Molecular Microbiology (2002) **44**(1), 9–19

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

Sayera Banu,[1,2] Nadine Honoré,[1] Brigitte Saint-Joanis,[1] Dana Philpott[3], Marie-Christine Prévost[4] and Stewart T. Cole[1]*

[1]*Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.*
[2]*International Centre for Diarrhoeal Disease Research, Bangladesh, GPO-128, Dhaka-1000, Bangladesh.*
[3]*Unité de Pathogenie Microbienne Moléculaire, and*
[4]*Plateforme de Micoscopie 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. gordonae*) 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.**

Accepted 20 November, 2001. *For correspondence. E-mail stcole@pasteur.fr; Tel. (+33) 1 45 68 84 46; Fax (+33) 1 40 61 35 83.

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-

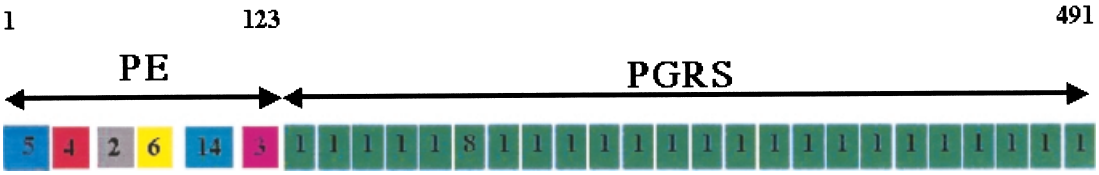

Fig. 1. Domain organization of PE-PGRS protein Rv1441. The figure is not to scale, and the most significant amino acid motifs, detected using the MEME and MAST algorithms (Bailey and Elkan, 1994; Bailey and Gribskov, 1998; Tekaia *et al*., 1999), have the following consensus sequence: 1, GNGGAGG; 2, YQALSAQAAAFH; 3, INAPTQALLGRPLI; 4, DEVSAAIAALF; 5, MSFVVAAPEMLAAAAADL; 6, QFVQALTAAAGAYAAAE; 8, GGWLYGNG; 14, SPLQTMQQ.

logical importance and could even represent the principal source of antigenic variation. It is known that one PE-PGRS member, Rv1759c, a fibronectin-binding protein of 55 kDa produced during infection and disease, elicits a variable antibody response, suggesting that either individuals mount different immune responses or this PE-PGRS protein may not be produced by all strains of *M. tuberculosis* (Abou-Zeid *et al*., 1991; Espitia *et al*., 1999). Indirect support in favour of the latter is provided by restriction fragment length polymorphism (RFLP) for various PE-PGRS genes in clinical isolates (Cole and Barrell, 1998), whereas direct evidence of allelic variation in the PE-PGRS protein encoded by Rv0746 has also been reported between *M. tuberculosis* H37Rv and *Mycobacterium bovis* BCG Pasteur (Cole *et al*., 1998).

Despite their numerical abundance, no PE proteins have yet been identified by biochemical means including extensive analysis of the *M. tuberculosis* proteome by means of two-dimensional electrophoresis and mass spectrometry (Jungblut *et al*., 1999; Rosenkrands *et al*., 2000). However, as pointed out by Betts *et al*. (2000), this may result from the relative paucity of trypsin cleavage sites in these acidic proteins. Other explanations, such as localization in cell membranes, are possible, as proteins from this cellular compartment are notoriously refractory to analysis by two-dimensional electrophoresis. In this paper, we present the findings of a study designed to establish the subcellular location and to investigate the antigenicity and biological functions of the PE proteins.

Results

PE-PGRS gene expression

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 1) with the aim of monitoring expression at the transcriptional level in *in vitro*-cultured *M. tuberculosis* cells by reverse transcription–polymerase chain reaction (RT–PCR) (Fig. 2). Smaller genes were picked in order to make PCR

Table 1. Selected PE-PGRS genes and primers used in RT–PCR.

				Primers used for RT–PCR[b]	
Gene[a]	Size (kb)	Mol. wt (Da)	pI	Name of primer	Sequence
Rv0980c	1.3	42 059	4.1	RT-Rv0980cL	5′-GGCGACCACCCAGGTGTTG-3′
				RT-Rv0980R	5′-GGTCCCGTCGATCCCGTTC-3′
Rv1067c	2.0	54 683	4.9	RT-Rv1067cL	5′-TGTTGGTGTCTCCGTCGCAG-3′
				RT-Rv1067cR	5′-TGGTCGACATTGGTCGCCTC-3′
Rv1068c	1.3	39 294	4.8	RT-Rv1068cL	5′-GTGCCAGACATGCTCAGCTC-3′
				RT-Rv1068cR	5′-GCACCTGCTGTTCGACGTTG-3′
Rv1243c	1.6	47 270	3.6	RT-Rv1243L	5′-GGTGCCGACGAGGTATCAG-3′
				RT-Rv1243R	5′-TGCCGTACAACAAGCCACCG-3′
Rv1325c	1.8	49 565	3.9	RT-Rv1325L	5′-CGGATTTGGCGAACATCGGG-3′
				RT-Rv1325R	5′-GGGTGGGCGTGTTCACAATG-3′
Rv1396c	1.7	47 890	5.4	RT-Rv1396L	5′-CGGATCTGGCAAGCATCGG-3′
				RT-Rv1396R	5′-AGCTGCTGCTCGACGTTGG-3′
Rv1441c	1.4	40 797	3.9	RT-Rv1441cL	5′-GGAATGTTGTCAGCGGCGG-3′
				RT-Rv1441cR	5′-GCGAGCACCTGTTGTTCGAG-3′
Rv1818c	1.4	40 743	3.9	RT-Rv1881L	5′-GCGATGGCGGCATTGTTCTC-3′
				RT-Rv1881R	5′-GGCTCCGTTACCGATCAGTG-3′
Rv1983	1.6	53 692	3.9	RT-Rv1983L	5′-CCGGCATGAGTGGTGGAG-3′
				RT-Rv1983R	5′-GCGAAGATGTAGTACAGCCC-3′
Rv2162c	1.5	44 464	4.3	RT-Rv2162L	5′-TCTCGGCGTTGTTGGGGTC-3′
				RT-Rv2162R	5′-CGTTGGGATTGGCGGTGTC-3′

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

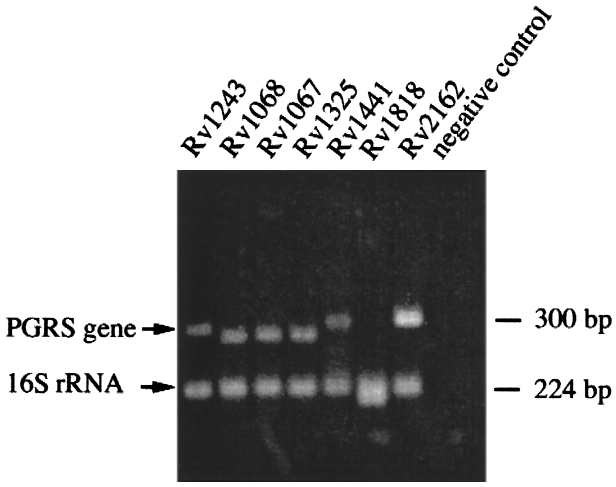

Fig. 2. Duplex RT–PCR analysis of total RNA extracted from *Mycobacterium tuberculosis* H37Rv. Primers were designed to amplify between 200 and 300 bp segments of 10 different PE-PGRS genes. A 224 bp segment of 16S rRNA of *M. tuberculosis* was amplified in the same PCR as the internal control. Seven 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.

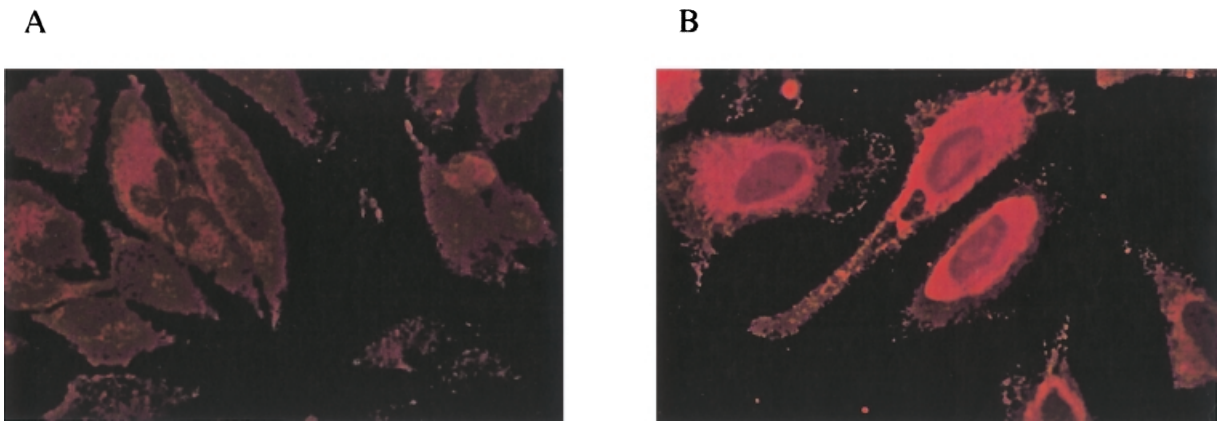

Fig. 3. Expression of Rv1441c in HeLa epithelial cells. HeLa cells were transfected with the vector construct alone (A) or the Rv1441c construct (B). Forty-eight hours after transfection, cells were fixed and stained with the Rv1441c-specific antibody and visualized by fluorescence microscopy. Strong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

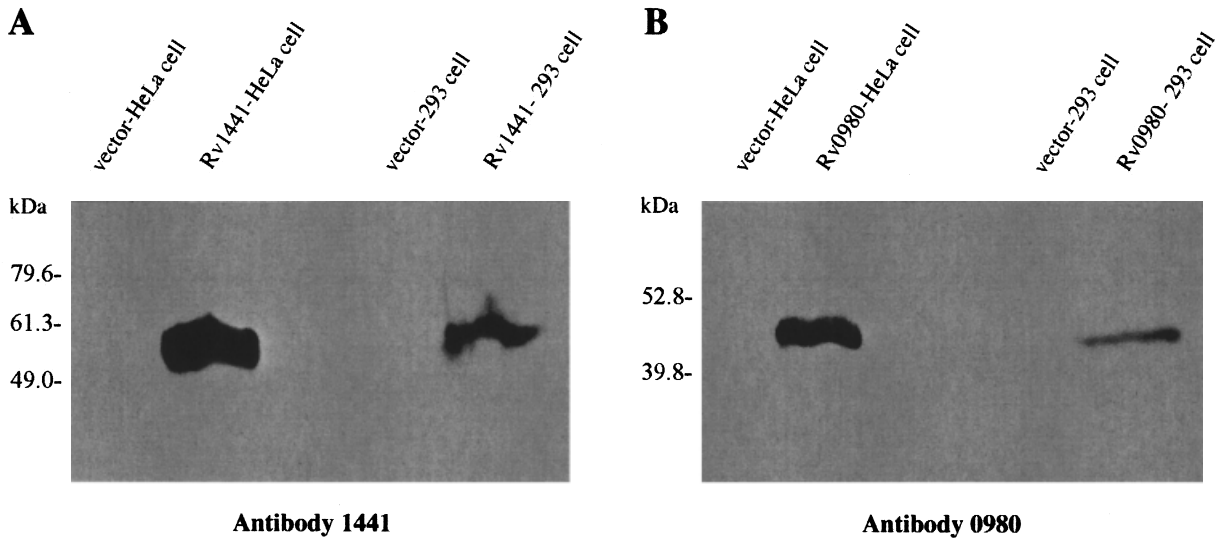

Fig. 4. Western blot analysis of Rv1441c and Rv0980c proteins expressed in eukaryotic cells. The same plasmid constructs used for DNA vaccination were transfected into HeLa and HEK 293 epithelial cells. Proteins were separated by SDS–PAGE and identified by immunoblot analysis using murine antibodies produced by DNA vaccination. Vector without insert was transfected for use as a negative control.
A. Protein extracts from HeLa and 293 cells transfected with plasmid Rv1441c immunoblotted with antibody 1441c.
B. A band of about 42 kDa was identified by antibody 0980c with the Rv0980c protein expressed in both HeLa and 293 cells. The protein markers (in kDa) are indicated on the left.

Specificity control of antibodies

To test the specificity of the antibodies, proteins were also extracted from HeLa and HEK 293 cells transfected with the above five plasmids, expressing the PE-PGRS proteins, or the vector alone, and 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 1). Figure 4 shows that antibody 0980c and antibody 1441c specifically recognized single proteins in both transfected cell lines but not in control extracts. However, in contrast to Rv0980c, the Rv1441c protein displayed an aberrantly high apparent molecular weight (≈ 60 kDa). To examine whether this size discrepancy 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 *Escherichia coli*. In both systems, the protein appeared as an ≈ 60 kDa band, suggesting that its biased amino acid composition might influence its electrophoretic mobility. 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 (data not shown).

Cross-reactivity of the anti-PGRS antibodies

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. Different extracts containing recombinant Rv0980c, Rv1067c, Rv1068c, Rv1441c and Rv 1983 proteins were dot-blotted onto nylon membranes (Table 2) 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, whereas antibody to Rv1441c also cross-reacted with the Rv1983 protein. As expected from the extensive amino acid sequence conservation among

Table 2. Cross-reactivity of PE-PGRS proteins.

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

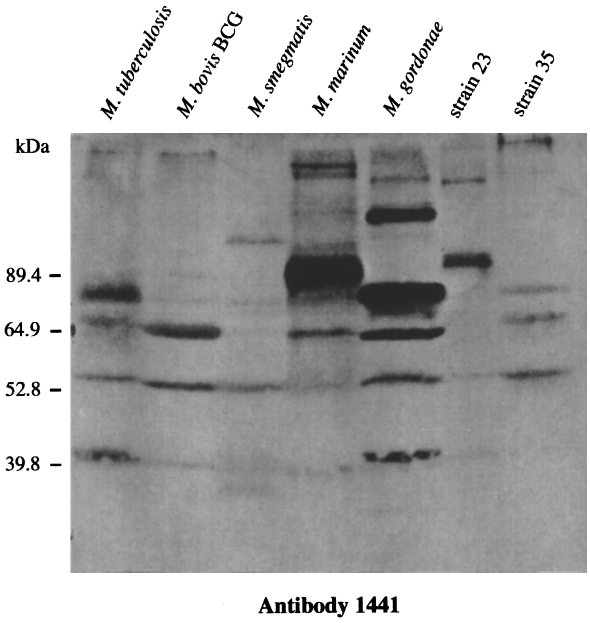

Fig. 5. Western blot analysis of protein extracts from various mycobacteria using antibody 1441c. Total proteins were extracted from *M. tuberculosis*, *M. bovis* BCG, *M. smegmatis*, *M. marinum*, *M. gordonae* and two clinical isolates of *M. tuberculosis* (strain 23 and strain 35), separated by SDS–PAGE and analysed by immunoblotting with antibody 1441c produced by DNA vaccination. The molecular weight markers (in kDa) are indicated on the left.

the PE family members (Cole and Barrell, 1998; Cole *et al*., 1998), these results clearly indicate that the different antibodies tested cross-react, implying that some PE-PGRS proteins share common epitopes.

Immunodetection of PE-PGRS proteins in M. tuberculosis *and other Mycobacteria*

After confirming the specificity of the murine antibodies against different PE-PGRS proteins produced in surrogate expression systems, immunoblot experiments were performed to detect PE-PGRS proteins in cell-free extracts of *M. tuberculosis* and other mycobacteria that had been shown previously to possess PGRS by Southern blotting (Poulet and Cole, 1995), namely *Mycobacterium bovis* BCG, *Mycobacterium smegmatis*, *Mycobacterium gordonae* and *Mycobacterium marinum.* Figure 5 shows the results obtained with antiserum Rv1441c, and it can be seen that from six to 10 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. Fewer proteins (about four) were detected when the other four antisera were used. When serum was used from control mice, immunized with the vector alone, none of these proteins was recognized (data not shown).

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.

Comparison of M. tuberculosis *clinical isolates*

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. Figure 6 shows Western blot results comparing proteins of the H37Rv strain of *M. tuberculosis* and six clinical isolates (strains 1, 9, 23, 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, as described above for antibody Rv1441c (Fig. 5). Fewer protein species were detected when antibodies raised against Rv1067c, Rv1068c and Rv1983 were used (data not shown).

The predicted molecular weight of the Rv0980c protein is ≈ 42 kDa (Table 1) 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 proteins. A strong band of about 65 kDa was found

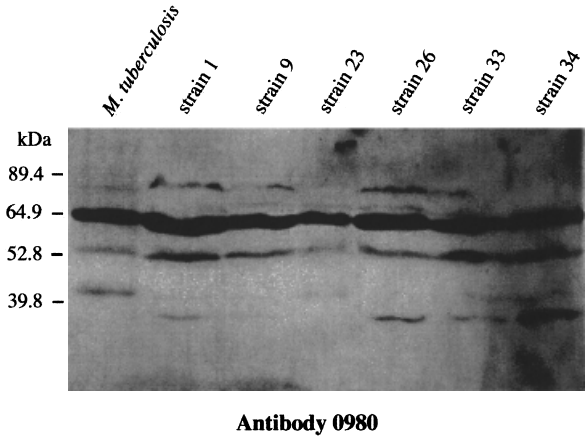

Fig. 6. Immunoblot analysis of protein extracts from different clinical isolates of *M. tuberculosis* using antibody 0980c. Proteins from the H37Rv strain of *M. tuberculosis*, strains 1, 9, 23, 26, 33 and 34 were separated by SDS–PAGE. Several proteins of various sizes from each strain were identified by immunoblotting using antibody 0980c. The molecular weight markers (in kDa) are indicated on the left.

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.

Subcellular localization of PE-PGRS proteins of M. tuberculosis *H37Rv and CDC1551*

The subcellular 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*, as the genome sequences of both organisms are available, and subjected to Western blot analysis using the antibody to Rv1441c, to the membrane protein OmpA (Senaratne *et al*., 1998) and to catalase-peroxidase, KatG (Saint-Joanis *et al*., 1999). In both strains, several putative PGRS proteins were recognized with differing intensities by the Rv1441c antibody (Fig. 7), and the majority 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. 7).

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. 8). In similar experiments using 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.

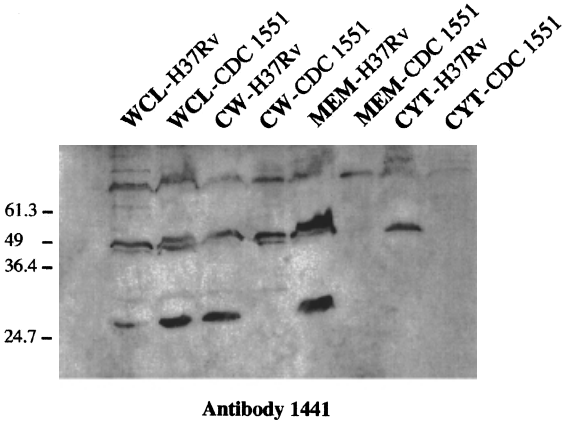

Fig. 7. Subcellular localization of PE-PGRS proteins in two strains of *M. tuberculosis.* Cell wall, cell membrane and cytosol fractions were prepared and proteins separated by SDS–PAGE. After transfer to nylon membranes, antibody 1441c was used to locate the PE-PGRS protein in the cell. The antibody recognized several proteins in each cell fraction: WCL, whole-cell lysate; CW, cell wall fraction; MEM, cytoplasmic membrane fraction; CYT, cytosol. The molecular weight markers (in kDa) are indicated on the left.

Discussion

Much insight into the biochemistry, general metabolism and physiology of tubercle bacilli has been obtained from the complete genome sequence of H37Rv, the most widely used strain of *M. tuberculosis* (Cole *et al*., 1998). However, a great deal remains to be learned about the biological roles of the proteins comprising the PE and PPE protein families. We have shown here that 10 genes encoding members of the PE-PGRS class are transcribed in cultured *M. tuberculosis* H37Rv cells and that five antibodies, raised by genetic immunization, recognized single proteins when the corresponding genes were expressed in eukaryotic cells. On examination of cell-free extracts of *M. tuberculosis*, proteins of the appropriate size were detected (in four cases), in addition to several other protein species. 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.

One of the most important findings of the present work concerns the subcellular location of some members of the PE-PGRS family. 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* (data not shown) 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.

Variations in the protein profiles were observed when clinical isolates were examined, suggesting that at least some PE-PGRS genes undergo sequence variation, providing experimental confirmation of the predictions of variability from earlier allelic comparisons between *M. tuberculosis* H37Rv and *M. bovis* BCG (Cole *et al*., 1998), and the extensive polymorphism in hybridization patterns

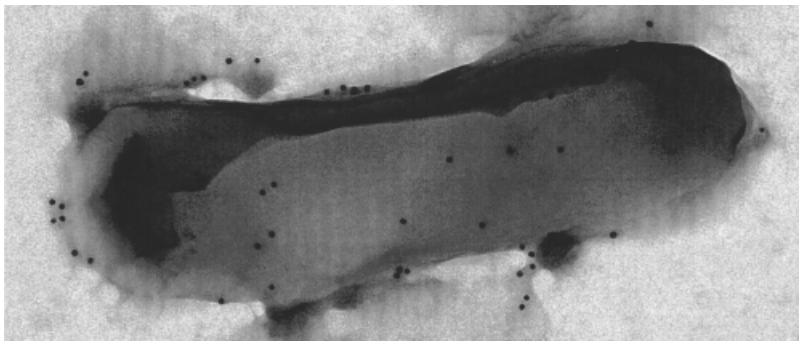

Fig. 8. Immunogold localization of PE-PGRS proteins in *M. tuberculosis* H37Rv. Intact cells were incubated with antibody 1441c and prepared for electron microscopy by negative contrast as described. No labelling of the surface was obtained in control experiments from which this antibody had been omitted.

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* (Stern and Meyer, 1987; Brunham, 1993; Gilsdorf, 1998; Long *et al*., 1998; McCrea *et al*., 1998; Zhang and Norris, 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 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 (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 9 bp 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 microorganisms, 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,

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 (Abou-Zeid *et al*., 1991; Espitia *et al*., 1999; Singh *et al*., 2001). In the case of Rv1759c, not all sera from tuberculosis patients contain antibodies that recognize the protein, suggesting that the Rv1759c protein may be absent or altered in some strains. Espitia *et al*. (1999) have also shown that one PE-PGRS protein is capable of binding fibronectin and could thus mediate bacterial attachment to host cells, but not all PE-PGRS proteins tested have this activity (Singh *et al*., 2001; our unpublished findings). In addition, members of the PE-PGRS families have been implicated in the pathogenesis of *M. marinum* (Ramakrishnan *et al*., 2000), where at least two genes were shown to be upregulated strongly after phagocytosis of the bacterium.

The immunogenicity of the PE-PGRS protein, Rv 1818c, was studied in great detail in mice, and the findings have been reported in a recent publication (Delogu and Brennan, 2001). Immunization 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 (Hinman *et al*., 2000). 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 organization 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, the dataset presented here suggests that the PE-PGRS genes may be contingency genes, some of which encode components of the cell envelope that can undergo extensive variation. Given the remarkable level of sequence conservation in the remainder of the proteome (Sreevatsan *et al*., 1997; Betts *et al*., 2000), 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.

Experimental procedures

Bacterial strains

Mycobacterium tuberculosis H37Rv, *M. bovis* BCG, *M. smegmatis*, *M. marinum* and *M. gordonae* strains were obtained from laboratory stocks. Clinical isolates of *M. tuberculosis* (strains 1, 9, 23, 26, 33, 34 and 35) have been described previously (Heym *et al*., 1994); strains 357 and 1627 are of Spanish origin, and strains 616 and 1349 were from San Francisco. The vectors and derived recombinants were maintained in *E. coli* DH5α.

RT–PCR

RNA for RT–PCR reactions was prepared from a 1-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 µl of 600-µm-diameter beads and a Mickle apparatus. After centrifugation at 12 000 *g* for 10 min, the supernatant was chloroform extracted, ethanol precipitated and resuspended in 60 µl of DEPC-treated water, before treatment with DNase I (Boehringer Mannheim). The mRNA was then reprecipitated and resuspended in DEPC-treated water. The Promega Access RT–PCR system was used to perform RT–PCR. 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 of dNTP mix (10 mM each dNTP), 1 µM concentration of each primer, 2 ml of 25 µM $MgSO_4$, 0.1 µl of AMV RT, 0.1 µl of *Tfl* DNA polymerase, 30–50 ng of RNA sample and nuclease-free water up to 50 µl. Thermal cycling was performed on a PTC-100 thermal cycler (MJ Research) 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 s at 94°C, 1 min at 60°C, 2 min at 68°C and one cycle for 7 min at 68°C.

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 (500 µg ml^{-1} lysozyme, 25% sucrose, 50 mM Tris, pH 8.0, 50 mM EDTA). After incubation for 4 h at 37°C, an equal volume of Proteinase K solution (100 mM Tris, pH 8.0, 400 µg ml^{-1} Proteinase K, 1% SDS) was added and reincubated for a further 4 h at 55°C, before phenol–chloroform extraction and ethanol precipitation to obtain chromosomal DNA. DNA from the Bac clones was prepared as described previously (Brosch *et al*., 1998). Cosmid DNA from the pYUB18 shuttle library of *M. tuberculosis* H37Rv was prepared as described previously (Eiglmeier *et al*., 1993).

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 anaesthesis 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 Rv 1983) 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 1×10^5 cells ml^{-1} and transfected the following day using Fugene 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 (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 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 [^{35}S]-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

g for 1 h and washed twice in PBS. The final cell wall pellet was suspended in PBS. The supernatant was subjected to a 100 000 *g* centrifugation for 4 h. 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 and dialysed against 0.01 M ammonium bicarbonate. Protein contents of all fractions were quantified using the BCA protein assay, and aliquots were stored at –80°C.

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% paraformaldehyde in PBS for 20 min at room temperature. Grids were incubated for 20 min in PBS containing 0.25% NH_4Cl then blocked by Aurion blocking solution (Aurion 905002) for 20 min. This was followed by incubation for 1 h at room temperature with antibody 1441 (1:5 dilution) in PBS BSA-C TM Aurion. After five 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 Jëol 1200 EX II electronic microscope at an accelerating voltage of 80 kV.

Acknowledgements

We thank France Lasnier and Gilles Fischer of the French Embassy in Dhaka, Professor S. I. Khan of Dhaka University and Dr Firdausi Qadri of ICDDR,B for their continued help and support, and Dr Kris Huyghen for advice on DNA vaccination. Sayera Banu is a recipient of a PhD fellowship from the French Government. This work was supported by grants from the European Union (QLRT-2000-02018), the Institut Pasteur and the Association Française Raoul Follereau.

References

Abou-Zeid, C., Garbe, T., Lathigra, R., Wiker, H.G., Harboe, M., Rook, G.A.W., and Young, D.B. (1991) Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect Immun* **59:** 2712–2718.

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36.

Bailey, T.L., and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14:** 48–45.

Betts, J.C., Dodson, P., Quan, S., Lewis, A.P., Thomas, P.J., Duncan, K., and McAdam, R.A. (2000) Comparison of the proteome of *Mycobacterium tuberculosis strain* H37Rv with clinical isolate CDC 1551. *Microbiology* **146:** 3205–3216.

Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., *et al.* (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome (BAC) library for genome mapping, sequencing and comparative genomics. *Infect Immun* **66:** 2221–2229.

Brosch, R., Gordon, S.V., Eiglmeier, K., Garnier, T., Tekaia, F., Yeramian, E., and Cole, S.T. (2000) Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. In *Molecular Genetics of Mycobacteria.* Hatfull, G.F., and Jacobs, W.R., Jr (eds). Washington, DC: American Society for Microbiology Press, pp. 19–36.

Brunham, R.C., Plummer, F.A., and Stephens, R.S. (1993) Bacterial antigenic variation, host immune response, and pathogen–host coevolution. *Infect Immun* **61:** 2273–2276.

Chausse, M.S., Wilson, J., and Hill, S.A. (1999) Characterization of the *recD* gene of *Neisseria gonorrhoeae* MS11 and the effect of *recD* inactivation on pilin variation and DNA transformation. *Microbiology* **145:** 389–400.

Cole, S.T. (1999) Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett* **452:** 7–10.

Cole, S.T., and Barrell, B.G. (1998) Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Genet Tuberculosis (Novartis Found Symp)* **217:** 160–172.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537–544.

Davis, H.L., Michel, M.-L., and Whalen, R.G. (1993) DNA-based immunization induces continuous secretion of hepatitis B surface antigen and high levels of circulating antibody. *Hum Mol Genet* **2:** 1847–1851.

Delogu, G., and Brennan, M.J. (2001) Comparative immune response to PE and PE-PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun* **69:** 5606–5611.

Donnelly, J.J., Ulmer, J.B., and Liu, M.A. (1994) Immunization with DNA. *J Immunol Methods* **176:** 145–152.

Eiglmeier, K., Honoré, N., Woods, S.A., Caudron, B., and Cole, S.T. (1993) Use of an ordered cosmid library to deduce the genomic organisation of *Mycobacterium leprae*. *Mol Microbiol* **7:** 197–206.

Espitia, C., Laclette, J.P., Mondragon-Palomino, M., Amador, A., Campuzano, J., Martens, A., *et al.* (1999) The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology* **145:** 3487–3495.

Fine, P.E.M. (1995) Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* **346:** 1339–1345.

Gilsdorf, J.R. (1998) Antigenic diversity and gene polymorphisms in *Haemophilus influenzae*. *Infect Immun* **66:** 5053–5059.

Heym, B., Honoré, N., Truffot-Pernot, C., Banerjee, A., Schurra, C., Jacobs, W.R., Jr, *et al.* (1994) The implications of multidrug resistance for the future of short course chemotherapy of tuberculosis: a molecular study. *Lancet* **344:** 293–298.

Hinman, M.B., Jones, J.A., and Lewis, R.V. (2000) Synthetic spider silk: a modular fiber. *Trends Biotechnol* **18:** 374–379.

Hirschfield, G.R., McNeil, M., and Brennan, P.J. (1990) Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. *J Bacteriol* **172:** 1005–1013.

Hollingshead, S.K., Fischetti, V.A., and Scott, J.R. (1987) Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* **207:** 196–203.

Huygen, K., Content, J., Denis, O., Montgomery, D.L., Yawman, A.M., Deck, R.R., *et al.* (1996) Immunogenicity

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., Sharipo, 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., Weldingh, K., Jacobsen, S., Hansen, C.V., Florio, W., Gianetri, I., 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., Raios, 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., Mobasheri, H., Papavinasasundaram, 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 Laal, 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., Felgner, 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 *vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun* **66:** 3698–3704.