the cell wall and cell membrane fractions together with OmpA, whereas KatG was predominantly cytoplasmic (data not shown). Interestingly, a 27 kDa protein present in the particulate fractions of *M. tuberculosis* H37Rv appeared to be less abundant in CDC1551 (Fig. 14).

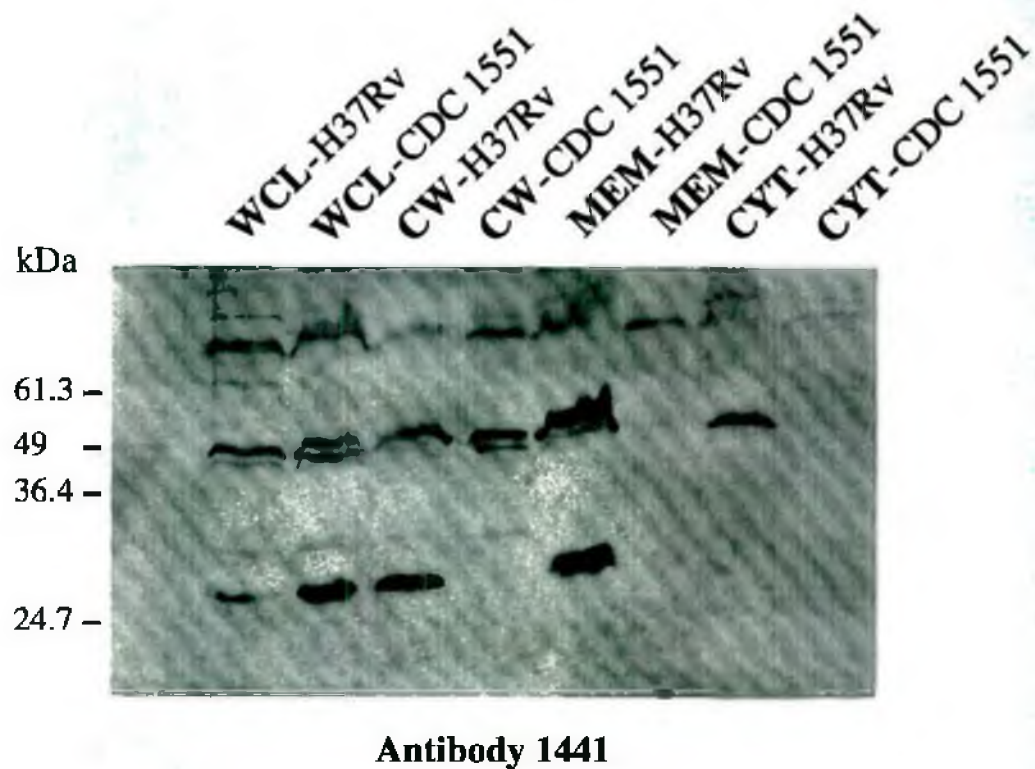


Fig. 14. Subcellular localization of PE-PGRS proteins in two strains of *M. tuberculosis*. Cell wall, cell membrane and cytosol proteins were separated by SDS-PAGE. Antibody 1441 was used to locate the PE-PGRS protein in the cell. The antibody recognized several proteins in each cell fraction: WCL, whole cell lysate; CW, cell wall fraction; MEM, cytoplasmic membrane fraction; CYT, cytosol.

From the subcellular fractionation experiments it was confirmed that some PE-PGRS proteins are located in the cell membrane and cell wall. To test the possibility that some PE-PGRS proteins were exposed on the surface of *M. tuberculosis*, electron microscopy was used to visualize immune complexes. When intact cells were reacted with antibody to Rv1441c, and examined by negative contrast, immunolabelling of the cell surface was observed (Fig. 15). In similar experiments employing thin sections of *M. tuberculosis*, immune complexes were found in both the cytoplasm and the cell envelope (data not shown) consistent with the cell fractionation results. In both experiments, no labelling was obtained when the antibody to Rv1441c was excluded or when pre-immune serum was used.

3.7 Detection of PE-PGRS proteins in other mycobacteria, subcellular fractions of *M. smegmatis*

Immunoblot experiments were also performed to detect PE-PGRS proteins in other mycobacteria that had been shown previously to possess PGRS by Southern blotting (140) namely, *M. bovis* BCG, *M. smegmatis*, *M. goodii*, and *M. marinum* and compared with *M. tuberculosis* H37Rv. Figure 16 (A) shows the results obtained with antiserum Rv1441c and it can be seen that from six to ten proteins were recognized in all five mycobacterial species and in two clinical isolates of *M. tuberculosis* (strain 23 and strain 35) with their apparent molecular weights ranging from ~25 to >150 kDa.

Figure 16 (B) shows Western blot analysis of protein extracts from different mycobacterial species with antiserum Rv0980c. In the blot, it is also seen that from three to eight proteins of different sizes were recognized in all five mycobacterial species and in same two clinical isolates of *M. tuberculosis* (strain 23 and strain 35). The number of proteins detected by antiserum Rv0980c was fewer than that detected by antiserum Rv1441c. Fewer proteins were also detected when the other three antisera were used (data not shown). In *M. marinum* a strong protein band of about 89 kDa was seen when hybridized with antiserum Rv1441. A similar strong band of about 85 kDa was also seen

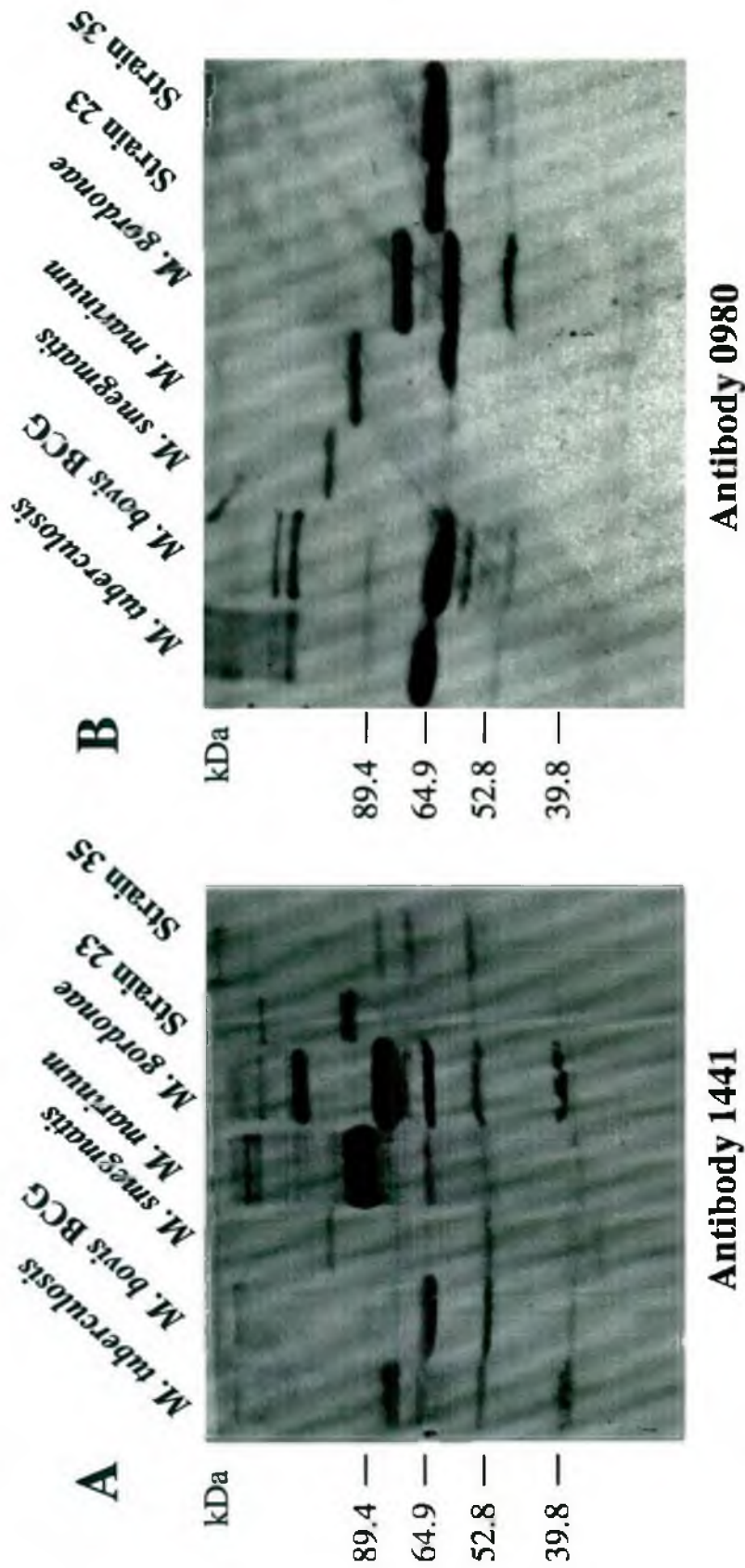


Fig. 16. Western blot analysis of protein extracts from various mycobacteria using antibody 1441 (A) and antibody 0980 (B). Total proteins were extracted from *M. tuberculosis*, *M. bovis* BCG, *M. smegmatis*, *M. marinum*, *M. goodii* and two clinical isolates of *M. tuberculosis* (strain 23 and strain 35), separated by SDS-PAGE and analysed by immunoblotting with antibody 1441 (A) and antibody 0980 (B).

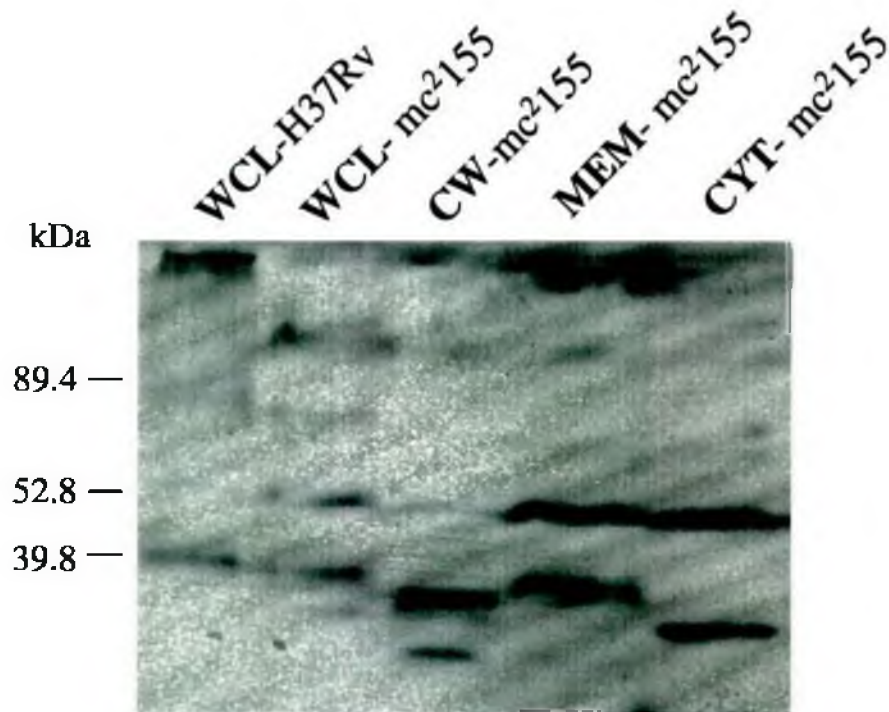
in *M. marinum* with antiserum Rv1067. As negative controls, mice without DNA injection, immunized with the vector alone were used. When serum was used from these control mice, none of these proteins was recognized.

The results indicate that the different sized proteins might be different PE-PGRS proteins and also that a given PE-PGRS protein may vary in size between mycobacteria, and within the same species. Differences in signal strength were also apparent, suggesting that some PE-PGRS proteins may be more abundant.

Subcellular localization of PE-PGRS proteins in *Mycobacterium smegmatis* was also done. Cell wall, cell membrane and cytosol fractions were prepared from the *M. smegmatis* mc² 155. To localize the PE-PGRS proteins, Western blot analysis was performed using the antibody Rv 1441 against the different fractions. Several PGRS proteins were recognized in all fractions of the cell (Fig. 17). A number of proteins of different sizes were seen in both the cell wall and cell membrane fractions.

3.8 Analysis of the *M. smegmatis* proteome *in silico*

When antibodies raised against the PE-PGRS proteins of *M. tuberculosis* Rv1441c and Rv0980c were used to probe Western blots bearing polypeptides from *M. smegmatis* several different species were detected suggesting the existence of related proteins in this mycobacterium (Fig. 16, 17). This result is consistent with the findings of Southern blotting using PGRS-specific probes (140). To gain more insight into the nature of these possible PE-PGRS proteins a BLASTP search was performed using the sequences of Rv1441c and Rv0980c to query the *M. smegmatis* protein database derived from the near complete genome sequence of strain mc²155 (http://tigrblast.tigr.org/ufmg/index.cgi?database=m_smegmatis|seq). This revealed the presence of >40 potential PE-PGRS proteins with scores lower than 10⁻⁵ and the top five hits are shown (Fig. 18 A). In both cases the same five proteins were detected although the order of relatedness varied. Better scores were obtained with Rv1441c which showed strong similarity to the PGRS part of the 882 aa *M. smegmatis* protein ORF06022 and



Antibody 1441

Fig. 17. Subcellular localization of PE-PGRS proteins in *M. smegmatis*. Cell wall, cell membrane and cytosol proteins were prepared and separated by SDS-PAGE. Antibody 1441 was used to locate the PE-PGRS proteins in the cell. The antibody recognized several proteins in each cell fraction: WCL, whole cell lysate; CW, cell wall fraction; MEM, cytoplasmic membrane fraction; CYT, cytosol.

Fig. (18) A. Results of Blast search against *M. smegmatis* protein database with Rv1441c and Rv0980c

Query= Rv1441c (491 aa)

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability	
		P(N)	N
ORF06022 (882 aa)	761	4.7e-79	1
ORF10726 (1274 aa)	602	3.9e-59	1
ORF01766 (513 aa)	540	1.0e-53	1
ORF06023 (564 aa)	528	2.0e-52	1
ORF02413 (450 aa)	283	3.8e-24	1

Query= Rv0980c (457 aa)

ORF10726 (1274 aa)	409	1.7e-38	1
ORF01766 (513 aa)	373	5.2e-36	1
ORF06022 (882 aa)	358	2.4e-33	1
ORF06023 (564 aa)	279	8.3e-24	2
ORF02413 (450 aa)	231	3.4e-18	1

Fig. 18. (B) Best alignments between HSP of Rv1441c and ORF06022

Score = 761 (272.9 bits), Expect = 4.7e-79, Sum P(2) = 4.7e-79
Identities = 193/402 (48%), Positives = 218/402 (54%)

```

Query:   118 GRP-LIGNGADGLPGQNGGAGG--LLWNGNGGCGAGDA-AHP-NGGNGGDAGMF-GNGGA 171
          GR  LIG G  G   NGG GG  LL+GNGG GG+G A A+  NGG+GG  G+F G GG
Sbjct:   304 GRSGLIGVGGAGF---NGGTGGSTLLFGNGGAGGSGVAWANSNGNGHGHGGLFLGTGGD 360

Query:   172 GGAGYSPAAGT-GAAGGAGGAGGAGGWSL-----GNGGAGGNGGTGASGADGGG-GLPPV 224
          GGAG +  AG  G  GGAGG GG  G  +      GNGGAG NGG GA+  + G  G  +
Sbjct:   361 GGAGAAGVAGVNGDGGAGGRRGGNSGAFALFAQAGNGGAGANGGNGATTTESGASGAAGI 420

Query:   225 PASPGNGGGGDAGGAAGMFGTGGAGGTGGDGGAGGAGDSPNSGANGARGGDGGNGAAGG 284
          GGNGG G + G   +F T GAGG GG GG GG G   G +G   G   AGG
Sbjct:   421 AGGRGGNGGNGGSSVGLFAIFATACAGGAGGAGGNGGNG---GDGVHGTATTAGE--AGG 475

Query:   285 AGGRLFGNGGAGGNGGTAGQGGDGGTALGAGGIGDGGTGGAGGTGGTAGIGGSSAGAGG 344
          AGG  G+GGAGG GG  G GG G   +  G  GDGG GG GG  G+ G GG  AGA G
Sbjct:   476 AGGN-GGDGGAGGQGGDGGAGGRGPLIVLRAGRTGDGGDGGDGAAGSPGNGG--AGAHG 532

Query:   345 AGG--DGGAGGTGGGSSMIGGKGGTGGNGGVGGTGGASALTIGNGSSAGAGGAGGAGGTG 402
          G  DGGAGG GG S   G GGTGG+G VGG  G   +G S   AGG GG GG G
Sbjct:   533 DAGHLDGGAGGNGGNSGA-AGSGGTGGSGTVGGANG-----VGGASVTAAGGHGGRGGDG 586

Query:   403 GTGGYIESLDGKGQAGNGGNGGNAAGGAGG-GGTGAGGNGGAGGNGGDGGPSQG----- 456
          +      +DG G+  GNGG+GGN   GG GG GG  + G  GGAGG+GGDG
Sbjct:   587 HSTTDA-GIDG-GRGGNGGDGGLNGGDDGGHGGNASSGRGGAGGDGGDGTIDNAATTSD 644

Query:   457 --GGNPGFGG-----DGGTGGPGGVGVDPGIGGANGAQGKHG 491
          GG  G GG   G GG GG  V  G G A G  G +G
Sbjct:   645 VTGGRGGNGGHSTTNSAGDGGDGGAAVNAGTGSALGAGGNG 686

```

Score = 37 (18.1 bits), Expect = 4.7e-79, Sum P(2) = 4.7e-79
Identities = 13/43 (30%), Positives = 19/43 (44%)

```

Query:    11 LSAAAADVASIGAALSAANGAAAPTATAGVLAAGADEVSAAIAS 53
          L  A  V +GA  A   A  +A V +A  DE +++  S
Sbjct:    19 LGRHARYVGRVVGALAVLALGIGLALGSAAVASADTDETTSSDTS 61

```

some similarity to the N-terminal PE domain (Fig. 18 B). This suggests that the PGRS domain should harbour more cross-reacting antigenic determinants than the PE domain and this agrees well with the immunological findings of Delogu and Brennan with the PE-PGRS protein Rv1818c (58). Furthermore, cross-reactivity can be expected between the highly related PGRS domains of different proteins.

3.9 Description of strains from Bangladesh

A total of 48 cases of TB were investigated (Table 3). Samples were collected randomly from patients who came to the IDCH between 1999 and 2000. The IDCH is a national facility for the diagnosis and treatment of TB in Dhaka city. Treatment is free, and the hospital, therefore, has to serve a large segment of the city's population. It also handles a substantial number of patients with complications referred from other hospitals and Thana Health Complexes in and outside Dhaka city. Twenty six of the patients were from different parts of Dhaka city. Others were from different regions of Bangladesh outside Dhaka. Of them five are from Naranyanganj, one from Patuakhali, one from Natore, one from Rangpur, two from Gazipur, one from Narsingdi, three from Kishoreganj, three from Shariatpur, one from Noakhali, one from Habiganj, one from Munshiganj and two from Comilla. Most of them (>70%) were of low socio-economic condition. Thirty six of them were male and twelve were female. The age ranges from 16 to 60 (the mean age was 34). Twenty two patients were vaccinated by BCG during their childhood and 20 were not vaccinated. The HIV status of the patients was not known. All of the cases were sputum microscopy positive. Culture of all sputum samples were done and they were found to be all positive. Drug sensitivity testing was not done for all of the samples. Thirty three were previously diagnosed as TB and had been treated with anti-tubercular drugs for certain periods. Fifteen of the isolates were primary isolates.

3.10 Principles of deletion analysis

Comparative genome hybridization arrays using Bacterial Artificial Chromosome (BAC) clones have been able to identify 20 variable regions among the members of the tubercle

bacilli (29, 83). DNA microarrays also contribute to comparing genomes of closely related strains of mycobacteria. Among the 20 variable genomic regions resulting from insertion-deletion events in the genomes of the tubercle bacilli, 14 regions (RD1-14), ranging in size from 2 to 12.7 kb, were absent from BCG Pasteur relative to *M. tuberculosis* H37Rv (18, 83). At the same time, six regions, RvD1-5, and 'M. tuberculosis specific deletion 1' (TbD1), that were absent from the *M. tuberculosis* H37Rv genome relative to other members of the *M. tuberculosis* complex were revealed by comparative genomics approaches employing pulsed-field gel electrophoresis (PFGE) techniques (31, 83) and *in silico* comparisons of the complete *M. bovis* AF2122/97 genome sequence and the *M. tuberculosis* H37Rv sequence. Based on the above knowledge a deletion analysis system has been developed to evaluate the distribution of 20 variable regions in the strains of *M. tuberculosis* complex.

Deletion analysis involves PCR screening assays using oligonucleotide primers internal to the RDs and RvDs, as well as oligonucleotides flanking these regions (29)(Table 7). This approach has been found to be highly reliable and internally controlled since PCR amplicons obtained with the internal primer pair correlated with the absence of an appropriately sized amplicon with the flanking primer-pair, and *vice-versa* (Fig. 19). *M. tuberculosis* strains are highly conserved with respect to RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD13, and RD14. These RDs, together with TbD1, can differentiate *M. tuberculosis* strains, independent of their geographical origin and their typing characteristics, from certain other members of the *M. tuberculosis* complex. The presence or absence of RvD regions in *M. tuberculosis* strains was variable and RvD2 region shows greatest variability.

3.11 Results obtained by deletion analysis

Forty eight strains isolated from pulmonary tuberculosis patients were investigated by PCR-based deletion analysis using the 20 variable regions. Most of the RD regions were

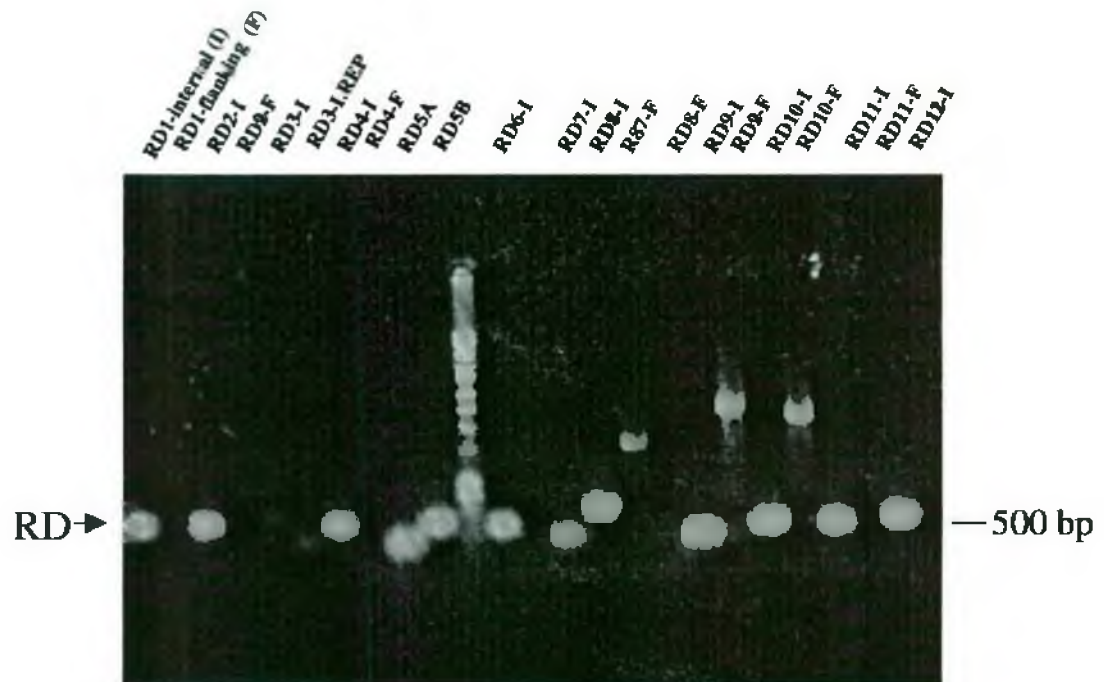


Fig. 19. Amplicons obtained from strain TB6 indicating presence or absence of genomic regions (RD1-12). Primers internal to RDs as well as flanking these regions were used in the PCR. PCR amplicons obtained with the internal primerpair correlated with the absence of an appropriate sized amplicon with the flanking primer pair.

present in all the strains tested. RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD13, RD14 and RvD1 were present in all strains (Table 7). The presence of these regions indicates that all the strains tested were *M. tuberculosis*. This result also confirms the previous observation that the above regions are highly conserved in *M. tuberculosis* strains and they might be involved in the pathogenesis of human tuberculosis (29). In 29 strains, the RD3 region was absent and in two strains the amplified product using the second set of internal primers was of smaller size. RD11 was also absent in eight of 48 strains tested. RD3 and RD11 regions correspond to the two prophages phiRv1 and phiRv2 of *M. tuberculosis* H37Rv (46). RD5, which is flanked by a copy of insertion sequence IS6110, was absent from some strains and RD6, containing IS1532 (83), was absent from 10 strains. These four regions were found to be absent from some of the *M. tuberculosis* strains investigated earlier (29).

A large number of the strains (35) tested had deletion of the TbD1 region indicating that they are all “modern” strains of *M. tuberculosis*. But 13 out of 48 strains were found to harbour the TbD1 region (Table 7). As was expected, the presence or absence of RvD regions were variable. RvD1 was present in all 48 strains. RvD2 and RvD3 were highly variable. RvD2 was absent from 27 strains while RvD3 was absent from 24 strains. Both RvD2 and RvD3 were deleted from 21 strains. Strains belonging to the Beijing cluster were shown to lack both the regions RvD2 and RvD3 (29). RvD4 was present in 45 strains and was absent from three strains. RvD5 was absent from one of the strains tested.

3.12 Results obtained by spoligotyping

All 48 isolates were analysed by spoligotyping (Table 7). A cluster was defined as two or more isolates from different patients with identical spoligotype patterns, whereas nonclustered patterns were referred to as unique. Thirty three (68%) strains could be grouped into 8 different clusters (Fig. 20). The largest cluster (cluster 1) consists of 15 strains, one cluster contains five strains (cluster 2), one consists of three (cluster 3) and there are five other clusters (clusters 4, 5, 6, 7 & 8) comprising two strains each. Fifteen (31%) isolates exhibited unique (nonclustered) patterns. All strains of the largest cluster

comprising 15 isolates contained only nine of the 43 spacer sequences tested. They showed a spoligotype pattern with hybridization only to the 3' –terminal spacer 35 to 43, that is characteristic of the Beijing genotype. In all strains identified as Beijing type, RD3, RvD2 and RvD3 regions were found to be absent by deletion analysis. The Beijing genotype specific IS6110 in *dnaA-dnaN* region was amplified in all 48 strains. This region was only present in the largest cluster strains identified as Beijing members by spoligotyping and deletion analysis (Table 7).

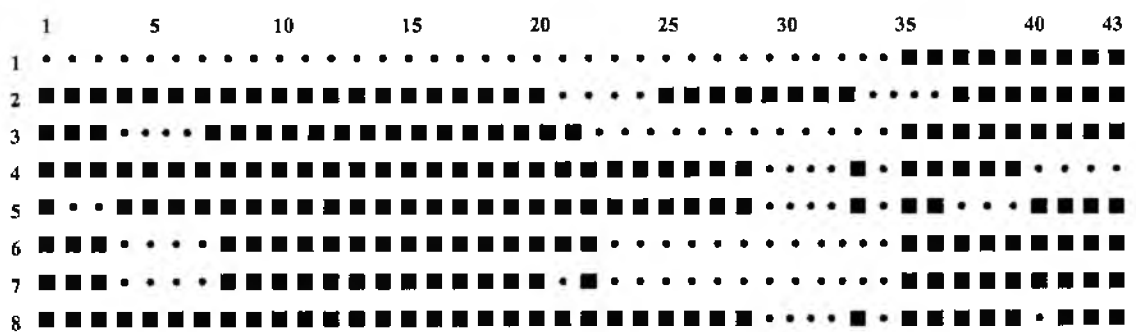


Fig. 20. Spoligotype patterns of clustered isolates. The filled boxes represent the presence of spacers, and the periods represent the absence of spacers. The numbers along the left of the figure are cluster numbers.

The 48 isolates analysed by deletion analysis and spoligotyping were organized in four blocks according to their degree of relatedness (Table 7). Block 1 comprised the 13 “ancestral” strains having TbD1 present. *M. bovis* strains usually lack spacer sequences 39-43. Two of the clustered strains (MA25 and MA31) in block 1 lacking spacers 40 to 43 looked like *M. bovis*, but the presence of RD9 confirmed that they are *M. tuberculosis*. According to the findings showed by Brosch *et al.* (169) these are suspected to belong to the phylogenetic group 1 defined by Sreevatsan *et al.* (169) based on the *katG* codon 463 and *gyrA* codon 95 sequence polymorphisms. Block 2 contains 9 TbD1 deleted strains having spacers 35-43 present indicating that they all belong to group 1. All 15 Beijing strains are also in group 1 characterized by their unique spoligotype. And finally block 4 contains the strains lacking spacers 33-36 and having the flanking spacers present. These are the typical characteristics for strains of phylogenetic groups 2 and 3 of Sreevatsan. Strains in blocks 2, 3 and 4 are all modern strains.

4. Discussion

A) The availability of complete genome sequence of H37Rv, the most widely used strain of *M. tuberculosis*, has provided us with a wealth of new information about the biochemistry, general metabolism, genetics, immunology and physiology of this much-feared bacterium (46). However, a great deal remains to be learned about the biological roles of the proteins comprising the PE and PPE protein families, two novel multigene families. It has been shown here that all ten PE-PGRS genes investigated are transcribed in cultured *M. tuberculosis* H37Rv cells. It indicates that probably all or most of the members of the PE-PGRS class are being expressed at the transcriptional level. Out of the ten selected PE-PGRS genes, five were taken for further investigations. Five antibodies, raised by genetic immunization, recognized single proteins when the corresponding genes were expressed in eukaryotic cells and hybridized with the antibodies produced. All five PE-PGRS genes were well expressed in HeLa cells than in 293 eukaryotic cells. In all five cases, proteins of the appropriate size were detected along with several other protein species when cell-free extracts of *M. tuberculosis* proteins were tested against the corresponding antibodies. This cross-reaction probably indicates that members of the PE-PGRS family have common epitopes, consistent with their extensive sequence conservation and predicted numerical abundance. Dot-blot experiments showing cross-reactivity among the five antibodies tested confirm the above results. In the dot-blot experiments, 2 proteins reacted with all five antibodies suggesting that the number of common epitopes are more in some PE-PGRS proteins.

There is evidence that suggests that the presence of PE-PGRS genes are restricted to members of the *M. tuberculosis* complex and a few other mycobacterial species which cause diseases in humans and other mammals (44, 46, 140, 152). This suggests that the PE family might have a specific role in mycobacterial infection and survival within host tissues or the numerous PE-PGRS genes could be a source of antigenic variability for *M. tuberculosis* in order to evade the host immune response.

After the discovery as a novel multigene family, it is of great interest to know the location of the PE-PGRS proteins in the mycobacterial cell. The first step in bacterial infection of the host is the manifestation of specific surface components by the pathogen to facilitate its entry of host cells and tissues. In a recent study, there has been evidence that certain PE-PGRS proteins including Rv1818 might be located at the cell surface and are involved in the pathogenesis of tuberculosis infection (26). One of the members of PPE gene family, Rv1917c, which is highly polymorphic in clinical isolates of *M. tuberculosis* was expressed in *M. smegmatis* and *M. bovis* BCG and the expression product was found to be hydrophobic and surface exposed (159). In the present study, experiments were carried out to determine the subcellular location of some of the members of the PE-PGRS family and the findings of the work provided important clues regarding the possible role of PE-PGRS proteins in antigenic variation in *M. tuberculosis*. On examination of various cellular compartments, distinct proteins were detected in different sites with many of them occurring in the cell wall and cell membrane fractions of *M. tuberculosis*, *M. smegmatis* and, presumably, the other PGRS-reactive mycobacteria. At least some of the PE-PGRS proteins are exposed on the cell surface as revealed by immunogold labelling. On examination of the primary sequences, hydrophobic stretches capable of forming transmembrane α -helices were only predicted in two PE-PGRS proteins, Rv0151c and Rv1430, and none of the proteins was found to contain an N-terminal signal peptide typical of Gram positive bacteria. This finding suggests that PE-PGRS genes are transported to membranes using a system other than the general secretory pathway and raises questions as to how these polypeptides become associated with the cell envelope.

Protein extracts from the clinical isolates were examined and there were variations in the protein profiles suggesting that at least some PE-PGRS genes undergo sequence variation. These experimental results confirm the predictions of variability from earlier allelic comparisons between *M. tuberculosis* H37Rv and *M. bovis* BCG (46), and the extensive polymorphism in hybridization patterns revealed with PGRS probes (140, 154, 187). More recently, comparisons of the genome sequences of *M. tuberculosis* strains H37Rv and CDC1551 have shown that apart from some ten insertion-deletion events, the

major source of genetic diversity occurs in the PE and PPE families (19). Pairwise comparisons of the respective PE and PE-PGRS gene sequences in these two strains were undertaken. These revealed that all 37 PE genes were identical in both strains. By contrast, 39 of the 62 common PE-PGRS genes showed extensive divergence that would result in the absence of the protein or a difference in size as a result of frameshift mutations, or in-frame insertion and deletion events, respectively. This is in good agreement with the findings of our Western blot analysis using the most cross-reactive antibodies (Figs 13, 14, 16 & 17) as these uncovered two to three differences in the size or presence among the five to ten PE-PGRS proteins detected. The loss of immunoreactivity could result either from inactivation of the corresponding gene or from amino acid sequence changes that alter B-cell epitopes.

Genetic variation results from spontaneous changes in the genes of a species and, in pathogens, can lead to phenotypic alterations which include phase and antigenic variation (150). This remarkable phenomenon, which has been intensively studied in a small number of model micro-organisms, is probably widely distributed in the prokaryotic kingdom. Genetic variation has been well documented in a number of pathogenic bacteria such as *Neisseria spp*, *Haemophilus spp*, *Streptococcus pneumoniae*, *Borrelia hermsii*, and *B. burgdorferi* (33, 35, 79, 116, 123, 170, 200). The principle mechanisms involved are recombinational events and slipped-strand mispairing during replication, both associated with repetitive sequences (150). Homologous recombination between intragenic tandem repeats, leading to deletion or duplication of runs of amino acids, is a frequent source of size and antigenic variation in the M protein present on the Streptococcal cell surface and results in evasion of the immune system (95). Variation of *opa* gene expression occurs in *Neisseria spp*. by slipped strand mispairing (124) involving CTCTT repeat units present, and leads to the *opa* gene being shifted in or out of frame.

The PE-PGRS element is the most abundant repetitive sequence in the tuberculosis complex. A characteristic feature of a PE-PGRS gene is its GC-richness, stemming from a 9 bp repeat CGGCGGCAA arranged in tandem copies (141). It is tempting to speculate

that similar recombinational events involving these sequences, or strand slippage during replication, might be responsible for PE-PGRS variation in *M. tuberculosis*. The data presented here show that the PE-PGRS proteins vary in size in different mycobacteria and also in different clinical isolates of *M. tuberculosis* (Figs. 5, 6). Comparative genomics of *M. tuberculosis* strains H37Rv and CDC1551 underlines the role of insertion/deletion events in generating diversity in the PE-PGRS family (Fig. 1) (19, 46). These data strongly suggest that some PE-PGRS proteins may correspond to variable surface antigens.

In order to escape the host's immune system, bacteria mutate extensively to generate rare variants. The role of DNA polymerases IV and V, also known as mutases, in this process has been recognised recently in microorganisms and the targets are the 'contingency genes' often encoding the surface antigens recognized by the immune system (131, 145). The corresponding genes contain simple, repeated sequences, which are preferential sites for DNA polymerase IV-mediated slippage mutations resulting in genetic diversity. *M. tuberculosis* contains genes for both Pol IV and Pol V (Rv1537, Rv3056) and should thus produce the enzymes that might generate variability in the PE-PGRS family.

Functional information remains scarce for the PE-PGRS proteins. On the basis of sequence similarity to EBNA, the Gly- and Ala-rich Epstein Barr nuclear antigen, it has been proposed that PE-PGRS proteins could act as proteasome inhibitors and block antigen presentation by the MHC I pathway (46, 114, 115). Our finding that some PE-PGRS proteins are located in the cell envelope is of relevance in this context as the proteins would readily be shed into the lumen of the phagosome. At least two PE-PGRS proteins correspond to antigens expressed during infection as antibody responses to Rv1759c and Rv3367 have been reported in humans and rabbits, respectively (4, 69, 165). In the case of Rv1759c, not all sera from tuberculosis patients contain antibodies that recognise the protein suggesting that the Rv1759c protein may be absent or altered in some strains. Espitia et al. have also shown that one PE-PGRS protein is capable of binding fibronectin and could thus mediate bacterial attachment to host cells (69) but not all PE-PGRS proteins tested have this activity (165). In addition, members of the PE-

PGRS families have been implicated in the pathogenesis of *M. marinum* (146), where at least two genes were shown to be upregulated strongly following phagocytosis of the bacterium.

The immunogenicity of the PE-PGRS protein, Rv1818c, was studied in great detail in mice and the findings reported in a very recent publication (58). Immunisation with the PE domain induced Th1-type responses that were not found when the complete PE-PGRS protein was used. Rather, an antibody response was induced with strong recognition of the PGRS part of the protein. During infection with *M. tuberculosis*, a humoral immune response to the PGRS domain of Rv1818c was also observed.

There are many similarities between the PE-PGRS proteins and the silk proteins produced by various insects. In silk, a structural protein that forms fibres, glycine, alanine and serine account for >40%, >20% and >10% of the amino acid residues, respectively (93). The predicted glycine and alanine content of the PE-PGRS proteins is very similar although asparagine (>5%) is much more abundant than in silk. Both families of proteins are modular in organisation, and rich in Gly-Gly-X repeats. These intriguing parallels raise the possibility that the role of the PE-PGRS proteins could be purely structural. If this were the case, the insertion and deletion of repeat modules would be readily accommodated unlike the situation in an enzyme where loss of activity would probably occur.

In conclusion, this study results suggests that some of the PE-PGRS genes encode components of the cell envelope, that can undergo extensive variation. Given, the remarkable level of sequence conservation in the remainder of the proteome (19, 169), and the consequent lack of amino acid diversity, it is conceivable that the major source of antigenic drift in tubercle bacilli is associated with the PE-PGRS proteins. Given the obvious ramifications for tuberculosis vaccine development it is imperative that detailed immunological studies of these proteins be performed and this could lead to novel interventions for controlling and treating tuberculosis.

B) The aim of the study was to characterize the strains causing tuberculosis in Bangladesh. A total of 48 strains were studied which were clinical isolates from adult pulmonary tuberculosis patients treated in a hospital in Dhaka during different time periods. Deletion analysis using RD1 to RD14 revealed that all the samples investigated were *M. tuberculosis*. The presence of RD9 region confirmed that the strains were not *M. bovis*. Bovine tuberculosis has been eradicated or almost under control in most of the developed countries. However, very little or no information is available from developing countries including Bangladesh. There has been a close association between *M. bovis* infection in workers in the meat and livestock industries and the infection in cattle (53, 54). Another potential source of *M. bovis* infection is unpasteurized dairy products. The information about the incidence of bovine tuberculosis in cattle in Bangladesh is scarce. Although people from urban area can take pasteurized milk, many people from the rural area of Bangladesh still drink unpasteurized milk from a herd with bovine tuberculosis. The clinical isolates tested in this study were mainly from the patients of urban area who are unlikely to have contact with cattle having bovine tuberculosis and less chance of taking unpasteurized milk. All isolates tested in this study were from pulmonary tuberculosis patients. The frequency of pulmonary and extrapulmonary TB is related to potential exposure and route of transmission of *M. bovis* from animals (191). *M. bovis* infection in humans is more commonly associated with extra-pulmonary cases including lymph gland and gastrointestinal TB (104). Therefore, in order to determine the extent of disease burden caused by *M. bovis*, it is important to test the strains recovered from extra-pulmonary samples. Although largely eradicated in the developed nations, epidemiological data on the impact of bovine tuberculosis on human health is scanty, but in the light of the increasing incidence of tuberculosis worldwide, it is urgently needed.

An interesting finding of the study was that 13 of the 48 strains tested had “*M. tuberculosis* specific deletion 1” (TbD1) region intact indicating that there are a considerable number of ‘ancestral’ strains taking part in the transmission of tuberculosis in Bangladesh (29). The ancestral strains are predominantly originated from endemic foci (144, 189) and this TbD1 region was found to be deleted in some of the strains of Indian origin (29). Bangladesh has been an endemic region for tuberculosis from long before

and our results suggest that tuberculosis was prevalent in this region before TbD1 deletion occurred, that is, long before the worldwide spread of modern *M. tuberculosis* strains causing tuberculosis epidemic. Unfortunately, it is impossible to calculate the time that has elapsed since the divergence from the putative common ancestor. Absence of TbD1 region in 35 strains supports the fact that 'modern' *M. tuberculosis* strains (also called epidemic strains) have also been introduced in Bangladesh contributing to the worldwide spread of tuberculosis epidemic.

The "modern" strains comprise representatives of major epidemics like the Beijing, Haarlem, and African *M. tuberculosis* clusters. The majority *M. tuberculosis* strains of the Beijing family originated from the province of Beijing in China and strains of this family were found to dominate in neighboring countries such as Mongolia, South Korea, Thailand and Vietnam (7, 190). In contrast to the dominance of the Beijing genotype in many Asian countries, a low frequency (3%) of this genotype was reported among the strains from India (126). Strains of the Beijing family have also been found in distant countries in Europe, Africa and the USA. The "W" strain, which caused a large outbreak of multidrug-resistant tuberculosis in New York City and other U.S. cities, belongs to the Beijing family (110). The strains of the Beijing family pertain to a closely related group shown by IS6110 fingerprints of *Pvu*II-restricted chromosomal DNAs which is distinct from other *M. tuberculosis* strains (110). Strains grouped in the Beijing family were found to be identical by spoligotyping, which is based on the DNA polymorphism in the DR region. Strains of the Beijing family contained only 9 of the 43 spacer sequences. These spacer sequences correspond to nine contiguous spacers located near the 3' end of the DR region of strain H37Rv. RvD2 and RvD3 regions were found to be deleted in the strains belonging to the Beijing cluster. Twenty-two out of 48 Bangladeshi strains tested had deleted RvD2 and RvD3 regions and were suspected to be members of the Beijing family. In order to confirm that these strains are indeed related to the Beijing family of strains we amplified the IS6110 between the *dnaA-dnaN* region. Fifteen of the strains lacking RvD2 and RvD3 region had the Beijing strain specific IS6110 region present indicating that they are members of Beijing family. Spoligotyping of these strains confirmed the results. A total of 33 (68%) strains were grouped into 8 clusters (Fig. 20),

the largest of which contains 15 strains. Though the high prevalence of Beijing family strains has been reported in many Southeast Asian countries, there is virtually no information on the strains from Bangladesh. Our investigations using deletion analysis and spoligotyping among the isolates from TB patients in Bangladesh led to the identification of a substantial number of strains belonging to the Beijing family. In the neighboring countries in Asia rates of infection with the Beijing family strains are higher than those in the more distant countries, suggesting that the Beijing family may have radiated from the Beijing area to other regions. The factor which might be responsible for the selection and dissemination of the Beijing strains is not known but there has been evidence that the prevalence of these strains are more common in the areas where BCG vaccination coverage is high. Most of the countries in Southeast Asia are covered by BCG vaccination which has been used for the past two to six decades. The hypothesis is supported by the study results where more DNA polymorphisms were found in a high incidence country where BCG vaccination is not a common practice than in a place of low incidence where BCG vaccination coverage is very low suggesting that vaccination may have favored the selection of *M. tuberculosis* strains that resist BCG-induced immunity (87). Our study results indicates that the prevalence of the Beijing type *M. tuberculosis* strains is more common in Bangladesh where BCG vaccination is an essential component of Expanded Program of Immunization (EPI). The samples for this study were collected from the patients coming to a hospital from different parts of Dhaka city and also from areas outside Dhaka. Further studies are needed with a larger sample size to investigate the incidence of Beijing type strains in Bangladesh and to know the transmission dynamics of tuberculosis caused by these virulent strains in the community. It is important because of its association with drug resistance.

5. Concluding Remarks

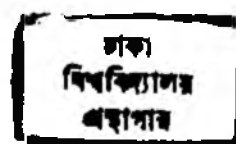
A) Complete genome sequencing of the much feared bacterium, *Mycobacterium tuberculosis* has revealed two novel multigene families occupying 8% of the whole genome. PE-PGRS genes, members of one of these families, are of great interest. Initially 10 of the PE-PGRS genes were selected for investigations and five of them were examined in detail. The results obtained show that:

- All ten genes were found to be expressed *in vitro* by RT-PCR indicating that all or most of the PE-PGRS genes are expressed at the transcription level.
- Antibodies against five PE-PGRS proteins were produced using DNA vaccination. These antibodies reacted with the corresponding proteins expressed in epithelial cells or in reticulocytes, confirming that the PE-PGRS proteins are antigenic.
- Each antibody identified several proteins of various sizes in protein extracts of five different mycobacteria (*M. tuberculosis*, *M. bovis* BCG, *M. smegmatis*, *M. marinum* and *M. goodii*) and 11 clinical isolates of *M. tuberculosis* indicating that the same PE-PGRS protein varies in size in different mycobacteria and also in clinical isolates of *M. tuberculosis*. The results also suggest that different PE-PGRS proteins share common epitopes.
- PE-PGRS proteins were localized in both the cell wall and cell membrane by subcellular fractionation studies and immunoelectron microscopy.

From these results, it can be concluded that some PE-PGRS proteins may be surface antigens and could generate antigenic diversity within the repetitive sequence of the PGRS domain. Further studies are needed to investigate the exact role of PE-PGRS proteins in the host immune response. The antigenic and immunological characterization of the PE-PGRS proteins needs to be investigated with the emphasis on clinical isolates from Bangladesh.

B) A total of 48 clinical samples were collected from the adult pulmonary tuberculosis patients attending a hospital in Dhaka city and investigated appropriately. Patients were mainly from different parts of the Dhaka city as well as from other parts of Bangladesh. All of the strains tested were *M. tuberculosis* strains. There were no *M. bovis* strain identified indicating that bovine TB in humans may not be a problem in Bangladesh. However, since *M. bovis* infection in humans is more commonly associated with extrapulmonary tuberculosis, it is important to test the strains from such cases in order to come to a firm conclusion. Deletion analysis revealed that both 'ancestral' and 'modern' *M. tuberculosis* strains cause tuberculosis in Bangladesh. A total of 33 (69%) strains were grouped into 8 different clusters by spoligotyping. Unique (nonclustered) spoligotype patterns were seen in 15 (31%) strains. The largest cluster comprised 15 isolates and the spoligotype pattern of these isolates is characteristic of the Beijing genotype. The Beijing type strains have originated from China and are found to be dominant in many Asian countries. Most of them were associated with drug resistance. Our results indicate that the 'Beijing' strains are contributing to the transmission of drug-resistant tuberculosis in Bangladesh. Further investigations are needed to determine the incidence of tuberculosis caused by 'Beijing' strains using clinical samples from a larger population of patients in Bangladesh.

400598



Acknowledgements

This dissertation is the result of a doctoral study conducted through a collaborative program between the University of Dhaka, the Institut Pasteur in Paris, and the ICDDR,B: Centre for Health and Population Research. A major part of my experimental work was done at the Institut Pasteur, which was supported by the Embassy of France in Dhaka. I would like to express my sincere gratitude to people belonging to the above mentioned institutions who have helped and supported me during these years.

In particular, I would like to express my heartiest gratitude to my supervisor Dr. S. I. Khan, Professor, Department of Microbiology, University of Dhaka for generously helping me in the University and for valuable advice and encouragement during the study. It would not have been possible for me to complete my study without his support.

Special thanks and gratitude are due to my supervisor at ICDDR,B, Dr. Firdausi Qadri, Senior Scientist, Laboratory Sciences Division, ICDDR,B who introduced me to this program. Her contribution to what ever I have achieved so far is unfathomable. She has always supported and encouraged me from the very moment I entered the program till today. I experienced many problems during my study - academic and otherwise, and I always found her ready with a helping hand.

Prof. Stewart Cole, Head of the Mycobacteriology Unit at the Institut Pasteur, was my supervisor and mentor at Pasteur who first introduced me to the exciting field of mycobacteriology. It was a great opportunity for me to work directly under him. During my work at the Institut Pasteur, he understood my problems readily and tried his best to solve them. He was a constant source of inspiration and encouragement. I owe him lot of thanks for giving me time in teaching computer programs and for frequent discussion regarding the experiments despite his extremely busy schedule.

I am indebted to Professor David A Sack, Director, ICDDR,B for giving me the opportunity to work at ICDDR,B and also for his support and encouragement.

My sincere thanks are due to Dr. Balakrish Nair, Associate Director and Head, Laboratory Sciences Division, who always encouraged and helped me in my efforts. His pleasant personality and friendly attitude helped me overcome the stress and worries associated with graduate studies.

I am thankful to the scientists of Laboratory Sciences Division, ICDDR,B, especially the members of the 'tea club' who accepted me in their midst were extremely helpful in cheering me up.

It is difficult to name everyone but I acknowledge the help of the staff, friends and colleagues at ICDDR,B during my work.

All my friends and colleagues at Pasteur who always strived to make the atmosphere in the laboratory lively and comfortable deserve special mention. My sincere gratitude and thanks to my friends Pillar, Alex, Roland and Steve for their help and for making my life in Paris enjoyable. I remain grateful to Nadine, Brigitte and many others for their guidance and help, be it research related or personal. I must thank Anne-Marie, the secretary of our unit at Pasteur, for her generous and never failing help and support.

My heartiest thanks and gratitude to my husband, Tahmeed, for his support, inspiration and above all for all his sacrifice without which this study would not have been possible. I remember my dearest daughter, Arnaaz who has been a source of joy and pleasure during my work.

My sincere gratitude to my parents for their never ending support, encouragement and sacrifice. I also recall the affection and encouragement of the other members of my family. My special thanks to my youngest sister, Taposhi, for taking care of my daughter during this stressful period.

My study in Paris was made possible by a fellowship from the French Government. I must thank France Lasnier, Gilles Fischer, Andre Raynouard, Caroline Ladous and Ms. Maria of the French Embassy in Dhaka. This work was supported by the grants from the European Union (QLRT-2000-02081), the Institut Pasteur, the Association Francaise Raoul Follereau and ICDDR,B: Centre for Health and Population Research, Bangladesh, UNOCAL, Bangladesh and Aventis, Bangladesh.

6. References

1. National Guidelines for Tuberculosis Control. 1997. Second ed. Tuberculosis & Leprosy Control Services, Dhaka.
2. World Health Organization Global Tuberculosis Programme. 1997. WHO Report on the Tuberculosis Epidemic. WHO.
3. World Health Organization. The world health report. 1998. WHO.
4. Abou-Zeid, C., T. Garbe, R. Lathigra, H. G. Wiker, M. Harboe, G. A. Rook, and D. B. Young. 1991. Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect Immun* 59:2712-8.
5. Ali, M. S. 1999. Studies on Primary Drug Resistance Pattern of *Mycobacteria*. BSM Medical University, Dhaka.
6. Amara, R. R., S. Shanti, and V. Satchidanandam. 1998. Characterization of novel immunodominant antigens of *Mycobacterium tuberculosis*. *Microbiology* 144 (Pt 5):1197-203.
7. Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen. 2000. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg Infect Dis* 6:302-5.
8. Arruda, S., 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* 261:1454-7.
9. Bailey, T. L., and M. Gribskov. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14:48-54.
10. Baily, T. L., and Elkan, C. 1994. Presented at the Second International Conference on Intelligent Systems for Molecular Biology.
11. Banatvala, N., and G. G. Peremitin. 1999. Tuberculosis, Russia, and the Holy Grail. *Lancet* 353:999-1000.
12. Barbour, A. G., S. L. Tessier, and H. G. Stoenner. 1982. Variable major proteins of *Borrellia hermsii*. *J Exp Med* 156:1312-24.

13. Barnes, P. F., H. el-Hajj, S. Preston-Martin, M. D. Cave, B. E. Jones, M. Ota, J. Pogoda, and K. D. Eisenach. 1996. Transmission of tuberculosis among the urban homeless. *Jama* 275:305-7.
14. Barnwell, J. W., R. J. Howard, H. G. Coon, and L. H. Miller. 1983. Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned *Plasmodium knowlesi* malaria. *Infect Immun* 40:985-94.
15. Baruch, D. I., B. L. Pasloske, H. B. Singh, X. Bi, X. C. Ma, M. Feldman, T. F. Taraschi, and R. J. Howard. 1995. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* 82:77-87.
16. Bauer, J., A. B. Andersen, K. Kremer, and H. Miorner. 1999. Usefulness of spoligotyping To discriminate IS6110 low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. *J Clin Microbiol* 37:2602-6.
17. Bayer, R., and D. Wilkinson. 1995. Directly observed therapy for tuberculosis: history of an idea. *Lancet North Am Ed* 345:1545-8.
18. Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284:1520-3.
19. Betts, J. C., P. Dodson, S. Quan, A. P. Lewis, P. J. Thomas, K. Duncan, and R. A. McAdam. 2000. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. *Microbiology* 146 Pt 12:3205-16.
20. Beyers, N., R. P. Gie, H. L. Zietsman, M. Kunneke, J. Hauman, M. Tatley, and P. R. Donald. 1996. The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. *S Afr Med J* 86:40-1, 44.
21. Bigi, F., C. Espitia, A. Alito, M. Zumarraga, M. I. Romano, S. Cravero, and A. Cataldi. 1997. A novel 27 kDa lipoprotein antigen from *Mycobacterium bovis*. *Microbiology* 143 (Pt 11):3599-605.
22. Bishai, W. R., N. M. Graham, S. Harrington, D. S. Pope, N. Hooper, J. Astemborski, L. Sheely, D. Vlahov, G. E. Glass, and R. E. Chaisson. 1998.

- Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *Jama* 280:1679-84.
23. Bonato, V. L., V. M. Lima, R. E. Tascon, D. B. Lowrie, and C. L. Silva. 1998. Identification and characterization of protective T cells in hsp65 DNA-vaccinated and *Mycobacterium tuberculosis*-infected mice. *Infect Immun* 66:169-75.
 24. Bonora, S., M. C. Gutierrez, G. Di Perri, F. Brunello, B. Allegranzi, M. Ligozzi, R. Fontana, E. Concia, and V. Vincent. 1999. Comparative evaluation of ligation-mediated PCR and spoligotyping as screening methods for genotyping of *Mycobacterium tuberculosis* strains. *J Clin Microbiol* 37:3118-23.
 25. Brandt, L., J. Feino Cunha, A. Weinreich Olsen, B. Chilima, P. Hirsch, R. Appelberg, and P. Andersen. 2002. Failure of the *Mycobacterium bovis* BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. *Infect Immun* 70:672-8.
 26. Brennan, M. J., G. Delogu, Y. Chen, S. Bardarov, J. Kriakov, M. Alavi, and W. R. Jacobs, Jr. 2001. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* 69:7326-33.
 27. Brosch, R., S. V. Gordon, A. Billault, T. Garnier, K. Eiglmeier, C. Soravito, B. G. Barrell, and S. T. Cole. 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Immun* 66:2221-9.
 28. Brosch, R., S. V. Gordon, K. Eiglmeier, T. Garnier, and S. T. Cole. 2000. Comparative genomics of the leprosy and tubercle bacilli. *Res Microbiol* 151:135-42.
 29. Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99:3684-9.

30. Brosch, R., Gordon, S.V., Eiglmeier, K., Garnier, T., Tekaia, F., Yeramian, E. and Cole, S.T. 2000. Molecular genetics of mycobacteria. ASM Press, Washington, D.C.
31. Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon. 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect Immun* 67:5768-74.
32. Brosch, R., A. S. Pym, S. V. Gordon, and S. T. Cole. 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol* 9:452-8.
33. Brunham, R. C., F. A. Plummer, and R. S. Stephens. 1993. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun* 61:2273-6.
34. Chaisson, R. E., J. S. Coberly, and K. M. De Cock. 1999. DOTS and drug resistance: a silver lining to a darkening cloud. *Int J Tuberc Lung Dis* 3:1-3.
35. Chaussee, M. S., J. Wilson, and S. A. Hill. 1999. Characterization of the *recD* gene of *Neisseria gonorrhoeae* MS11 and the effect of *recD* inactivation on pilin variation and DNA transformation. *Microbiology* 145 (Pt 2):389-400.
36. Chaves, F., Z. Yang, H. el Hajj, M. Alonso, W. J. Burman, K. D. Eisenach, F. Dronda, J. H. Bates, and M. D. Cave. 1996. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. *J Clin Microbiol* 34:1118-23.
37. Chevrel-Dellagi, D., A. Abderrahman, R. Haltiti, H. Koubaji, B. Gicquel, and K. Dellagi. 1993. Large-scale DNA fingerprinting of *Mycobacterium tuberculosis* strains as a tool for epidemiological studies of tuberculosis. *J Clin Microbiol* 31:2446-50.
38. Chintu, C., and A. Mwinga. 1999. An African perspective on the threat of tuberculosis and HIV/AIDS--can despair be turned to hope? *Lancet* 353:997.
39. Chowdhury AMR, C. S. A. Islam . In: Programme and abstract, 7th ASCON. 1998. 55.

40. Cirillo, J. D., C. K. Stover, B. R. Bloom, W. R. Jacobs, Jr., and R. G. Barletta. 1995. Bacterial vaccine vectors and bacillus Calmette-Guerin. *Clin Infect Dis* 20:1001-9.
41. Cock, K. M. D. 1994. Impact of Interaction with HIV. *In* J. D. H. P. a. K. P. W. J. McAdam (ed.), *Tuberculosis: Back to the Future*. John Wiley & Sons Ltd.
42. Cohn, D. L., F. Bustreo, and M. C. Raviglione. 1997. Drug-resistant tuberculosis: review of the worldwide situation and the WHO/TUATLD Global Surveillance Project. *International Union Against Tuberculosis and Lung Disease. Clin Infect Dis* 24 Suppl 1:S121-30.
43. 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. Meta-analysis of the published literature. *Jama* 271:698-702.
44. Cole, S. T. 1999. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett* 452:7-10.
45. Cole, S. T. 1996. Rifamycin resistance in mycobacteria. *Res Microbiol* 147:48-52.
46. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead & B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537-44.
47. Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares,

- S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. *Nature* 409:1007-11.
48. Cole, S. T. a. B., B.G. 1998. Presented at the Genetics and tuberculosis (Novartis Foundation Symposium 217).
49. Collins, C. H., M. D. Yates, and J. M. Grange. 1982. Subdivision of *Mycobacterium tuberculosis* into five variants for epidemiological purposes: methods and nomenclature. *J Hyg (Lond)* 89:235-42.
50. Contamin, H., T. Fandeur, C. Rogier, S. Bonnefoy, L. Konate, J. F. Trape, and O. Mercereau-Puijalon. 1996. Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical malaria episodes in Senegalese children. *Am J Trop Med Hyg* 54:632-43.
51. Council, M. 1955. Various combinations of isoniazid with streptomycin or with PAS in the treatment of pulmonary tuberculosis. *Br Med J* 1:435-445.
52. Cousins, D., S. Williams, E. Liebana, A. Aranaz, A. Bunschoten, J. Van Embden, and T. Ellis. 1998. Evaluation of four DNA typing techniques in epidemiological investigations of bovine tuberculosis. *J Clin Microbiol* 36:168-78.
53. Cousins, D. V. 2001. *Mycobacterium bovis* infection and control in domestic livestock. *Rev Sci Tech* 20:71-85.
54. Cousins, D. V., and D. J. Dawson. 1999. Tuberculosis due to *Mycobacterium bovis* in the Australian population: cases recorded during 1970-1994. *Int J Tuberc Lung Dis* 3:715-21.
55. Cross, G. A. 1996. Antigenic variation in trypanosomes: secrets surface slowly. *Bioessays* 18:283-91.
56. Davis, H. L., M. L. Michel, and R. G. Whalen. 1993. DNA-based immunization induces continuous secretion of hepatitis B surface antigen and high levels of circulating antibody. *Hum Mol Genet* 2:1847-51.
57. Deitsch, K. W., E. R. Moxon, and T. E. Wellems. 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* 61:281-93.
58. Delogu, G., and M. J. Brennan. 2001. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun* 69:5606-11.

59. Diaz, R., K. Kremer, P. E. de Haas, R. I. Gomez, A. Marrero, J. A. Valdivia, J. D. van Embden, and D. van Soolingen. 1998. Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994-June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. *Int J Tuberc Lung Dis* 2:743-50.
60. Diwan, V. K., and A. Thorson. 1999. Sex, gender, and tuberculosis. *Lancet* 353:1000-1.
61. Dolin, P. J., M. C. Raviglione, and A. Kochi. 1994. Global tuberculosis incidence and mortality during 1990-2000. *Bull World Health Organ* 72:213-20.
62. Donald, P. R. 1999. Children and tuberculosis: protecting the next generation? *Lancet* 353:1001-2.
63. Donnelly, J. J., J. B. Ulmer, and M. A. Liu. 1994. Immunization with DNA. *J Immunol Methods* 176:145-52.
64. Doran, T. J., A. L. Hodgson, J. K. Davies, and A. J. Radford. 1993. Characterisation of a highly repeated DNA sequence from *Mycobacterium bovis*. *FEMS Microbiol Lett* 111:147-52.
65. Dye, C., G. P. Garnett, K. Sleeman, and B. G. Williams. 1998. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet* 352:1886-91.
66. Eiglmeier, K., H. Fsihi, B. Heym, and S. T. Cole. 1997. On the catalase-peroxidase gene, *katG*, of *Mycobacterium leprae* and the implications for treatment of leprosy with isoniazid. *FEMS Microbiol Lett* 149:273-8.
67. Eiglmeier, K., N. Honore, S. A. Woods, B. Caudron, and S. T. Cole. 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol Microbiol* 7:197-206.
68. Ellner, J. J. 1996. Immunosuppression in tuberculosis. *Infect Agents Dis* 5:62-72.
69. Espitia, C., J. P. Lacleite, M. Mondragon-Palomino, A. Amador, J. Campuzano, A. Martens, M. Singh, R. Cicero, Y. Zhang, and C. Moreno. 1999. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology* 145 (Pt 12):3487-95.

70. Fine, P. E. 1995. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* 346:1339-45.
71. Fine, P. E. M. 1994. Immunities in and to tuberculosis: implications for pathogenesis and vaccination. *In* P. D. H. J. a. M. P. W. J. Keith (ed.), *Tuberculosis: Back to the Future*. John Wiley & Sons, Chichester. New York. Brisbane. Toronto. Singapore.
72. Flynn, J. L., M. M. Goldstein, K. J. Triebold, B. Koller, and B. R. Bloom. 1992. Major histocompatibility complex class I-restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci U S A* 89:12013-7.
73. Fomukong, N. G., T. H. Tang, S. al-Maamary, W. A. Ibrahim, S. Ramayah, M. Yates, Z. F. Zainuddin, and J. W. Dale. 1994. Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of IS6110. *Tuber Lung Dis* 75:435-40.
74. Fonseca, D. P., B. Benaissa-Trouw, M. van Engelen, C. A. Kraaijeveld, H. Snippe, and A. F. Verheul. 2001. Induction of cell-mediated immunity against *Mycobacterium tuberculosis* using DNA vaccines encoding cytotoxic and helper T-cell epitopes of the 38-kilodalton protein. *Infect Immun* 69:4839-45.
75. French, A. L., S. F. Welbel, S. E. Dietrich, L. B. Mosher, P. S. Breall, W. S. Paul, F. E. Kocka, and R. A. Weinstein. 1998. Use of DNA fingerprinting to assess tuberculosis infection control. *Ann Intern Med* 129:856-61.
76. Frieden, T. R., T. Sterling, A. Pablos-Mendez, J. O. Kilburn, G. M. Cauthen, and S. W. Dooley. 1993. The emergence of drug-resistant tuberculosis in New York City. *N Engl J Med* 328:521-6.
77. Frothingham, R., and W. A. Meeker-O'Connell. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 144 (Pt 5):1189-96.
78. Genewein, A., A. Telenti, C. Bernasconi, C. Mordasini, S. Weiss, A. M. Maurer, H. L. Rieder, K. Schopfer, and T. Bodmer. 1993. Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 342:841-4.

79. Gilsdorf, J. R. 1998. Antigenic diversity and gene polymorphisms in *Haemophilus influenzae*. *Infect Immun* 66:5053-9.
80. Goble, M., M. D. Iseman, L. A. Madsen, D. Waite, L. Ackerson, and C. R. Horsburgh, Jr. 1993. Treatment of 171 patients with pulmonary tuberculosis resistant to isoniazid and rifampin. *N Engl J Med* 328:527-32.
81. Goguet de la Salmoniere, Y. O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel. 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. *J Clin Microbiol* 35:2210-4.
82. Gonzalez, N., M. J. Torres, J. Aznar, and J. C. Palomares. 1999. Molecular analysis of rifampin and isoniazid resistance of *Mycobacterium tuberculosis* clinical isolates in Seville, Spain. *Tuber Lung Dis* 79:187-90.
83. Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole. 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol* 32:643-55.
84. Goyal, M., N. A. Saunders, J. D. van Embden, D. B. Young, and R. J. Shaw. 1997. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. *J Clin Microbiol* 35:647-51.
85. Griffin, J. F., C. G. Mackintosh, L. Slobbe, A. J. Thomson, and G. S. Buchan. 1999. Vaccine protocols to optimise the protective efficacy of BCG. *Tuber Lung Dis* 79:135-43.
86. Groenen, P. M., A. E. Bunschoten, D. van Soolingen, and J. D. van Embden. 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 10:1057-65.
87. Hermans, P. W., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, D. Frommel, and et al. 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. *J Infect Dis* 171:1504-13.

88. Hermans, P. W., D. van Soolingen, and J. D. van Embden. 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J Bacteriol* 174:4157-65.
89. Heym, B., P. M. Alzari, N. Honore, and S. T. Cole. 1995. Missense mutations in the catalase-peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Mol Microbiol* 15:235-45.
90. Heym, B., N. Honore, C. Truffot-Pernot, A. Banerjee, C. Schurra, W. R. Jacobs, Jr., J. D. van Embden, J. H. Grosset, and S. T. Cole. 1994. Implications of multidrug resistance for the future of short-course chemotherapy of tuberculosis: a molecular study. *Lancet* 344:293-8.
91. Heym, B., W. Philipp, and S. T. Cole. 1996. Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *Curr Top Microbiol Immunol* 215:49-69.
92. Heym, B., S. T. Cole. 1997. Multidrug resistance in *Mycobacterium tuberculosis*. *International Journal of Antimicrobial Agents* 8:61-70.
93. Hinman, M. B., J. A. Jones, and R. V. Lewis. 2000. Synthetic spider silk: a modular fiber. *Trends Biotechnol* 18:374-9.
94. Hirschfield, G. R., M. McNeil, and P. J. Brennan. 1990. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. *J Bacteriol* 172:1005-13.
95. Hollingshead, S. K., V. A. Fischetti, and J. R. Scott. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* 207:196-203.
96. Hommel, M., P. H. David, and L. D. Oligino. 1983. Surface alterations of erythrocytes in *Plasmodium falciparum* malaria. Antigenic variation, antigenic diversity, and the role of the spleen. *J Exp Med* 157:1137-48.
97. Honore, N., and S. T. Cole. 1993. Molecular basis of rifampin resistance in *Mycobacterium leprae*. *Antimicrob Agents Chemother* 37:414-8.
98. Honore, N., and S. T. Cole. 1994. Streptomycin resistance in mycobacteria. *Antimicrob Agents Chemother* 38:238-42.

99. Hopewell, P. C. 1992. Impact of human immunodeficiency virus infection on the epidemiology, clinical features, management, and control of tuberculosis. *Clin Infect Dis* 15:540-7.
100. Huygen, K. 1998. DNA vaccines: application to tuberculosis. *Int J Tuberc Lung Dis* 2:971-8.
101. Huygen, K., J. Content, O. Denis, D. L. Montgomery, A. M. Yawman, R. R. Deck, C. M. DeWitt, I. M. Orme, S. Baldwin, C. D'Souza, A. Drowart, E. Lozes, P. Vandebussche, J. P. Van Vooren, M. A. Liu, and J. B. Ulmer. 1996. Immunogenicity and protective efficacy of a tuberculosis DNA vaccine. *Nat Med* 2:893-8.
102. Jungblut, P. R., U. E. Schaible, H. J. Mollenkopf, U. Zimny-Arndt, B. Raupach, J. Mattow, P. Halada, S. Lamer, K. Hagens, and S. H. Kaufmann. 1999. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol* 33:1103-17.
103. Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35:907-14.
104. Kazwala, R. R., C. J. Daborn, J. M. Sharp, D. M. Kambarage, S. F. Jiwa, and N. A. Mbembati. 2001. Isolation of *Mycobacterium bovis* from human cases of cervical adenitis in Tanzania: a cause for concern? *Int J Tuberc Lung Dis* 5:87-91.
105. Kent, L., T. D. McHugh, O. Billington, J. W. Dale, and S. H. Gillespie. 1995. Demonstration of homology between IS6110 of *Mycobacterium tuberculosis* and DNAs of other *Mycobacterium* spp.? *J Clin Microbiol* 33:2290-3.
106. Kilburn, J. O., and J. Greenberg. 1977. Effect of ethambutol on the viable cell count in *Mycobacterium smegmatis*. *Antimicrob Agents Chemother* 11:534-40.
107. Kolk, A. H., A. R. Schuitema, S. Kuijper, J. van Leeuwen, P. W. Hermans, J. D. van Embden, and R. A. Hartskeerl. 1992. Detection of *Mycobacterium tuberculosis* in clinical samples by using polymerase chain reaction and a nonradioactive detection system. *J Clin Microbiol* 30:2567-75.

108. Kopanoff, D. E., J. O. Kilburn, J. L. Glassroth, D. E. Snider, Jr., L. S. Farer, and R. C. Good. 1978. A continuing survey of tuberculosis primary drug resistance in the United States: March 1975 to November 1977. A United States Public Health Service cooperative study. *Am Rev Respir Dis* 118:835-42.
109. Kox, L. F., D. Rhienthong, A. M. Miranda, N. Udomsantisuk, K. Ellis, J. van Leeuwen, S. van Heusden, S. Kuijper, and A. H. Kolk. 1994. A more reliable PCR for detection of *Mycobacterium tuberculosis* in clinical samples. *J Clin Microbiol* 32:672-8.
110. Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* 37:2607-18.
111. Kumaresan, J. A., A. K. Ahsan Ali, and L. M. Parkkali. 1998. Tuberculosis control in Bangladesh: success of the DOTS strategy. *Int J Tuberc Lung Dis* 2:992-8.
112. Lagranderie, M. R., A. M. Balazuc, E. Deriaud, C. D. Leclerc, and M. Gheorghiu. 1996. Comparison of immune responses of mice immunized with five different *Mycobacterium bovis* BCG vaccine strains. *Infect Immun* 64:1-9.
113. Lee, B. Y., S. A. Hefta, and P. J. Brennan. 1992. Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect Immun* 60:2066-74.
114. Levitskaya, J., M. Coram, V. Levitsky, S. Imreh, P. M. Steigerwald-Mullen, G. Klein, M. G. Kurilla, and M. G. Masucci. 1995. Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* 375:685-8.
115. Levitskaya, J., A. Sharipo, A. Leonchiks, A. Ciechanover, and M. G. Masucci. 1997. Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc Natl Acad Sci U S A* 94:12616-21.

116. Long, C. D., R. N. Madraswala, and H. S. Seifert. 1998. Comparisons between colony phase variation of *Neisseria gonorrhoeae* FA1090 and pilus, pilin, and S-pilin expression. *Infect Immun* 66:1918-27.
117. Lowrie, D. B., R. E. Tascon, V. L. Bonato, V. M. Lima, L. H. Faccioli, E. Stavropoulos, M. J. Colston, R. G. Hewinson, K. Moelling, and C. L. Silva. 1999. Therapy of tuberculosis in mice by DNA vaccination. *Nature* 400:269-71.
118. Lozes, E., K. Huygen, J. Content, O. Denis, D. L. Montgomery, A. M. Yawman, P. Vandenbussche, J. P. Van Vooren, A. Drowart, J. B. Ulmer, and M. A. Liu. 1997. Immunogenicity and efficacy of a tuberculosis DNA vaccine encoding the components of the secreted antigen 85 complex. *Vaccine* 15:830-3.
119. Magdalena, J., P. Supply, and C. Locht. 1998. Specific differentiation between *Mycobacterium bovis* BCG and virulent strains of the *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 36:2471-6.
120. Magdalena, J., A. Vachee, P. Supply, and C. Locht. 1998. Identification of a new DNA region specific for members of *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 36:937-43.
121. Manchester, K. 1984. Tuberculosis and leprosy in antiquity: an interpretation. *Med Hist* 28:162-73.
122. Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A* 98:1901-6.
123. McCrea, K. W., J. L. Sauver, C. F. Marrs, D. Clemans, and J. R. Gilsdorf. 1998. Immunologic and structural relationships of the minor pilus subunits among *Haemophilus influenzae* isolates. *Infect Immun* 66:4788-96.
124. Meyer, T. F., C. P. Gibbs, and R. Haas. 1990. Variation and control of protein expression in *Neisseria*. *Annu Rev Microbiol* 44:451-77.
125. Michel, M. L., H. L. Davis, M. Schleef, M. Mancini, P. Tiollais, and R. G. Whalen. 1995. DNA-mediated immunization to the hepatitis B surface antigen in mice: aspects of the humoral response mimic hepatitis B viral infection in humans. *Proc Natl Acad Sci U S A* 92:5307-11.

126. Mistry, N. F., A. M. Iyer, T. D'Souza D, G. M. Taylor, D. B. Young, and N. H. Antia. 2002. Spoligotyping of *Mycobacterium tuberculosis* isolates from multiple-drug-resistant tuberculosis patients from Bombay, India. *J Clin Microbiol* 40:2677-80.
127. Mizrahi, V., and S. J. Andersen. 1998. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol* 29:1331-9.
128. Moazed, D., and H. F. Noller. 1987. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* 327:389-94.
129. Montgomery, D. L., J. W. Shiver, K. R. Leander, H. C. Perry, A. Friedman, D. Martinez, J. B. Ulmer, J. J. Donnelly, and M. A. Liu. 1993. Heterologous and homologous protection against influenza A by DNA vaccination: optimization of DNA vectors. *DNA Cell Biol* 12:777-83.
130. Morse D, B. D., Ucko PJ. 1964. Tuberculosis in ancient Egypt. *American Review of Respiratory Disease* 90:524-541.
131. Moxon, E. R. a. W., C. 1999. DNA microsatellites: agents of evolution? *Scien Amer*:72-77.
132. Murray, C. J., and J. A. Salomon. 1998. Modeling the impact of global tuberculosis control strategies. *Proc Natl Acad Sci U S A* 95:13881-6.
133. Murray, C. J., K. Styblo, and A. Rouillon. 1990. Tuberculosis in developing countries: burden, intervention and cost. *Bull Int Union Tuberc Lung Dis* 65:6-24.
134. Mushtaque, A., and R. Chowdhury. 1999. Success with the DOTS strategy. *Lancet* 353:1003-4.
135. Olive, D. M., and P. Bean. 1999. Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* 37:1661-9.
136. Orme, I. M. 1997. Progress in the development of new vaccines against tuberculosis. *Int J Tuberc Lung Dis* 1:95-100.
137. Park, H., H. Jang, C. Kim, B. Chung, C. L. Chang, S. K. Park, and S. Song. 2000. Detection and identification of mycobacteria by amplification of the internal transcribed spacer regions with genus- and species-specific PCR primers. *J Clin Microbiol* 38:4080-5.

138. Pays, E. 1985. Gene conversion in trypanosome antigenic variation. *Prog Nucleic Acid Res Mol Biol* 32:1-26.
139. Philipp, W. J., S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole. 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc Natl Acad Sci U S A* 93:3132-7.
140. Poulet, S., and S. T. Cole. 1995. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* 163:87-95.
141. Poulet, S., and S. T. Cole. 1995. Repeated DNA sequences in mycobacteria. *Arch Microbiol* 163:79-86.
142. Prod'hom, G., C. Guilhot, M. C. Gutierrez, A. Varnerot, B. Gicquel, and V. Vincent. 1997. Rapid discrimination of *Mycobacterium tuberculosis* complex strains by ligation-mediated PCR fingerprint analysis. *J Clin Microbiol* 35:3331-4.
143. Qian, L., J. D. Van Embden, A. G. Van Der Zanden, E. F. Weltevreden, H. Duanmu, and J. T. Douglas. 1999. Retrospective analysis of the Beijing family of *Mycobacterium tuberculosis* in preserved lung tissues. *J Clin Microbiol* 37:471-4.
144. Radhakrishnan, I., M. Y. K., R. A. Kumar, and S. Mundayoor. 2001. Implications of low frequency of IS6110 in fingerprinting field isolates of *Mycobacterium tuberculosis* from Kerala, India. *J Clin Microbiol* 39:1683.
145. Radman, M. 1999. Enzymes of evolutionary change. *Nature* 401:866-7, 869.
146. Ramakrishnan, L., N. A. Federspiel, and S. Falkow. 2000. Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* 288:1436-9.
147. Ramaswamy, S., and J. M. Musser. 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* 79:3-29.
148. Raviglione, M. C., D. E. Snider, Jr., and A. Kochi. 1995. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. *Jama* 273:220-6.

149. Rieder, H. L., G. M. Cauthen, G. W. Comstock, and D. E. Snider, Jr. 1989. Epidemiology of tuberculosis in the United States. *Epidemiol Rev* 11:79-98.
150. Robertson, B. D., and T. F. Meyer. 1992. Genetic variation in pathogenic bacteria. *Trends Genet* 8:422-7.
151. Romano, M. I., A. Alito, J. C. Fisanotti, F. Bigi, I. Kantor, M. E. Cicuta, and A. Cataldi. 1996. Comparison of different genetic markers for molecular epidemiology of bovine tuberculosis. *Vet Microbiol* 50:59-71.
152. Rook, G. A., and B.R. Bloom. 1994. Mechanisms of pathogenesis of tuberculosis, p. 485-501. *In* B. R. Bloom (ed.), *Tuberculosis: pathogenesis, protection, and control*. ASM Press, Washington, D.C.
153. Rosenkrands, I., K. Weldingh, S. Jacobsen, C. V. Hansen, W. Florio, I. Gianetri, and P. Andersen. 2000. Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* 21:935-48.
154. Ross, B. C., K. Raio, K. Jackson, and B. Dwyer. 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J Clin Microbiol* 30:942-6.
155. Sahadevan, R., S. Narayanan, C. N. Paramasivan, R. Prabhakar, and P. R. Narayanan. 1995. Restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, India, by use of direct-repeat probe. *J Clin Microbiol* 33:3037-9.
156. Saint-Joanis, B., H. Souchon, M. Wilming, K. Johnsson, P. M. Alzari, and S. T. Cole. 1999. Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from *Mycobacterium tuberculosis*. *Biochem J* 338 (Pt 3):753-60.
157. Sales, M. P., G. M. Taylor, S. Hughes, M. Yates, G. Hewinson, D. B. Young, and R. J. Shaw. 2001. Genetic diversity among *Mycobacterium bovis* isolates: a preliminary study of strains from animal and human sources. *J Clin Microbiol* 39:4558-62.

158. Sambrook, J., Fritscii, E.F., and Maniatis,T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, New York.
159. Sampson, S. L., P. Lukey, R. M. Warren, P. D. van Helden, M. Richardson, and M. J. Everett. 2001. Expression, characterization and subcellular localization of the Mycobacterium tuberculosis PPE gene Rv1917c. Tuberculosis (Edinb) 81:305-17.
160. Scorpio, A., and Y. Zhang. 1996. Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat Med 2:662-7.
161. Senaratne, R. H., H. Mobasheri, K. G. Papavinasasundaram, P. Jenner, E. J. Lea, and P. Draper. 1998. Expression of a gene for a porin-like protein of the OmpA family from Mycobacterium tuberculosis H37Rv. J Bacteriol 180:3541-7.
162. Seth, V. 1991. Tuberculosis in Children: epidemiology, diagnosis and treatment, p. 1-7, Tuberculosis in children. Indian Academy of Pediatrics, New Delhi.
163. Shiver, J. W., Perry, H.C., and Davies, M.E. 1995. Immune responses to HIV gp120 elicited by DNA vaccination., p. 95-98. *In* R. M. Chanock, Brown, F., Ginsberg, H.S., and Norrby, E. (ed.), Vaccines. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
164. Silva, C. L., M. F. Silva, R. C. Pietro, and D. B. Lowrie. 1994. Protection against tuberculosis by passive transfer with T-cell clones recognizing mycobacterial heat-shock protein 65. Immunology 83:341-6.
165. Singh, K. K., X. Zhang, A. S. Patibandla, P. Chien, Jr., and S. Laal. 2001. Antigens of Mycobacterium tuberculosis expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect Immun 69:4185-91.
166. Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med 330:1703-9.

167. Small, P. M., and A. Moss. 1993. Molecular epidemiology and the new tuberculosis. *Infect Agents Dis* 2:132-8.
168. Snider, D. E., Jr. 1994. Tuberculosis: the world situation. History of the disease and efforts to combat it. *In* P. J. D. H. a. M. P.W.J (ed.), *Tuberculosis: Back to the Future*. John Wiley & Sons.
169. Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 94:9869-74.
170. Stern, A., and T. F. Meyer. 1987. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. *Mol Microbiol* 1:5-12.
171. Stoenner, H. G., T. Dodd, and C. Larsen. 1982. Antigenic variation of *Borrelia hermsii*. *J Exp Med* 156:1297-311.
172. Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht. 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 39:3563-71.
173. Supply, P., J. Magdalena, S. Himpens, and C. Locht. 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol Microbiol* 26:991-1003.
174. Supply, P., E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* 36:762-71.
175. Takahashi, M., Y. Kazumi, Y. Fukasawa, K. Hirano, T. Mori, J. W. Dale, and C. Abe. 1993. Restriction fragment length polymorphism analysis of epidemiologically related *Mycobacterium tuberculosis* isolates. *Microbiol Immunol* 37:289-94.
176. Takayama, K., and J. O. Kilburn. 1989. Inhibition of synthesis of arabinogalactan by ethambutol in *Mycobacterium smegmatis*. *Antimicrob Agents Chemother* 33:1493-9.

177. Tascon, R. E., M. J. Colston, S. Ragno, E. Stavropoulos, D. Gregory, and D. B. Lowrie. 1996. Vaccination against tuberculosis by DNA injection. *Nat Med* 2:888-92.
178. Tekaia, F., S. V. Gordon, T. Garnier, R. Brosch, B. G. Barrell, and S. T. Cole. 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* 79:329-42.
179. Toossi, Z. 1996. Cytokine circuits in tuberculosis. *Infect Agents Dis* 5:98-107.
180. Torrea, G., C. Offredo, M. Simonet, B. Gicquel, P. Berche, and C. Pierre-Audigier. 1996. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* 34:1043-9.
181. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76:4350-4.
182. Tsuyuguchi, I. 1996. Regulation of the human immune response in tuberculosis. *Infect Agents Dis* 5:82-97.
183. Ulmer, J. B., J. J. Donnelly, S. E. Parker, G. H. Rhodes, P. L. Felgner, V. J. Dwarki, S. H. Gromkowski, R. R. Deck, C. M. DeWitt, A. Friedman, and et al. 1993. Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* 259:1745-9.
184. Ulmer, J. B., M. A. Liu, D. L. Montgomery, A. M. Yawman, R. R. Deck, C. M. DeWitt, J. Content, and K. Huygen. 1997. Expression and immunogenicity of *Mycobacterium tuberculosis* antigen 85 by DNA vaccination. *Vaccine* 15:792-4.
185. van Deutekom, H., J. J. Gerritsen, D. van Soolingen, E. J. van Ameijden, J. D. van Embden, and R. A. Coutinho. 1997. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. *Clin Infect Dis* 25:1071-7.
186. van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and et al. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31:406-9.

187. van Soolingen, D., P. E. de Haas, P. W. Hermans, P. M. Groenen, and J. D. van Embden. 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* 31:1987-95.
188. van Soolingen, D., P. W. Hermans, P. E. de Haas, D. R. Soll, and J. D. van Embden. 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 29:2578-86.
189. van Soolingen, D., T. Hoogenboezem, P. E. de Haas, P. W. Hermans, M. A. Koedam, K. S. Teppema, P. J. Brennan, G. S. Besra, F. Portaels, J. Top, L. M. Schouls, and J. D. van Embden. 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 47:1236-45.
190. van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol* 33:3234-8.
191. Vekemans, M., M. Cartoux, S. Diagbouga, M. Dembele, B. Kone, A. Delafosse, and A. Dera. 1999. Potential source of human exposure to *Mycobacterium bovis* in Burkina Faso, in the context of the HIV epidemic. *Clin Microbiol Infect* 5:617-621.
192. Wallis, R. S., M. Vjecha, M. Amir-Tahmasseb, A. Okwera, F. Byekwaso, S. Nyole, S. Kabengeru, R. D. Mugerwa, and J. J. Ellner. 1993. Influence of tuberculosis on human immunodeficiency virus (HIV-1): enhanced cytokine expression and elevated beta 2-microglobulin in HIV-1-associated tuberculosis. *J Infect Dis* 167:43-8.
193. Wang, B., K. E. Ugen, V. Srikantan, M. G. Agadjanyan, K. Dang, Y. Refaeli, A. I. Sato, J. Boyer, W. V. Williams, and D. B. Weiner. 1993. Gene inoculation generates immune responses against human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* 90:4156-60.

194. Wolff, J. A., J. J. Ludtke, G. Acsadi, P. Williams, and A. Jani. 1992. Long-term persistence of plasmid DNA and foreign gene expression in mouse muscle. *Hum Mol Genet* 1:363-9.
195. Wolff, J. A., R. W. Malone, P. Williams, W. Chong, G. Acsadi, A. Jani, and P. L. Felgner. 1990. Direct gene transfer into mouse muscle in vivo. *Science* 247:1465-8.
196. Yang, Z., P. F. Barnes, F. Chaves, K. D. Eisenach, S. E. Weis, J. H. Bates, and M. D. Cave. 1998. Diversity of DNA fingerprints of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol* 36:1003-7.
197. Yang, Z. H., I. Mtoni, M. Chonde, M. Mwasekaga, K. Fuursted, D. S. Askgard, J. Bennedsen, P. E. de Haas, D. van Soolingen, J. D. van Embden, and et al. 1995. DNA fingerprinting and phenotyping of *Mycobacterium tuberculosis* isolates from human immunodeficiency virus (HIV)-seropositive and HIV-seronegative patients in Tanzania. *J Clin Microbiol* 33:1064-9.
198. Yates, M. D., C. H. Collins, and J. M. Grange. 1978. Differentiation of BCG from other variants of *Mycobacterium tuberculosis* isolated from clinical material. *Tubercle* 59:143-6.
199. Yuen, L. K., B. C. Ross, K. M. Jackson, and B. Dwyer. 1993. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. *J Clin Microbiol* 31:1615-8.
200. Zhang, J. R., and S. J. Norris. 1998. Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun* 66:3698-704.
201. Zhang, Y., B. Heym, B. Allen, D. Young, and S. Cole. 1992. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* 358:591-3.
202. Zwarenstein, M., J. H. Schoeman, C. Vundule, C. J. Lombard, and M. Tatley. 1998. Randomised controlled trial of self-supervised and directly observed treatment of tuberculosis. *Lancet* 352:1340-3.

Publications

Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?

Sayera Banu,^{1,2} Nadine Honoré,¹ Brigitte Saint-Joanis,¹ Dana Philpott³, Marie-Christine Prévost⁴ and Stewart T. Cole^{1*}

¹Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.

²International Centre for Diarrhoeal Disease Research, Bangladesh, GPO-128, Dhaka-1000, Bangladesh.

³Unité de Pathogénie Microbienne Moléculaire, and

⁴Plateforme de Microscopie Electronique, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.

Summary

Mycobacterium tuberculosis H37Rv contains 67 PE-PGRS genes, with multiple tandem repetitive sequences, encoding closely related proteins that are exceptionally rich in glycine and alanine. As no functional information was available, 10 of these genes were selected and shown to be expressed *in vitro* by reverse transcription–polymerase chain reaction (RT–PCR). Antibodies against five PE-PGRS proteins, raised in mice by DNA vaccination, detected single proteins when the same plasmid constructs used for immunization were expressed in epithelial cells or in reticulocyte extracts, confirming that the PE-PGRS proteins are antigenic. As expected from the conserved repetitive structure, the antibodies cross-reacted with more than one PE-PGRS protein, suggesting that different proteins share common epitopes. PE-PGRS proteins were detected by Western blotting in five different mycobacterial species (*M. tuberculosis*, *M. bovis* BCG, *M. smegmatis*, *M. marinum* and *M. goodii*) and 11 clinical isolates of *M. tuberculosis*. Whole-genome comparisons of *M. tuberculosis* predicted allelic diversity in the PE-PGRS family, and this was confirmed by immunoblot studies as size variants were detected in clinical strains. Subcellular fractionation studies and immunoelectron microscopy localized many PE-PGRS proteins in the cell wall and cell membrane of *M. tuberculosis*. The data suggest that some PE-PGRS proteins are variable surface antigens.

Introduction

Although tuberculosis remains one of the leading causes of death worldwide, the only vaccine currently available is BCG, the 'bacille de Calmette et Guérin', which has been found to impart highly variable levels of efficacy, particularly against pulmonary disease (Fine, 1995). There is therefore an urgent need to identify protective antigens of the tubercle bacillus so that an effective vaccine can be developed to control this devastating epidemic and also to understand the reasons for the failure of BCG to confer uniform protection.

Deciphering the entire genome sequence of *Mycobacterium tuberculosis* H37Rv has opened a new era in mycobacteriology and revealed many unsuspected aspects of the biology of this important pathogen (Cole *et al.*, 1998). Among the major surprises was the finding that 8% of the potential coding capacity was accounted for by two large unrelated gene families encoding the PE and PPE proteins (Cole, 1999). The names PE and PPE are derived from the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) found in most cases near the N-terminus of these glycine- and alanine-rich proteins in conserved domains of ~ 110 and 180 amino acid residues respectively (Cole and Barrell, 1998). There are ~ 100 members of the PE family, whereas the PPE protein family, which will not be discussed further here, comprises 68 members. The PE family can be subdivided into three classes on the basis of their domain structure (Fig. 1; Brosch *et al.*, 2000). There are 29 proteins that contain the PE domain alone, and a further eight in which the PE domain is followed by C-terminal segments of unique sequence. The largest class of the PE family, with 67 members in *M. tuberculosis* H37Rv, is referred to as the PE-PGRS subfamily, as these proteins consist of the PE domain followed by a C-terminal extension with multiple tandem repetitions of Gly-Gly-Ala or Gly-Gly-Asn encoded by the PGRS motif (polymorphic GC-rich repetitive sequences). PE-PGRS proteins are predicted to contain as many as 1900 amino acids, of which up to 50% can be glycine.

As the name implies, the PGRS genes are highly GC rich (about 80%) and a major source of polymorphism in the *M. tuberculosis* complex, which otherwise displays remarkable genetic homogeneity and very few single-nucleotide polymorphisms (Sreevatsan *et al.*, 1997). A number of observations related to the PE and PPE families inspired us to speculate that they might be of immuno-

Accepted 20 November 2001. *For correspondence. E-mail stcole@pasteur.fr. Tel.: (+33) 1 45 88 84 46. Fax: (+33) 1 40 61 35 83.

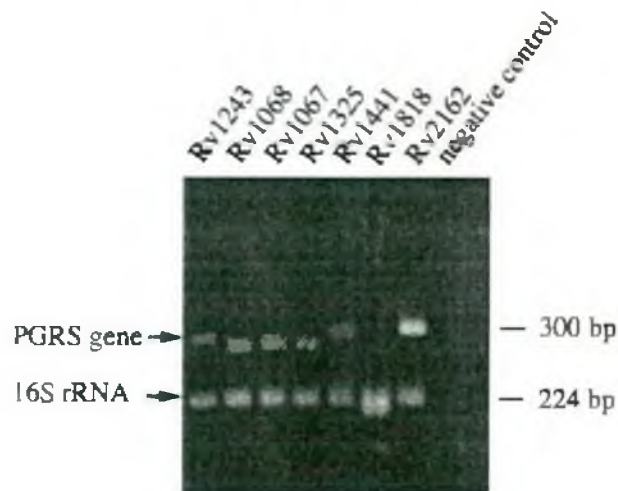


Fig. 2. Duplex RT-PCR analysis of total RNA extracted from *Mycobacterium tuberculosis* H37Rv. Primers were designed to amplify between 200 and 300 bp segments of 10 different PE-PGRS genes. A 224 bp segment of 16S rRNA of *M. tuberculosis* was amplified in the same PCR as the internal control. Seven of the 10 selected PE-PGRS genes are shown. The same PCR mixture without reverse transcriptase was used as the negative control. The sizes (in bp) of the RT-PCR product of the Rv2162 gene and the 16S rRNA are indicated on the right.

amplification easier, with the size of the selected PE-PGRS genes varying from 1 to 2 kb. Suitable primers were designed to amplify regions of between 200 and 300 bp of each PE-PGRS gene after cDNA production (Table 1). Duplex RT-PCR results for seven of the 10 selected PE-PGRS genes are shown in Fig. 2. A 224 bp segment of 16S rRNA was amplified in the same PCR as the internal control, and this was generally seen below the products derived from the PE-PGRS mRNA on agarose gel electrophoresis (Fig. 2). The size of the PCR product

of Rv2162 is 219 bp, and the band appears just below the PCR product of 16S rRNA. No amplicons were obtained in the absence of reverse transcription. These experiments demonstrated that all 10 PE-PGRS genes examined were transcribed in cultured mycobacteria (Fig. 2; data not shown).

Antibody production

After confirming that the selected PE-PGRS genes were expressed in *M. tuberculosis* H37Rv (shown above), DNA vaccination was used to produce antibodies against selected PE-PGRS proteins in mice. DNA vaccination has been demonstrated to be effective in inducing both antibody and cell-mediated immune responses (Davis *et al.*, 1993; Ulmer *et al.*, 1993; Donnelly *et al.*, 1994; Huygen *et al.*, 1996). The genes for five PE-PGRS proteins (Rv0980c, Rv1067c, Rv1068c, Rv1441c and Rv1983) were cloned downstream of the cytomegalovirus promoter in the V1Jns.tPA vector (Montgomery *et al.*, 1993), and the recombinant plasmids were used to immunize mice and induce antibody responses. To determine the specificity of the resultant murine antibodies against the PE-PGRS proteins, the same plasmids were used to express proteins in eukaryotic cells. The PE-PGRS constructs were transfected into HeLa and HEK 293 epithelial cells, and cell extracts were then probed with the corresponding antibody and immune complexes visualized by fluorescence microscopy. Typical results for the construct producing Rv1441c are presented (Fig. 3), and similar results were obtained with Rv0980c, Rv1067c, Rv1068c and Rv1983 (data not shown). These experiments showed that the PE-PGRS genes were strongly expressed in both HeLa and HEK 293 (data not shown) cell lines.

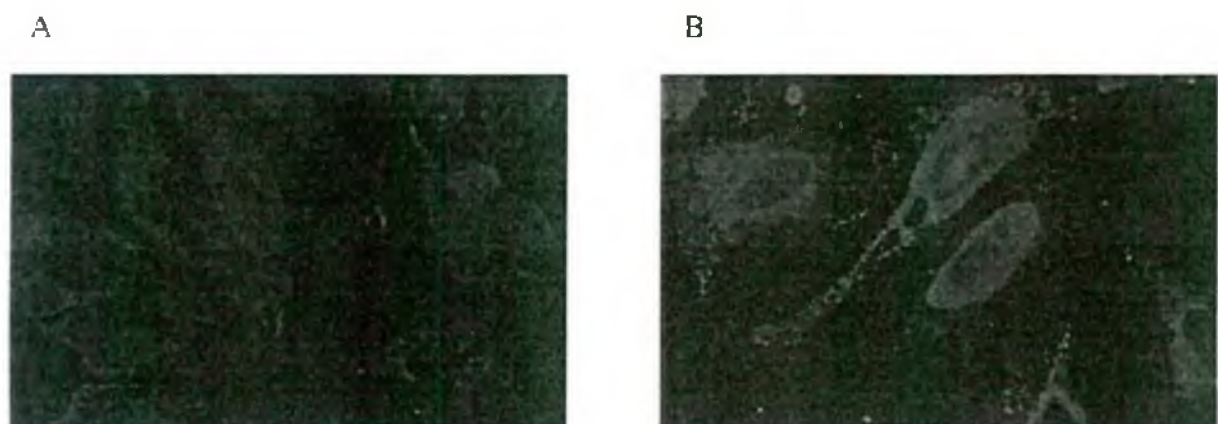


Fig. 3. Expression of Rv1441c in HeLa epithelial cells. HeLa cells were transfected with the vector construct alone (A) or the Rv1441c construct (B). Forty-eight hours after transfection, cells were fixed and stained with the Rv1441c-specific antibody and visualized by fluorescence microscopy. Strong perinuclear staining of Rv1441c is seen in (B).



Fig. 8. Immunogold localization of PE-PGRS proteins in *M. tuberculosis* H37Rv. Intact cells were incubated with antibody 1441c and prepared for electron microscopy by negative contrast as described. No labelling of the surface was obtained in control experiments from which this antibody had been omitted.

revealed with PGRS probes (Ross *et al.*, 1992; van Soolingen *et al.*, 1993; Poulet and Cole, 1995). More recently, comparisons of the genome sequences of *M. tuberculosis* strains H37Rv and CDC1551 have shown that, apart from some 10 insertion-deletion events, the major source of genetic diversity occurs in the PE and PPE families (Betts *et al.*, 2000). In the course of the present study, pairwise comparisons of the respective PE and PE-PGRS gene sequences in these two strains have been undertaken. These revealed that all 37 PE genes were identical in both strains. In contrast, 39 of the 62 common PE-PGRS genes showed extensive divergence that would result in the absence of the protein or a difference in size as a result of frameshift mutations or in frame insertion and deletion events respectively. This is in good agreement with the findings of our Western blot analysis using the most cross-reactive antibodies (Figs 5–7), as these uncovered two or three differences in the size or presence among the five to 10 PE-PGRS proteins detected. The loss of immunoreactivity could result from either inactivation of the corresponding gene or amino acid sequence changes that alter B-cell epitopes.

Genetic variation results from spontaneous changes in the genes of a species and, in pathogens, can lead to phenotypic alterations that include phase and antigenic variation (Robertson and Meyer, 1992). Genetic variation has been well documented in a number of pathogenic bacteria such as *Neisseria* spp., *Haemophilus* spp., *Streptococcus pneumoniae*, *Borrelia hermsii* and *Borrelia burgdorferi* (Stem and Meyer, 1987; Brunham, 1993; Gilsdorf, 1998; Long *et al.*, 1998; McCrear *et al.*, 1998; Zhang and Norns, 1998; Chausse *et al.*, 1999). The principle mechanisms involved are recombinational events and slipped-strand mispairing during replication, both associated with repetitive sequences (Robertson and Meyer, 1992). Homologous recombination between intra-genic tandem repeats, leading to deletion or duplication of runs of amino acids, is a frequent source of size and antigenic variation in the M protein present on the *Streptococcal* cell surface and results in evasion of the immune system (Hollingshead *et al.*, 1987). Variation in *opa* gene expression occurs in *Neisseria* spp. by slipped-strand

mispairing (Meyer *et al.*, 1990), involving CTCTT repeat units present, and leads to the *opa* gene being shifted in or out of frame.

A characteristic feature of a PE-PGRS gene is its GC richness, stemming from a 9bp repeat CGGCGGCAA arranged in tandem copies (Poulet and Cole, 1995). It is tempting to speculate that similar recombinational events involving these sequences, or strand slippage during replication, might be responsible for PE-PGRS variation in *M. tuberculosis*. The data presented here show that the PE-PGRS proteins vary in size in different mycobacteria and also in different clinical isolates of *M. tuberculosis* (Figs 5 and 6). Comparative genomics of *M. tuberculosis* strains H37Rv and CDC1551 underlines the role of insertion-deletion events in generating diversity in the PE-PGRS family (Fig. 1; Cole *et al.*, 1998; Betts *et al.*, 2000). Taken together, these data strongly suggest that some PE-PGRS proteins may correspond to variable surface antigens.

In order to escape the host's immune system, bacteria mutate extensively to generate rare variants. The role of DNA polymerases IV and V, also known as mutases, in this process has been recognized recently in micro-organisms, and the targets are the 'contingency genes' often encoding the surface antigens recognized by the immune system (Moxon and Wills, 1999; Radman, 1999). The corresponding genes contain simple, repeated sequences, which are preferential sites for DNA polymerase IV-mediated slippage mutations resulting in genetic diversity. *M. tuberculosis* contains genes for both Pol IV and Pol V (Rv1537, Rv3056) and should thus produce the enzymes that might generate variability in the PE-PGRS family.

Functional information remains scarce for the PE-PGRS proteins. On the basis of sequence similarity to EBNA, the Gly- and Ala-rich Epstein-Barr nuclear antigen, it has been proposed that PE-PGRS proteins could act as proteasome inhibitors and block antigen presentation by the major histocompatibility complex (MHC) I pathway (Levitskaya *et al.*, 1995; 1997; Cole *et al.*, 1998). Our finding that some PE-PGRS proteins are located in the cell envelope is of relevance in this context.

Plasmid construction

PE-PGRS genes from *M. tuberculosis* H37Rv strain were amplified by PCR from H37Rv chromosomal DNA, cosmids (Phillip *et al.*, 1996) or Bac clones (Brosch *et al.*, 1998) using customized primers with *Bgl*II or *Bam*HI sites and cloned in V1Jns.tPA vector (Shiver *et al.*, 1995). Briefly, the following 30-cycle procedure was used with a PTC-100 thermal cycler (MJ Research): 2 min at 92°C, 2 min at 70°C and 2 min at 72°C. To ensure complete synthesis of products, the final elongation step was performed for 10 min at 72°C. Amplified DNA was digested with *Bgl*II or *Bam*HI, isolated on agarose gels (1%) containing TBE buffer, detected with ethidium bromide and extracted using a GeneClean II kit (Bio 101). Fragments were cloned into the *Bgl*II or *Bam*HI site of the V1Jns.tPA vector, as described previously (Montgomery *et al.*, 1993). PCR products were also cloned in pUC19 and pET vectors and transformed into *E. coli*.

Vaccination

Five PE-PGRS plasmid constructs based on the DNA vaccination vector, V1Jns.tPA (Montgomery *et al.*, 1993; Shiver *et al.*, 1995), were purified, then injected intramuscularly in BALB.c mice (6- to 8-week-old females, purchased from IFFA-CREDO). After anaesthesia by intraperitoneal injection of ketamine (100 mg kg⁻¹), groups of five to eight mice were immunized by intramuscular injection (three times at 3 week intervals) in both quadriceps with 100 µg of PE-PGRS DNA in 100 µg of saline using a 0.3 cc insulin syringe (Huygen *et al.*, 1996). As negative controls, mice were uninjected, injected with saline or with control plasmid DNA (not containing a gene insert) in saline. Pooled sera from blood samples were taken 3 weeks after the third immunization.

Expression in HeLa and HEK 293 cells

To verify the expression of PE-PGRS proteins and the specificity of the corresponding antibodies, HeLa and HEK 293 epithelial cells were transfected with the same constructs used for DNA vaccination, and expression of the five PE-PGRS (Rv0980c, Rv1068c, Rv1067c, Rv1441c and Rv1983) proteins was observed by both immunofluorescence and Western blot analysis. HeLa and HEK 293 cells were plated onto coverslips or in six-well plates at a density of 1 × 10⁵ cells ml⁻¹ and transfected the following day using Eugene reagent (Roche Biochemicals) with 500 ng of plasmid PE-PGRS DNA or vector DNA alone, according to the manufacturer's directions. Analyses were performed 48 h after transfection. For immunofluorescence, cells were fixed with 4% paraformaldehyde in PBS for 15 min, permeabilized with 0.2% Triton X-100 in PBS and subsequently stained with the corresponding antibody. After labelling with Cy3-linked anti-rabbit antibodies (Jackson ImmunoResearch Laboratory), coverslips were mounted and viewed with a conventional immunofluorescence microscope (BX50; Olympus Optical). For Western blot analysis, HeLa and HEK 293 cells grown in six-well plates were rinsed once with PBS and lysed with lysis buffer [20 mM HEPES, pH 7.8, 10 mM KCl, 1 mM dithiothreitol, 0.5 mM EDTA, 0.5 mM EGTA, 1% Nonidet P-40

and a protease inhibitor tablet (mini-Complete; Roche)]. Samples were vortexed and centrifuged at 14 000 r.p.m. for 5 min. The resulting supernatant was then run on a 10% SDS-PAGE gel, transferred to nitrocellulose, and specific antibody binding was revealed using an enhanced chemiluminescence detection system (Amersham).

Expression in reticulocyte extracts

Two of the PE-PGRS genes were expressed in reticulocyte extracts using TNT coupled transcription/translation systems (Promega). Rv1068c and Rv1441c genes were cloned in the pET22 vector. TNT T7 quick master mix (40 µl), 2 µl of [³⁵S]-methionine (1000 Ci mmol⁻¹), 1 µg of plasmid DNA, nuclease-free water to a final volume of 50 µl were placed in a 1.5 ml microcentrifuge tube and gently mixed by pipetting. The reaction mixture was incubated at 30°C for 60–90 min. The synthesized proteins were analysed by SDS-PAGE and autoradiography.

Western blotting

Different mycobacteria and clinical isolates of *M. tuberculosis* were grown in Middlebrook 7H9 medium. Harvested bacteria were resuspended in 1/20th of the original culture volume in PBS, lysed and stored at -20°C until use. Extracted proteins were separated by SDS-PAGE in PBS on slab gels of 12% polyacrylamide. Total protein (10 µg) was loaded in each lane of the gel. A mixture of standard protein markers was used for the determination of molecular mass. For immunoblotting analysis, proteins were transferred from polyacrylamide gels onto nitrocellulose membranes (Towbin *et al.*, 1979). To test the immune responses of murine antibodies to different mycobacterial proteins, blots were hybridized with serum samples diluted 1:100. Bound antibody was revealed by incubation with peroxidase-conjugated anti-mouse immunoglobulin.

Subcellular fractionation and immunoelectron microscopy

Subcellular fractions of *M. tuberculosis* (produced through funds from the National Institutes of Health, National Institute of Allergy and Infectious Diseases, contract NO1-AI-75320) were obtained from the Department of Microbiology, Colorado State University, CO, USA. Cell wall, cell membrane, cytosol and whole-cell lysate were separated using the procedure described previously (Hirschfield *et al.*, 1990; Lee *et al.*, 1992). In brief, each strain was grown to late log phase (day 14) in glycerol-alanine-salts (GAS) medium, washed with PBS (pH 7.4) and inactivated by gamma irradiation. The bacilli were suspended (2 g ml⁻¹) in PBS containing DNase, RNase and a proteinase inhibitor cocktail [phenylmethylsulphonyl fluoride (PMSF), pepstatin A and leupeptin] and broken in a French press pressure cell at 4°C. Unbroken cells were removed by low-speed (3000 g) centrifugation. The supernatant isolated is a crude or whole-cell isolate containing the bacterial cell, including cell wall, cytosol and membrane. The cell wall was isolated by centrifugation at 27 000

- and protective efficacy of a tuberculosis DNA vaccine (see comments). *Nature Med* 2: 893–898.
- Jungblut, P.R., Schaible, U.E., Mollenkopf, H.-J., Zimny-Arndt, U., Raupach, B., Mattow, J., et al. (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol* 33: 1103–1117.
- Lee, B.-Y., Hefta, S.A., and Brennan, P.J. (1992) Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect Immun* 60: 2066–2074.
- Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald, M.P., Klein, G., et al. (1995) Inhibition of antigen processing by the internal repeat region of the Epstein–Barr virus nuclear antigen-1. *Nature* 375: 685–688.
- Levitskaya, J., Sharpo, A., Leonchiks, A., Ciechanover, A., and Masucci, M.G. (1997) Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein–Barr virus nuclear antigen 1. *Proc Natl Acad Sci USA* 94: 12616–12621.
- Long, C.D., Madraswala, R.N., and Seifert, H.S. (1998) Comparisons between colony phase variation of *Neisseria gonorrhoeae* FA1090 and pilus, pilin, and S-pilin expression. *Infect Immun* 66: 1918–1927.
- McCrea, K.W., Sauver, J.L.S., Marrs, C.F., Clemans, D., and Gilsdorf, J.R. (1998) Immunologic and structural relationships of the minor pilus subunits among *Haemophilus influenzae* isolates. *Infect Immun* 66: 4788–4796.
- Meyer, T.F., Gibbs, C.P., and Haas, R. (1990) Variation and control of protein expression in *Neisseria*. *Annu Rev Microbiol* 44: 451–477.
- Montgomery, D.L., Shiver, J.W., Leander, K.R., Perry, H.C., Friedman, A., Martinez, D., et al. (1993) Heterologous and homologous protection against influenza A by DNA vaccination: optimization of DNA vectors. *DNA Cell Biol* 12: 777–783.
- Moxon, E.R., and Wills, C. (1999) DNA microsatellites: agents of evolution? *Sci Am Jan*: 72–77.
- Philipp, W.J., Nair, S., Guglielmi, G., Lagranderie, M., Gicquel, B., and Cole, S.T. (1996) An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc Natl Acad Sci USA* 93: 3132–3137.
- Poulet, S., and Cole, S.T. (1995) Characterisation of the polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* 163: 87–95.
- Radman, M. (1999) Enzymes of evolutionary change. *Nature* 401: 866–869.
- Ramakrishnan, L., Federspiel, N.A., and Falkow, S. (2000) Granuloma-specific expression of mycobacterium virulence proteins from the glycine-rich PE-PGRS family. *Science* 288: 1436–1439.
- Robertson, B.D., and Meyer, T.F. (1992) Genetic variation in pathogenic bacteria. *Trends Genet* 8: 422–427.
- Rosenkrands, I., Weidigh, K., Jacobsen, S., Hansen, C.V., Florio, W., Gianetti, L., and Andersen, P. (2000) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* 21: 935–948.
- Ross, B.C., Raio, K., Jackson, K., and Dwyer, B. (1992) Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J Clin Microbiol* 30: 942–946.
- Saint-Joanis, B., Souchon, H., Wilming, M., Johnsson, K., Alzari, P.M., and Cole, S.T. (1999) Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from *Mycobacterium tuberculosis*. *Biochem J* 338: 753–760.
- Senaratne, R.H., Mobasher, H., Papavinasundaram, K.G., Jenner, P., Lea, E.J., and Draper, P. (1998) Expression of a gene for a porin-like protein of the OmpA family from *Mycobacterium tuberculosis* H37Rv. *J Bacteriol* 180: 3541–3547.
- Shiver, J.W., Perry, H.C., and Davies, M.E. (1995) Immune responses to HIV gp120 elicited by DNA vaccination. In *Vaccines*. Chanock, R.M., Brown, F., Ginsberg, H.S., and Norrby, E. (eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 95–98.
- Singh, K.K., Zhang, X., Patibandla, A.S., Chien, P., Jr, and Laai, S. (2001) Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. *Infect Immun* 69: 4185–4191.
- van Soolingen, D., de Haas, P.E.W., Hermans, P.W.M., Groenen, P.M.A., and van Embden, J.D.A. (1993) Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* 31: 1987–1995.
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., and Musser, J.M. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 94: 9869–9874.
- Stern, A., and Meyer, T.F. (1987) Common mechanism controlling phase and antigenic variation in pathogenic. *Mol Microbiol* 1: 5–12.
- Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell, B.G., and Cole, S.T. (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle Lung Dis* 79: 329–342.
- Towbin, H., Staehelin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* 76: 4350–4354.
- Ulmer, J.B., Donnelly, J.J., Parker, S.E., Rhodes, G.H., Feigner, P.L., Dwarki, V.J., et al. (1993) Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* 259: 1745–1749.
- Zhang, J.-R., and Norris, S.J. (1998) Genetic variation of the *Borrelia burgdorferi* gene *visE* involves cassette-specific, segmental gene conversion. *Infect Immun* 66: 3698–3704.

Outline

Introduction

The *Mycobacterium tuberculosis* complex

The genome sequence and biology of *M. tuberculosis* H37Rv

The genome sequence of *M. tuberculosis* CDC1551 and *M. bovis* AF2122/97

Initial comparison of the three genome sequences

Other methods for genome-wide comparisons

A summary of genomic variability

Variability associated with the PE and PPE multigene families

Clues from the genome for improved BCG vaccine development

Concluding remarks

INTRODUCTION

Genomics, the systematic analysis of the complete genetic material found in an organism by means of DNA sequencing and bioinformatics, is undoubtedly the most powerful tool now available to biological research. A massive return can be obtained for a single investment, since genomics not only provides the DNA sequence of an organism but also generates complete datasets of its genes, proteins, enzymes and antigens. Armed with this information the metabolic pathways, physiology and evolution can be reconstructed *in silico* and rich biological, evolutionary and medical insight obtained. At the time of writing, four complete genome sequences are available for pathogenic mycobacteria (two strains of *Mycobacterium tuberculosis*, one strain each of *Mycobacterium bovis* and *Mycobacterium leprae*) and several others are at different stages of completion. (For an overview see <http://www.pasteur.fr/unites/recherche/Lgmb/mycogenomics.html/>). In this chapter we will briefly describe the various approaches that have been employed, present some selected findings and show how this information can be used to improve diagnostic procedures.

THE *Mycobacterium tuberculosis* COMPLEX

The most predominant human pathogen is *M. tuberculosis*, the aetiologic agent of human tuberculosis, that accounts world-wide for eight million new infections and two million deaths per year [1]. Together with other highly related bacteria, *M. tuberculosis* forms a tightly-knit complex, a single species as defined by DNA/DNA hybridization studies [2], and characterised by a singular lack of genetic diversity [3]. The complex includes: *M. tuberculosis*, the causative agent in the vast majority of human tuberculosis cases; *M. canettii*, a smooth variant that is very rarely encountered; *M. africanum*, an agent of human tuberculosis in sub-Saharan Africa; *M. microti*, the agent of tuberculosis in voles; *M. bovis*, which infects a wide variety of mammalian species including humans, and BCG (bacille Calmette-Guérin), an attenuated derivative of *M. bovis*, used as a live vaccine against human tuberculosis.

The genome sequence and biology of *M. tuberculosis* H37Rv

For the genome project of the widely used reference strain *M. tuberculosis* H37Rv, a combined strategy consisting of sequencing selected cosmid and Bacterial Artificial Chromosome (BAC) clones, as well as whole genome shotgun sequencing, was employed. The minimally overlapping set of BAC clones containing large inserts of *M. tuberculosis* H37Rv DNA [4] was of critical importance for the timely completion of the *M. tuberculosis* H37Rv genome sequence, as it allowed the extremely G+C rich areas of the genome, corresponding to the PE-PGRS genes (discussed further below), to be obtained as these were under-represented in the small insert shotgun libraries. The complete genome sequence of *M. tuberculosis* H37Rv comprises 4,411,532 bp and has an average G+C content of 65.6 %. As the findings of the analysis have been described extensively elsewhere [5-8], only a brief description of selected features of the sequence will be presented here. The genome contains ~4,000 genes distributed fairly evenly between the two strands. Over 51% of the genes have arisen as a result of gene duplication or domain shuffling events, and 3.4% of the genome is composed of insertion sequences (IS) and phages (phiRv1, phiRv2). There are 56 copies of IS elements belonging to the well-known IS3, IS5, IS21, IS30, IS110, IS256, and ISL3 families, as well as a new IS family, IS1535, that appears to employ a frameshifting mechanism to

produce its transposase [9]. *IS6110*, a member of the IS3 family, is the most abundant element and has played an important role in shaping the genome.

The information deduced from the genome sequence provided new and valuable insight into the biology of the tubercle bacilli, and highlighted the importance of lipid metabolism to its lifestyle as at least 8% of the genome is dedicated to this activity. While *M. tuberculosis* was known to contain a remarkable array of lipids, glycolipids, lipoglycans and polyketides [10] and the genome sequence revealed many of the genes required for their production, it was surprising to note the presence of numerous genes and proteins that could confer lipolytic functions. Estimates of the concentrations of potential substrates available to a pathogen in host tissues suggest that lipids and sterols are more abundant than carbohydrates. While *M. tuberculosis* has the prototype β -oxidation cycle required for lipid catabolism, catalysed by the multifunctional FadA/FadB proteins, it also appears to have ~100 enzymes potentially involved in alternative lipid oxidation pathways in which exogenous lipids from host cells could be degraded.

One of the major findings of the *M. tuberculosis* genome project was the identification of large gene families which were either unknown previously or poorly understood. Foremost among these were the novel PE and PPE families comprising 100 and 67 members, respectively [5, 11]. Members of each family share a conserved N-terminal domain of ~110 and 180 amino acid residues, with the characteristic motifs ProGlu (PE in single letter code), or ProProGlu (PPE) at positions 8-9, or 8-10, respectively. The N-terminal domain is generally followed by a C-terminal extension that is often of highly repetitive sequence, and of particular interest are the PE proteins belonging to the PGRS (polymorphic GC-rich sequence) subclass [12] as nearly half of their amino acid content is composed of glycine and this occurs in tandem repetitions of the motif AsnGlyGlyAlaGlyGlyAla, or variants thereof. As there has been extensive recent progress in the characterisation of these proteins, and their genes represent a major source of diversity, they will be discussed in detail in a subsequent section.

The genome sequence of *M. tuberculosis* CDC1551 and *M. bovis* AF2122/97

The CDC1551 strain of *M. tuberculosis* is of interest because it appears to be highly transmissible, causing widespread skin-test conversion in a rural area of the United States [13]. Contrary to what was first thought, CDC1551 is less virulent in animal models of the disease than the H37Rv strain of *M. tuberculosis* but it does induce a more rapid and robust immunological response by the host [14]. Its genome sequence was obtained by the whole genome shotgun method and is nearly identical to that of *M. tuberculosis* H37Rv but is slightly smaller at 4,403,836 bp (Table 1).

The *M. bovis* strain AF2122/97 was responsible for numerous cases of tuberculosis in cattle, in the county of Devon in the United Kingdom in the 1990s, and has also been isolated from diseased badgers, and other wildlife, in the same area [15]. No association with human tuberculosis has yet been reported. The genome sequence of *M. bovis* AF2122/97 has also been obtained through a combined approach involving the sequencing of whole genome shotgun clones and selected BACs. This genome is nearly 70 kb smaller than those of the two *M. tuberculosis* isolates (Table 1) and contains about 80 genes less. Of particular interest was the presence of a stretch of additional DNA in *M. bovis* AF2122/97, known as TbD1, as this has helped to understand the evolution of the *M. tuberculosis* complex by providing a valuable diagnostic marker [16]. The TbD1 region of *M. bovis* (and all other tubercle bacilli except modern *M. tuberculosis*), contains two genes *mmpS6* and *mmpL6* (Fig. 1), that may be involved in the transport of complex (glyco)lipids by analogy with *MmpL7*, which is required for the translocation of phthiocerol-dimycocerosate across the cytoplasmic membrane [17, 18].

Initial comparison of the three genome sequences

The pioneering work of Musser and his colleagues on the diversity and population genetics of the *M. tuberculosis* complex indicated that single nucleotide polymorphisms (SNP) were unusually rare and that synonymous substitutions occurred infrequently. This led these workers to suggest that the tubercle bacillus may have become a human pathogen as recently as 15,000 years ago and that, having adopted this highly specialised niche, it encountered an evolutionary bottleneck that restricted its divergence [3, 19].

However, the lack of diversity is somewhat surprising given the absence of the MutL-S mis-match repair system from the proteome [20].

With the availability of three complete genome sequences, finer comparisons of nucleotide sequence divergence can be undertaken compared to those of Sreevatsan *et al.* that targeted 26 loci [3]. The initial analysis, presented in Table 1, reveals that SNP are very uncommon and account for less than 0.02% of the differences between the *M. tuberculosis* strains H37Rv and CDC1551, or 0.05% of the differences between *M. bovis* and the two *M. tuberculosis* strains. This confirms and extends the previous findings [3]. In contrast, the major source of diversity may be found in the form of insertions or deletions (InDels), ranging in size from 2 bp to 13 kb (Table 1). Ten of these result from IS6110 copy number differences whereas others correspond to some of the well-characterised "regions of difference" (RD) described further below. On comparison of the two *M. tuberculosis* strains, 55% of these InDels were found to affect genes, especially those encoding PE-PGRS proteins, while the remainder were intergenic. Similar findings and values were obtained when the *M. bovis* genome was examined. Among the intergenic sequences that display variation are the mycobacterial interspersed repetitive units or MIRUs [21, 22]. There are 41 loci in *M. tuberculosis* H37Rv that contain MIRUs and these often occur as tandem repeats, situated between genes in operons. Variability in MIRU copy number has been observed at certain loci and this has been exploited to develop an epidemiological test to distinguish between strains [23].

OTHER METHODS FOR GENOME-WIDE COMPARISONS

One of the first methods that enabled whole genome comparisons was pulsed-field gel electrophoresis (PFGE) of macro-restriction fragments of chromosomal DNA. PFGE has been used as a tool for molecular epidemiologic and population genetic studies of the *M. tuberculosis* complex [24, 25] and for the construction of integrated genome maps of *M. tuberculosis* H37Rv and *M. bovis* BCG Pasteur [26, 27]. These maps were of the utmost importance for estimating genome size (~4.4 Mb) and structure (circular chromosome, absence of plasmids), for establishing ordered cosmid and BAC libraries [4, 27] and for guiding and controlling assembly of the *M. tuberculosis* H37Rv genome sequence [5].

Subtractive hybridization, which is based on the hybridization of DNA from two genomes and removal of any common sequences [28], is another powerful technique for whole genome comparisons and was applied to the genomes of *M. bovis* and *M. bovis* BCG. Three genomic regions of difference (RD1- RD3) were identified that were absent from *M. bovis* BCG relative to *M. bovis* [29]. This initial comparison of the BCG vaccine strain to virulent strains of the *M. tuberculosis* complex has recently been extended by the use of genome hybridization arrays based on a minimal set of BAC clones [4, 30], microarrays employing spotted PCR products [31] or oligonucleotide-based Affymetrix GeneChip technology [32, 33]. These combined approaches identified 18 RD regions that are absent from *M. bovis* BCG but present in H37Rv and other *M. tuberculosis* strains. Similar array analyses have been applied to clinical isolates and additional variable regions identified, some of these occur on multiple occasions whereas others are confined to a particular strain.

In silico comparison of genome sequences from related mycobacterial species represents an alternative strategy to the various hybridization and array techniques. With the growing number of completed genome sequences, and advances in bio-informatics, more refined comparisons of sequence variation among strains become possible such as those outlined above in which TbD1 was initially identified. More details of the techniques used in comparative mycobacterial genomics may be found in a recent review together with the principal findings [34].

A SUMMARY OF GENOMIC VARIABILITY

Genetic variability can arise as a result of mutations involving single nucleotide changes or from more complex events such as the insertion or deletion of stretches of nucleotides within genes. The transposition of IS elements can also lead to the inactivation or loss of genes depending on the site of insertion or even result in up-regulation if the transposon contains an outward pointing promoter and inserts upstream of a gene.

Although SNP occur in the genomes of members of the *M. tuberculosis* complex (Table 1) at the relatively low level for a bacterium [3] of 1 in every 2 - 4,000 bp, depending on

the species, InDels appear to be a more common means of generating diversity. Most of the insertions result from transposition events, generally involving IS6110, or more rarely from gene duplication. There is no evidence in favor of recent horizontal gene transfer and the closest example of this is provided by RD3 or RD11, the presence or absence of the prophage genomes, phiRv1 or phiRv2, respectively [8]. The deletions fall into two groups, ancient and recent. The ancient deletions occurred at different stages in the speciation process and are widespread whereas the recent deletions have a more restricted distribution. Examples of the latter are the IS6110-mediated deletion of the 7 kb locus RvD2 in *M. tuberculosis* H37Rv, still present in the closely related avirulent derivative H37Ra [35] that also undergoes great variability in clinical isolates of *M. tuberculosis* [36]. Other IS6110-mediated deletions are represented by RvD3-RvD5 that have removed three, two and four genes from *M. tuberculosis* H37Rv. An example of a recent deletion in BCG Pasteur is the loss of the 8 kb locus RD14, which is still present in all the other BCG substrains [31].

In contrast to these deletion events, the absence of regions RD7, RD8, RD9, and RD10 from *M. microti*, *M. bovis* and BCG seems to be a much older event in evolutionary terms. From close inspection of the sequences that border these RD-regions it is apparent that deletions occurred within coding regions. Genes that are present in *M. tuberculosis* in full length were found to be disrupted in BCG, *M. bovis* and *M. microti* at exactly the same location, whereas these genes are still intact in *M. tuberculosis* and *M. canettii* strains. This observation indicates that these RD regions result from real deletions and not the insertion of genes into *M. tuberculosis* as might be speculated. Based on the presence or absence of such conserved RD regions, a degree of relatedness to the last common ancestor of the *M. tuberculosis* complex was proposed that shows that the lineages of *M. tuberculosis* and *M. bovis* separated before the *M. tuberculosis* specific deletion TbD1 occurred (Fig. 2). This analysis has identified some evolutionarily "old" *M. canettii*, *M. tuberculosis* and *M. africanum* strains, most of them of African origin, as well as "modern" *M. tuberculosis* strains, the latter including representatives from major epidemic clusters like Beijing, Haarlem and Africa [16].

Some of these regions, primarily RD9 and TbD1 but also RD1, RD2, RD4, RD7, RD8, RD10, RD12 and RD13 represent very interesting candidates for the development of powerful diagnostic tools for the rapid and unambiguous identification of members of the *M. tuberculosis* complex [16]. Figure 2 presents a differential scheme for identifying individual species that relies on the presence of these markers in association with selected SNP such as *mmpL6* 551AAC→AAG. This diagnostic strategy offers great promise to studying the epidemiology and evolutionary biology of the tubercle bacilli. Its application has already shaken one of the tenets of our understanding of the origin of human tuberculosis, namely that *M. tuberculosis*, the etiological agent of human tuberculosis evolved from *M. bovis*, the agent of bovine disease. From Figure 2 it is clear that this cannot have been the case but instead that *M. canettii* and ancestral *M. tuberculosis* strains appear to be direct descendants of tubercle bacilli that existed before the *M. africanum*→*M. bovis* lineage separated from the *M. tuberculosis* lineage.

VARIABILITY ASSOCIATED WITH THE PE AND PPE MULTIGENE FAMILIES

Genome comparisons and functional genomics have shed new light on the possible roles of the PE and PPE proteins. When the PE genes of *M. tuberculosis* strains H37Rv and CDC1551 were compared *in silico* it was found that the genes encoding a PE domain or a PE domain followed by a unique protein sequence were identical in both cases [37, 38]. In contrast, 39 of the 62 common PE-PGRS proteins displayed variability as a result of frameshift mutations or the in-frame insertion or deletion of different Ala, Gly-rich coding sequences in the PGRS component of the gene (Fig. 3). Furthermore, consistent with this finding, size variation was also seen on Western blot analysis of protein samples, prepared from different clinical isolates, using PE-PGRS specific antibodies. As expected from the conserved repetitive structure, the antibodies cross-reacted with more than one PE-PGRS protein suggesting that different proteins share common epitopes and an example of this and variability between H37Rv and CDC1551 proteins may be seen in Figure 4.

Subcellular fractionation studies and immuno-gold- or fluorescent antibody-staining localized many PE-PGRS proteins in the cell wall and cell membrane of *M. tuberculosis* [37, 39]. Disruption of the gene encoding the PE-PGRS protein Rv1818c in *M. tuberculosis* led to greatly reduced bacterial clumping, suggesting that this protein may mediate cell-cell adhesion and altered phagocytosis by macrophages [39]. Another PE-PGRS protein, Rv1759c, is capable of binding fibronectin and could thus mediate bacterial attachment to host cells [40, 41]. In addition, members of the PE-PGRS families have been implicated in the pathogenesis of *M. marinum* [42], where at least two genes were shown to be upregulated strongly following phagocytosis of the bacterium.

The immunogenicity of the PE-PGRS protein, Rv1818c, has been studied in great detail in mice [43] where immunisation with the PE domain induced Th1-type responses that were not found when the complete PE-PGRS protein was used. Instead, the PGRS part of the protein elicits antibody production and suppresses the Th1-response induced by the PE domain. The PE-PGRS proteins bear some sequence similarity to EBNA, the Epstein-Barr virus nuclear antigens, that block antigen presentation by the MHC class I pathway, through their action as proteasome inhibitors [5]. It was speculated that PE-PGRS proteins may also have this activity and it has recently been shown that the PGRS domain, when fused to GFP, confers increased resistance to proteosomal attack [44]. If these immunological and adhesive properties are shared among other members of the family, it is conceivable that the extensive variation observed at the gene level could bestow very different phenotypes on the different strains.

The PPE proteins also fall into subfamilies and by far the most spectacular of these is the MPTR, or major polymorphic tandem repeat subfamily, which contains multiple repetitions of the motif AsnXGlyXGlyAsnXGly. The largest PPE-MPTR protein detected contains 3300 amino acid residues and, like the PE-PGRS proteins, these are also associated with polymorphic DNA sequences [11]. Extensive sequence variation, due to InDels, has been reported between the PPE proteins Rv1917c and Rv1753c of *M. tuberculosis* and *M. bovis* [15]. There is little evidence concerning their possible

function but Rv1917c was recently shown to be cell wall associated and surface exposed [45].

CLUES FROM THE GENOME FOR IMPROVED BCG VACCINE DEVELOPMENT

The only tuberculosis vaccine is BCG [46], one of the six components of the Expanded Program of Immunisation of WHO. Since 1921, more than three billion doses have been administered worldwide with negligible side effects in immuno-competent individuals. BCG is highly effective against disseminated forms of tuberculosis in children [47], such as miliary disease and meningitis, and also was shown to protect against leprosy in a recent controlled trial [48]. However, the protection conferred by BCG vaccination against pulmonary TB in adults, representing the majority of the disease burden, is highly variable and ranges from 0 to 80 %, depending on the population, the country and the BCG strain used [49].

BCG vaccination will probably continue until controlled long-term vaccine trials prove that new vaccines offer better protection. Consequently, molecular and immunological characterisation of the presently used BCG vaccine strains continues to be an important research topic as it will allow us to identify the genetic basis for the attenuation of BCG and to find the reasons why some BCG substrains offer better protection or cause more scarring than others [50]. Encouragingly, it has been found recently that engineering BCG to overproduce a 30-kDa major secretory protein, known as antigen 85-B, results in more effective protection against tuberculosis in guinea pigs [51].

By elucidating the genomic differences between the virulent and avirulent members of the *M. tuberculosis* complex, comparative genomics has the potential to define mycobacterial virulence determinants specific for humans and other host-range determining factors that may have important impact on vaccine design. In addition to the many RD regions described above, two large duplications (DU1 and DU2) have been found in BCG strains. Some important potentially protective antigens may have been lost from BCG during the attenuation process. For example, ESAT-6 and CFP10, two

strongly immunogenic proteins, are found in early culture filtrates of *M. tuberculosis* but not in BCG [52, 53]. Interestingly, the ESAT-6 and CFP10 genes are located in the RD1 region of *M. tuberculosis*, a 10 kb genomic region which is absent from all BCG strains. One attractive approach is to use recombinant BCG carrying the missing genes [54] to determine whether better protection can be obtained.

Besides loss of genes, gene duplication undoubtedly played an important role in the differentiation of BCG. Evidence that this phenomenon is still ongoing can be seen in the presence of two large tandem duplications (DU1, DU2) in BCG Pasteur, detected by comparative genomics [55]. These duplications of 29 and 36 kb seem to have appeared independently from each other as their presence and/or their size may vary in the different BCG substrains. While DU1 appears to be confined to BCG Pasteur, DU2 has been detected in all BCG substrains, tested so far but this duplication can exhibit size variation (our unpublished observations). Gene dosage effects may result from these duplications thereby accounting for some of the differences in immunogenicity displayed by the various BCG vaccine strains.

CONCLUDING REMARKS

Genomic analysis of *M. tuberculosis* has provided us with a rich source of information about the pathogen that is of inestimable value. Application of this new knowledge will surely result in better understanding of the molecular basis of pathogenesis and, hopefully, translate into new drugs, diagnostic tests and vaccines to combat the disease. Comparative genomics of the *M. tuberculosis* complex has revealed that, on the one hand, the majority of the genome is well conserved in sequence and highly stable, while on the other, extensive variation may occur as a result of a limited range of InDels. Some of these, especially the more numerous deletion events, may affect features such as host-range and persistence while others may have more subtle effects. Of particular interest, and some concern, is the finding that the major source of diversity between strains is due to variation in the sequences of the PE and PPE proteins. If these proteins serve as variable surface antigens, as seems increasingly likely, and affect interactions with host cells and phagocytes, development of a more effective vaccine will be most challenging.

Acknowledgements

We thank T. Garnier and S.V. Gordon for helpful discussions, and B. Saint-Joanis for technical assistance. Parts of this work were supported by grants from the European Union (QLRT-2000-02018), the Institut Pasteur (PTR35) and the Association Française Raoul Follereau.

FIGURE LEGENDS.

Fig. 1. The lineage of modern *M. tuberculosis* strains is characterized by a specific deletion (TbD1), resulting in an in-frame gene fusion of *mmpS6'* and '*mmpL61*' (lower portion of the figure) whereas all other members of the *M. tuberculosis* complex harbor *mmpS6* and *mmpL6* at this locus (upper portion of the figure). All six possible reading frames are shown. Vertical bars indicate stop codons.

Fig. 2. Scheme of the evolutionary pathway of the tubercle bacilli. Boxes indicate deletions (grey) and point mutations (white) that appeared in the various lineages during the evolution of the tubercle bacilli [16].

Fig. 3. Example of an in-frame 9 codon deletion in a orthologous PE-PGRS gene in *M. bovis* BCG corresponding to gene Rv3507 from *M. tuberculosis* H37Rv. The sequence which is shown originates from the complementary strain.

Fig. 4. Sub-cellular localization of PE-PGRS proteins in two strains of *M. tuberculosis*. Cell wall, cell membrane and cytosol fractions were prepared and proteins separated by SDS-PAGE. After transfer to nylon membranes, antibody 1441c was used to locate the PE-PGRS protein in the cell. The antibody recognized several proteins in each cell fraction. The molecular weight markers (in kDa) are indicated on the left. The white arrow indicates size-variation of PE-proteins that hybridize to antibody 1441 in the two tested *M. tuberculosis* strains.

REFERENCES

1. Dye C, Sheele S, Dolin P, et al. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. *JAMA*, 1999;282:677-686.
2. Imaeda T. Deoxyribonucleic acid relatedness among selected strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium micoti* and *Mycobacterium africanum*. *Int. J. Syst. Bact.*, 1985;35:147-150.
3. Sreevatsan S, Pan X, Stockbauer KE, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences USA*, 1997;94:9869-9874.
4. Brosch R, Gordon SV, Billault A, et al. Use of a *Mycobacterium tuberculosis* H37Rv Bacterial Artificial Chromosome (BAC) library for genome mapping, sequencing and comparative genomics. *Infection Immun.*, 1998;66:2221-2229.
5. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998;393:537-544.
6. Cole ST. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Letters*, 1999;452:7-10.
7. Tekaia F, Gordon SV, Garnier T, et al. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle Lung Disease*, 1999;79:329-342.
8. Brosch R, Gordon SV, Eiglmeier K, et al. Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. In: Hatfull GF and Jacobs WR, Jr. Molecular genetics of mycobacteria. Washington, D.C.: ASM Press, 2000; 19-36.
9. Gordon SV, Heym B, Parkhill J, et al. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, 1999;145:881-892.
10. Daffe M, Draper P. The envelope layers of mycobacteria with reference to their pathogenicity. *Advances in Microbial Physiology*, 1998;39:131-203.

11. Cole ST, Barrell BG. Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Genetics and tuberculosis (Novartis Foundation Symposium 217)*, 1998;160-172.
12. Poulet S, Cole ST. Characterisation of the polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch. Microbiol.*, 1995;163:87-95.
13. Valway SE, Sanchez MP, Shinnick TF, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *New England Journal of Medicine*, 1998;338:633-639.
14. Manca C, Tsenova L, Barry Cr, et al. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *Journal of Immunology*, 1999;162:6740-6746.
15. Gordon SV, Eiglmeier K, Garnier T, et al. Genomics of *Mycobacterium bovis*. *Tuberculosis (Edinb)*, 2001;81:157-163.
16. Brosch R, Gordon SV, Marmiesse M, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*, 2002;99:3684-3689.
17. Camacho LR, Ensergueix D, Perez E, et al. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Molecular Microbiology*, 1999;34:257-267.
18. Cox JS, Chen B, McNeil M, et al. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature*, 1999;402:79-83.
19. Kapur V, Whittam TS, Musser J. Is *Mycobacterium tuberculosis* 15,000 years old? *J. Inf. Dis.*, 1994;170:1348-1349.
20. Mizrahi V, Andersen SJ. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Molecular Microbiology*, 1998;29:1331-1339.
21. Supply P, Mazars E, Lesjean S, et al. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Molecular Microbiology*, 2000;36:762-771.
22. Supply P, Magdalena J, Himpens S, et al. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Molecular Microbiology*, 1997;26:991-1003.
23. Mazars E, Lesjean S, Banuls AL, et al. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A*, 2001;98:1901-1906.

24. Singh SP, Salamon H, Lahti CJ, et al. Use of pulsed-field gel electrophoresis for molecular epidemiologic and population genetic studies of *Mycobacterium tuberculosis*. *J Clin Microbiol.*, 1999; 37:1927-1931.
25. Zhang Y, Wallace RJ, Jr. Genetic differences between BCG substrains. *Tubercul Lung Dis*, 1995;76:43-50.
26. Philipp WJ, Nair S, Guglielmi G, et al. Physical mapping of *Mycobacterium bovis* BCG Pasteur reveals differences from the genome map of *Mycobacterium tuberculosis* H37Rv and from *Mycobacterium bovis*. *Microbiology*, 1996;142:3135-3145.
27. Philipp WJ, Poulet S, Eiglmeier K, et al. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proceedings of the National Academy of Sciences USA*, 1996;93:3132-3137.
28. Straus D, Ausubel FM. Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc Natl Acad Sci U S A*, 1990;87:1889-1893.
29. Mahairas GG, Sabo PJ, Hickey MJ, et al. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *Journal of Bacteriology*, 1996;178:1274-1282.
30. Gordon SV, Brosch R, Billault A, et al. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Molec Microbiol*, 1999;32:643-656.
31. Behr MA, Wilson MA, Gill WP, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarrays. *Science*, 1999;284:1520-1523.
32. Salamon H, Kato-Maeda M, Small PM, et al. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. *Genome Res*, 2000;10:2044-2054.
33. Kato-Maeda M, Rhee JT, Gingeras TR, et al. Comparing Genomes within the Species *Mycobacterium tuberculosis*. *Genome Res*, 2001;11:547-554.
34. Brosch R, Pym AS, Gordon SV, et al. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol*, 2001;9:452-458.

35. Brosch R, Philipp W, Stavropoulos E, et al. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra. *Infection and immunity*, 1999;67:5768-5774.
36. Ho TBL, Robertson BD, Taylor GM, et al. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast (Comparative and Functional Genomics)*, 2001;17:272-282.
37. Banu S, Honoré N, Saint-Joanis B, et al. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol*, 2002;in press:
38. Betts JC, Dodson P, Quan S, et al. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. *Microbiology*, 2000;146:3205-3216.
39. Brennan MJ, Delogu G, Chen Y, et al. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun*, 2001;69:7326-7333.
40. Espitia C, Laclette JP, Mondragon-Palomino M, et al. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology*, 1999;145:3487-3495.
41. Singh KK, Zhang X, Patibandla AS, et al. Antigens of *Mycobacterium tuberculosis* Expressed during Preclinical Tuberculosis: Serological Immunodominance of Proteins with Repetitive Amino Acid Sequences. *Infect Immun*, 2001;69:4185-4191.
42. Ramakrishnan L, Federspiel NA, Falkow S. Granuloma-specific expression of mycobacterium virulence proteins from the glycine-rich PE-PGRS family. *Science*, 2000;288:1436-1439.
43. Delogu G, Brennan MJ. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun*, 2001;69:5606-5611.
44. Brennan MJ, Delogu G. The PE multigene family: A molecular mantra for mycobacteria. *Trends Microbiol*, 2002;in press:
45. Sampson SL, Lukey P, Warren RM, et al. Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. *Tuberculosis (Edinb)*, 2001;81:305-317.
46. Calmette A. La vaccination preventive contre la tuberculose. 1927;250.

47. Hart PD, Sutherland I. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *Br Med J*, 1977;2 (6082):293-295.
48. Ponnighaus JM, Fine PE, Sterne JA, et al. Efficacy of BCG vaccine against leprosy and tuberculosis in northern Malawi. *Lancet*, 1992;339:636-639.
49. Tuberculosis Prevention Trial M. Trial of BCG vaccines in south India for tuberculosis prevention. *Indian J. Med. Res.*, 1980;72(Suppl):1-74.
50. Lagranderie MR, Balazuc AM, Deriaud E, et al. Comparison of immune responses of mice immunized with five different *Mycobacterium bovis* BCG vaccine strains. *Infect Immun*, 1996;64:1-9.
51. Horwitz MA, Harth G, Dillon BJ, et al. Recombinant bacillus calmette-guerin (BCG) vaccines expressing the *Mycobacterium tuberculosis* 30-kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model. *Proc Natl Acad Sci U S A*, 2000;97:13853-13858.
52. Brandt L. Key epitopes on the ESAT-6 antigen recognized in mice during the recall of protective immunity to *Mycobacterium tuberculosis*. *J.Immunol.*, 1996;157:3527-3533.
53. Brandt L, Elhay M, Rosenkrands I, et al. ESAT-6 subunit vaccination against *Mycobacterium tuberculosis*. *Infection and immunity*, 2000;68:791-795.
54. Stover CK, V.F. dIC, G.P. B, et al. Use of recombinant BCG as a vaccine delivery vehicle. *Adv Exp Med Biol*, 1992;327:175-182.
55. Brosch R, Gordon SV, Buchrieser C, et al. Comparative genomics uncovers tandem chromosomal duplications in some strains of *Mycobacterium bovis* BCG: Implications for vaccination. *Comparative and Functional Genomics (Yeast)*, 2000;17:111-123.

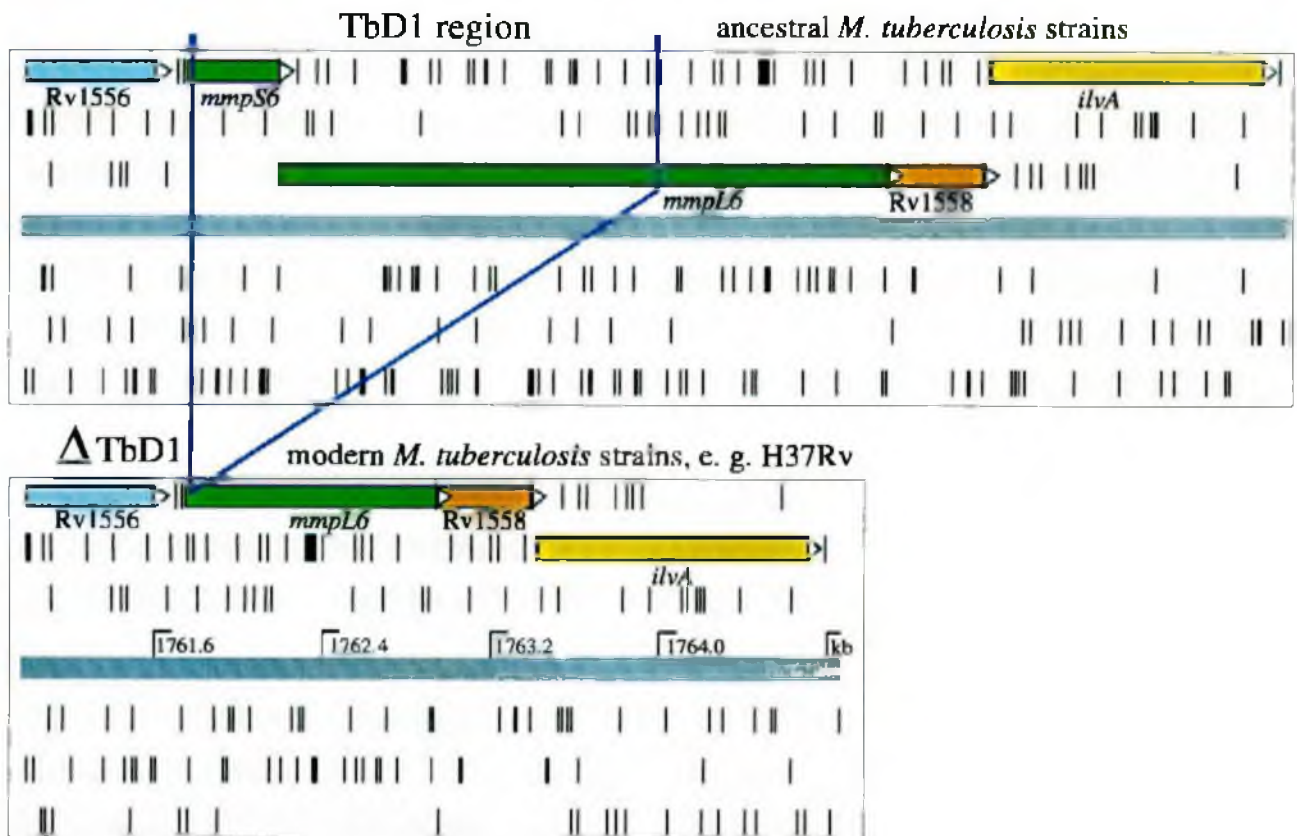


Figure 1

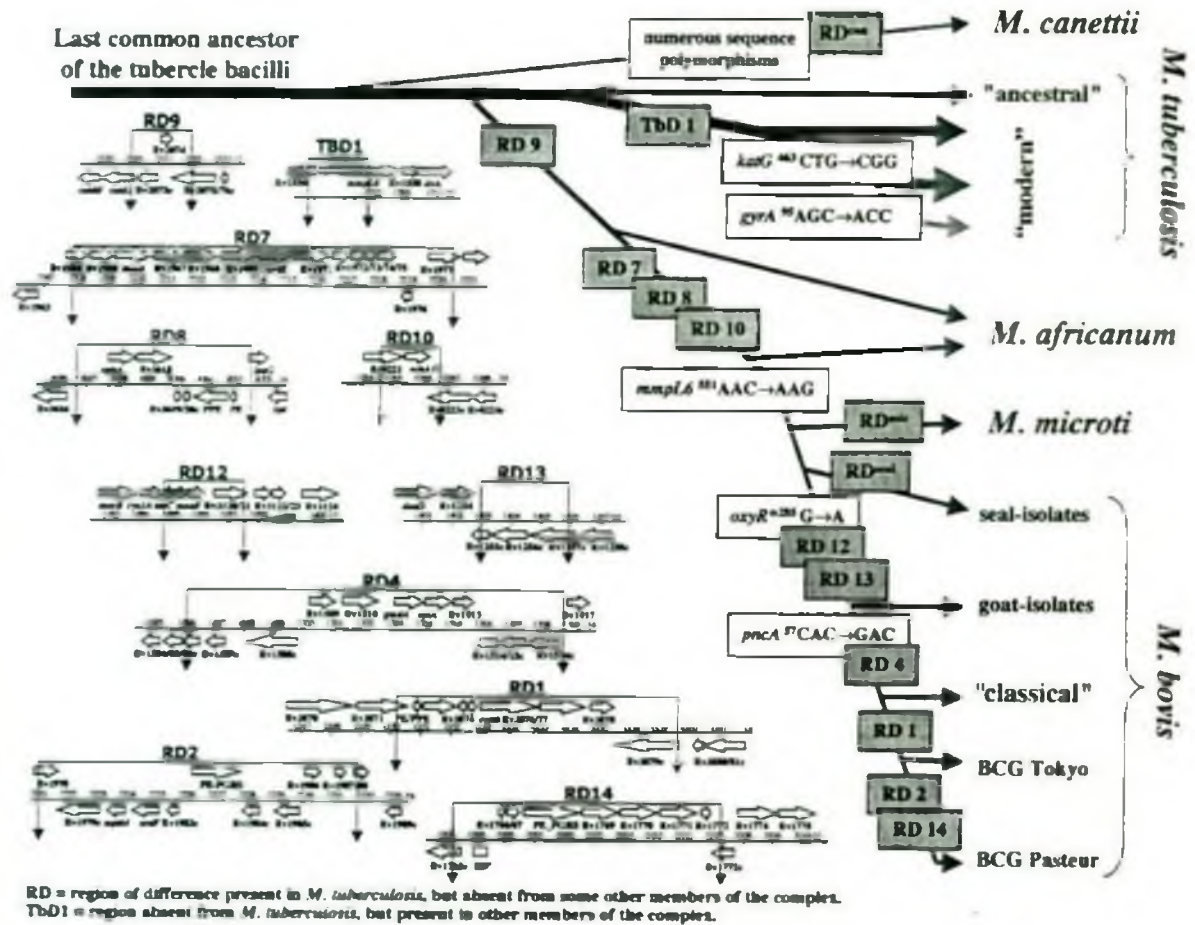


Figure 2

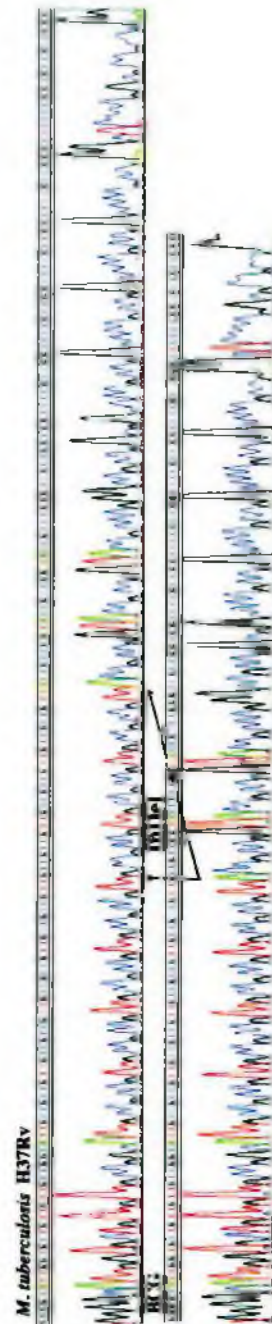


Figure 3

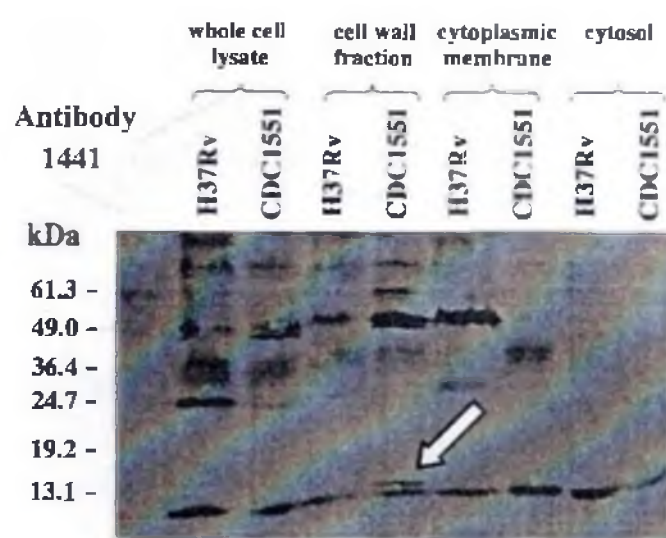


Figure 4