CHAPTER 4

DISCUSSION
DISCUSSION

One third, approximately 1.7 billion people, of the world's population is infected with tuberculosis. Tuberculosis is the number one cause of mortality from infectious agents world-wide, killing an estimated 3 million people each year, more than 10 times the number of deaths due to AIDS. Global tuberculosis control will unlikely come from development of new antituberculosis drugs because they will be too expensive to be used in developing countries. Control of tuberculosis depends on development of an effective antituberculosis vaccine. Development of a safer and more effective vaccine has been made feasible by two recent developments. First, it was found that immunization of guinea pigs with a crude mixture of proteins secreted by the tuberculosis bacterium protected against development of tuberculosis. Second, in June 1998, the genome of the tuberculosis bacterium, *Mycobacterium tuberculosis*, was sequenced, and the post genomic era may allow identification of new genes that may produce proteins that elicit protective immunity (Ahmed and Hasnain, 2004, Chakhaiyar and Hasnain, 2004). These advances will make it possible to make a "subunit" protein vaccine, which will be much safer than a live vaccine. However, it is essential to first identify the best vaccine candidate proteins, and second, to develop methods to produce large quantities of these proteins.
4.1. The PE and PPE gene families of *Mycobacterium tuberculosis*

Among the most interesting gene families found in the mycobacteria are the PE and PPE families. Ten percent of the *Mycobacterium tuberculosis* genome is devoted to these genes, encoding acidic glycine-rich proteins. The field is rife with speculation on the functions of these proteins, including their functions to provide antigenic variation and interfere with immune responses. The amino acid composition of these protein families differs radically from that of the bulk of the proteins of the Mycobacterial genome. Both the PE and PPE proteins are exceptionally glycine-rich while the PPE proteins also contain copious amounts of asparagine, an amino acid that is generally rare in the proteome. Curiously, asparagine is the preferred nitrogen source for *Mycobacterium tuberculosis*, and this raises the possibility that the PPE proteins may also serve as storage proteins. Prior to completion of the genome sequence, the existence of these protein families, whose genes occupy 10% of the total coding sequence, was unknown. It was clear, however, that the genome contained two dispersed simple sequence repeats referred to as PGRS (polymorphic G+C-rich sequence) and MPTR (major polymorphic tandem repeat) and these have since been shown to
Discussion

correspond to part of the 3’ ends of the PE and PPE genes (Cole et al, 1998).

Review of literature reveals substantial information regarding the PE families of proteins. Banu et al showed 10 of these genes to be variable surface antigens (Banu et al, 2002). A significant humoral immune response was observed against recombinant Rv1818, a member of the PE-PGRS family (Brennan et al, 2001). Differential expression of PE and PE_PGRS genes in Mycobacterium tuberculosis strains has also been reported (Flores and Espitia, 2003). Singh et al showed that Rv3367, a member of the PE-PGRS family is recognized by pooled sera from TB patients and not from healthy controls, confirming its in vivo expression during active infection in humans (Singh et al, 2001). pH dependent expression of Rv0834c, a member of the PE-PGRS family of Mycobacterium tuberculosis has been reported (Saviola et al, 2003). However very few reports are available regarding the PPE family of M. tuberculosis. The primary objective of this study therefore, was to identify novel immunodominant antigen(s) within the PPE family of M. tuberculosis.
4.2. *In silico* and web based analyses

The 69 members of the PPE protein family have a conserved N-terminal domain that comprises ~180 amino acids followed by C-terminal segments that vary markedly in sequence and length. Based on our pattern search analysis of the Tuberculist database (http://genolist.pasteur.fr/TubercuList/) these proteins were categorised into three groups. Subgroup 1, represented by 20 members, constitutes the MPTR class characterized by the presence of multiple, tandem copies of the motif Asn-X-Gly-X-Gly-Asn-X-Gly. The second subgroup, comprising of 21 members, contains a characteristic well conserved motif Gly-X-X-Ser-Val-Pro-X-X-Trp around position 350, and the third subgroup proteins, with 28 members, are unrelated except for the presence of the common PPE domain. Proteins belonging to the third sub group was the focus of the present study as members of these sub-group display significant sequence and length variation in their C-terminal region and are unrelated except for the presence of the common PPE domain.

ORFs belonging to the third subgroup with coding capacity equal to or less than 200 amino acids were shortlisted. This shortlist was further narrowed down based on two very important criteria -
namely, antigenic profile and the association of the ORF with pathological conditions.

Antigenic profiles were generated using the Protein Analysis Software, Protean, DNASTAR. Two ORFs, Rv2430c and Rv3425 displayed major antigenic stretches. Furthermore, DNA Microarray results demonstrated that Rv2430c was one of the genes induced in IdeR mutant of M. tuberculosis (Rodriguez et al, 2002), pointing to its possible role in pathogenesis. Rv2430c was accordingly shortlisted for the present study and was evaluated for its role as an antigen in a clinical setting. To confirm whether Rv2430c indeed represented a functional gene, mRNA extracted from H37Rv, the virulent strain of Mycobacterium tuberculosis was used in an RT-PCR reaction. The RT-PCR reaction was found to be positive indicating the expression of Rv2430c in liquid cultures of Mycobacterium tuberculosis.

The ORF corresponding to Rv2430c was amplified from the genomic DNA of H37Rv and cloned in an E.coli expression vector. The recombinant protein was purified from E.coli under denaturing conditions and was used for further studies.
4.3. Biophysical characterization of the PPE ORF Rv2430c

Though there have been several reports about immunodominant antigens of *Mycobacterium tuberculosis*, information about the structure of antigens is very little. Ronning *et al* have described the crystal structure of the secreted form of antigen 85C (Ronning *et al*, 2000). The antigen 85 (ag85) complex, composed of three proteins (ag85A, B and C), is a major protein component of the *Mycobacterium tuberculosis* cell wall. Each protein possesses a mycolyltransferase activity required for the biogenesis of trehalose dimycolate (cord factor), a dominant structure necessary for maintaining cell wall integrity. The crystal structure of recombinant ag85C from *M. tuberculosis*, refined to a resolution of 1.5 Å, reveals an alpha/beta-hydrolase polypeptide fold, and a catalytic triad formed by Ser 124, Glu 228 and His 260. ag85C complexed with a covalent inhibitor implicates residues Leu 40 and Met 125 as components of the oxyanion hole. Renshaw *et al* have determined the secondary structures of ESAT-6 and CFP-10, secretory proteins of *Mycobacterium tuberculosis*, which are potent B cell and T cell antigens (Renshaw *et al*, 2002). Using a combination of fluorescence, circular dichroism, and nuclear magnetic resonance spectroscopy they have shown that ESAT-6 contains up to 75% helical secondary structure, but little if any stable tertiary structure, and exists in a
molten globule-like state. In contrast, CFP-10 was found to form an unstructured, random coil polypeptide. An exciting discovery was that ESAT-6 and CFP-10 form a tight, 1:1 complex, in which both proteins adopt a fully folded structure, with about two-thirds of the backbone in a regular helical conformation. This clearly suggests that ESAT-6 and CFP-10 are active as a complex. Goulding et al have determined the crystal structure MPT63, a small, major secreted protein of unknown function from *Mycobacterium tuberculosis* at 1.5 Å resolution that has been shown to have immunogenic properties and has been implicated in virulence (Goulding et al, 2002). The structure of MPT63 is an antiparallel beta-sandwich immunoglobulin-like fold, with the unusual feature of the first beta-strand of the protein forming a parallel addition to the small antiparallel beta-sheet. MPT63 has weak structural similarity to many proteins with immunoglobulin folds, in particular, *Homo sapiens* beta2-adaptin, bovine arrestin, and *Yersinia pseudotuberculosis* invasin. Although the structure of MPT63 gives no conclusive evidence to its function, structural similarity suggests that MPT63 could be involved in cell-host interactions to facilitate endocytosis/phagocytosis. In another study, the purification and characterization of three immunodominant antigens of *Mycobacterium tuberculosis*, namely the 38kDa, 30kDa and 16kDa antigens have been described. The 38-kDa form was purified by
preparative isoelectric focusing, followed by preparative electrophoresis (Devi et al., 2002). Gel-permeation chromatography was employed for the isolation of the 16-kDa form, from the cytosol fraction of M. tuberculosis H37Rv. The purified proteins were characterized by Circular Dichroism studies. Analysis of the CD data revealed that the 38kDa and the 16kDa proteins have a beta sheet like structure.

No information, however, is available regarding the biophysical or structural features of the PE and PPE family of proteins. Hence, the present study was also carried out to gain insights into the structural characteristics of the recombinant PPE protein. Secondary structure prediction employing Protean 4.0 revealed a very high content of a-helical structure in Rv2430c. This was evident from both the Garnier-Robson and Chou-Fasman methods of secondary structure prediction. Web based analysis of Rv2430c using PSIPRED and Predict protein also suggested a predominant a helical composition. To confirm these predictions, the recombinant PPE protein was extracted under denaturing conditions from E.coli as expression of the protein led to its localization in the inclusion bodies. Initial attempts to refold the protein using dialysis resulted in massive precipitation of the protein. Therefore, an on-column refolding strategy (Pullakhandham et al., 2004) was used in the
presence of L-arginine that is known to act as a chemical chaperone (Srinivas et al, 2003), and glutathione was included to provide reducing equivalents during folding. The on-column refolded protein was soluble and was found to be pure by SDS-PAGE analysis. The purified protein displayed a CD spectra characteristic of a helical proteins, confirming the in silico predictions of secondary structure. The possible environment of the aromatic amino acid residues in refolded protein was also checked. The refolded protein exhibited emission maximum at 340nm. Interestingly, incubation of protein with urea resulted in the significant red-shift (340 to 350nm) in the emission maxima and also an increase in the fluorescence intensity was observed, indicating that the aromatic amino acid residues are present in a hydrophobic environment. Thus, there appears to be a well-formed hydrophobic core in the protein, which becomes exposed under the influence of 8M urea. In the absence of any functional assay described for this protein, conformation of the protein analyzed by CD and fluorescence studies compared with in silico predictions of protein based on primary structure suggest that the protein is properly folded.

The current method of refolding can be employed for other members of the PE/PPE protein family to obtain large quantities of protein for crystallization purposes. Such on-column refolding strategy has been
successfully used for a protein which is known to generate oligomers (Pullakhandharn et al., 2004). That such properly folded protein will likely display differential immunoreactivity to patient sera renders it possible to evaluate other members of this family for their likely biological roles.

4.4. Serological characterization of the PPE ORF Rv2430c

Having shown that the ORF Rv2430c was expressed at the mRNA level in liquid cultures of *M. tuberculosis*, experiments were designed to assess its role in eliciting a humoral immune response. The recombinant His-tagged protein was used to assess its immunological potential. Human sera obtained from a total of 101 patients were used in the present study. These patients belonged to Fresh infection cases (Category 1, n=32), Patients with relapsed tuberculosis (Category 2, n=30), Extrapulmonary cases (Category 3, n=30) and Multi Drug Resistant Cases (Category 4, n=9). Also included were 10 clinically healthy controls which were *M. bovis* BCG Vaccinated and sera from 15 non-TB patients (culture negative for *Mycobacterium tuberculosis*). Initial screening of the sera revealed that the recombinant protein coded for by Rv2430c was recognized by infected patients at a serum dilution of 1:200 whereas a poor reactivity was observed in clinically healthy controls. This result
confirmed that Rv2430c is expressed in active infection with *M. tuberculosis* and is involved in disease manifestation and progression.

Having demonstrated that the PPE ORF Rv2430c is associated with disease manifestation and progression, it was of interest to dissect the immune response to Rv2430c. For this the recombinant protein was used to screen the panel of sera obtained from all the four clinical groups. We used Hsp10, a well documented immunodominant antigen of *M. tuberculosis* (Young and Garbe, 1991) as a reference to compare the immune response against Rv2430c in all the four clinical categories. Screening of the sera revealed that recombinant protein Rv2430c was well recognized by the sera of all the four clinical categories. However, as compared to Hsp10, patients belonging to Category 1 mounted a strong immune response to Rv2430c (P<0.003). The immunoreactivity to Rv2430c in the remaining three categories, however was comparable to Hsp10 with values of P<0.8, P<0.06 and P<0.912 for Categories 2, 3, and 4, respectively. Serological sensitivity of Rv2430c as compared to Hsp10 was demonstrated by the fact that a higher percentage of individuals belonging to Category 1 mounted a strong antibody response to Rv2430c as compared to Hsp10. Similar results were obtained when IgM antibodies were assayed.
The Purified Protein Derivative has often been used to diagnose infection with TB. This test has remained essentially unchanged since it was developed 100 years ago by Robert Koch, who also discovered M. *tuberculosis*. Intradermal inoculation of purified protein derivative (PPD, or tuberculin), a crude precipitate of M. *tuberculosis* culture supernatant containing over 200 antigens widely shared among mycobacteria other than M. *tuberculosis*, including M. *bovis* BCG, elicits a local cutaneous delayed type hypersensitivity response in sensitised individuals. This in *vivo* cellular immune response results in a bump in the skin. The size of the bump, or induration, is measured using a scale three to seven days after the test is placed. A large bump is interpreted as indicating M. *tuberculosis* infection, while the absence of a bump implies no infection. Intermediate sized bumps are difficult to interpret. Several reports have emphasized the utility of PPD in the serodiagnosis of tuberculosis. Zeiss *et al* obtained sera from patients with active tuberculosis and sera from appropriate control individuals and screened the same for immunoglobulin G antibody activity to purified protein derivative by a polystyrene tube radioimmunoassay and an enzyme-linked immunosorbent assay. Both assays showed a marked increase in immunoglobulin G antibody activity in patients with active tuberculosis. There was no overlap between the values for the patient group and the values for the purified protein derivative skin
Discussion

test-positive control individuals. The replication of these assays was excellent, and both could provide quantitative measurements of immunoglobulin G antibody activity to purified protein derivative antigen within 24 h (Zeiss et al., 1982). Radin et al assayed the sera of patients with active tuberculosis (TB) and sera from control groups for IgG, IgA, secretory IgA, IgM and IgE antibody activity to purified protein derivative (PPD) using the enzyme-linked immunosorbert assay. Patients with active TB clearly had higher levels of IgG antibody activity to PPD antigen than did healthy patients who were skin test positive or negative. There was a clear separation between the diseased and healthy groups. Similar, but not as marked, increase was seen in IgA and secretory IgA antibody activity in diseased patients. No correlation between the presence of disease and antibody levels were found with IgM, and no IgE antibodies were found (Radin et al 1983). Viljanen et al developed an Enzyme-linked immunosorbert assay (ELISA) for IgM, IgA and IgG antibodies against PPD. PPD antibodies of 44 patients with active pulmonary tuberculosis were measured at admission to hospital. The control material consisted of 35 healthy blood donors. The mean antibody levels in all three immunoglobulin classes correlated with the extent of the tuberculous infection, i.e. the severer the disease the higher the mean antibody level. The mean antibody levels also were significantly higher in the tuberculosis patients than in the controls,
the only exception was specific IgA antibodies in the patients with minimal disease. **ELISA-positivity** of the patients, i.e. positive result in at least one immunoglobulin class was as follows: 33% of the grade I, 64% of the grade II and 100% of the grade II patients (Viljanen et al, 1982). Kalish et al studied three patients with culture-proven *Mycobacterium tuberculosis* meningitis. Analysis of cerebrospinal fluid with an enzyme-linked immunosorbent assay (ELISA) method measuring IgG antibody to purified protein derivative rapidly yielded positive results, whereas results of acid-fast smears were negative and cultures took several weeks before growth appeared. They carried out serial studies of cerebrospinal fluid and sera from one patient. Initially, greater amounts of IgG antibody to purified protein derivative were present in the cerebrospinal fluid than in the serum. The antibody level in the cerebrospinal fluid paralleled the patient's clinical course, cerebrospinal fluid cell count, protein level, and glucose level. Cerebrospinal fluid samples from 33 hospitalized control patients were negative for antibody to purified protein derivative (Kalish et al, 1983a; Kalish et al, 1983b).

### 4.5. Diagnostic potential of PPE ORF Rv2430c

The diagnostic utility of PPD has become a question of debate primarily because of false positives it generates in the case of healthy
controls. The broad antigenic cross-reactivity of PPD is responsible for the poor specificity of the Tuberculin Skin Testing (TST); a positive reaction is consistent with BCG vaccination as well as M. *tuberculosis* infection. Since one third of the world's population is believed to be infected with M. *tuberculosis* and the majority have been BCG-vaccinated, accurate identification of *M. tuberculosis* infected people for targeted chemoprophylaxis is very difficult. There are also numerous operational drawbacks with the TST: administration and reading of the TST are highly operator dependent, standardization of PPD is problematic and a strongly positive TST can cause painful ulceration and scarring. Taking into account the above facts the need for identification of novel antigens which can also serve the purpose of serodiagnosis has become one of the primary goals in *M. tuberculosis* research. We compared the immunopotentiality of Rv2430c with PPD. Interestingly, patients belonging to Category 1 mounted statistically significant immune response to Rv2430c as compared to PPD.

In order to confirm the diagnostic ability of Rv2430c, we compared the reactivity of Rv2430c in sera obtained from clinically healthy donors and non-TB patients. The non-TB patients were confirmed TB negative both by acid fast staining and culture and therefore they could be harboring any pathogen other than *M. tuberculosis*. ELISA
Discussion

Results revealed that Rv2430c displayed similar reactivity to sera obtained from clinically healthy donors and non-TB patients. However, the reactivity to the sera obtained from TB patients was statistically significant as compared to sera obtained from clinically healthy controls and non-TB patients indicating the presence of antibodies against Rv2430c in infected patients and their absence in non-TB patients. These results clearly indicate that Rv2430c does not cross react with the sera of non-TB patients and can differentiate between infected and non-TB patients.

Several reports have emphasized on the observation of lack of sufficient immune responses in TB patients against many promising serodiagnostic antigens of *M. tuberculosis*. The fact is more distressing in case of fresh infection or active infection where for majority of the cases the immune system is not sufficiently primed to elicit a strong antibody responses against most of the *M. tuberculosis* antigens. The recombinant Rv2430c protein was very strongly recognized by all the four categories of patients including the fresh infection or so called active infection group.

The development of a sensitive and rapid serodiagnostic test in tuberculosis (TB) would complement present methods of diagnosis including skin testing, DNA amplification, bacterial culture and
radiological imaging. The major focus of tuberculosis research also includes serodiagnosis of tuberculosis, since an early knowledge about the patient’s disease would be an effective measurement for global control of tuberculosis. The ELISA based serodiagnosis appears to be rapid and inexpensive assay which could reduce the cost of diagnosis too.

It is pertinent to note that although several antigens have been tested for their use in serodiagnosis (Amara et al, 1998, Batoni et al, 2002, Devi et al, 2002, Dillon et al, 2000, Florio et al, 2002, Houghton et al, 2002, Laal et al, 1997, Ljungqvist et al, 1990, Lim et al, 2000, Lodes et al, 2001, Mustafa et al, 2002), no single antigen has proved to be able to achieve sensitivity and specificity in a study population suitably large and heterogenous. The factors responsible include a) the stage of the disease, b) the location of the infection, and c) the genetic background. Our results reveal that Rv2430c is expressed during active infection with TB and shows better reactivity to sera from fresh infection or the so called active infection cases when compared to Hsp10 and PPD. These results clearly demonstrate the immunodominant as well as immunodiagnostic nature of Rv2430c. It would be interesting to speculate on the use of Rv2430c alongwith other immunodominant antigens for vaccine development (Dhar et al, 2000). The PPE ORF Rv2430c used in our
study belongs to the third subgroup of the PPE class of proteins. Members of this subgroup are unrelated. Given the known antigenic variability between the different members of this family the antigenic epitopes will not be common. Therefore, the chances of cross-reactivity would be minimum, if at all.

Data obtained from microarray studies reveal that Rv2430c is one of the genes upregulated in Rel Mtb knock out of \textit{Mycobacterium tuberculosis} \cite{Dahl et al, 2003}. Long-term survival of nonreplicating \textit{Mycobacterium tuberculosis} is ensured by the coordinated shutdown of active metabolism through a broad transcriptional program called the stringent response. In \textit{M. tuberculosis}, this response is initiated by the enzymatic action of RelMtb and deletion of RelMtb produces a strain (H37RvAre1Mtb) severely compromised in the maintenance of long-term viability. Microarray analysis revealed that H37RvAre1Mtb suffers from a generalized alteration of the transcriptional apparatus, as well as specific changes in the expression of virulence factors, cell-wall biosynthetic enzymes, heat shock proteins, and secreted antigens that may alter immune recognition of the recombinant organism. Hence, Rv2430c could be one of the genes responsible for the long term survival of non replicating \textit{Mycobacterium tuberculosis}. 
In order to ascertain the probable localization of Rv2430c, the ORF was cloned in pET23a expression vector and experiments were designed to ascertain its localization using a Rabbit Reticulocyte Lysate cell free coupled transcription and translation system. *In vivo* translation yielded a product of the expected size indicating lack of any post translational modification of the protein. Translation in the presence of Canine Pancreatic Microsomal membrane was set up to determine the localization of the protein. No reduction in size of the protein was observed in the above reaction pointing to the lack of a secretory signal. Digestion of the protein translated in the presence of Canine Pancreatic Microsomal membrane by Trypsin led to the complete digestion of this protein indicating the absence of any functional transmembrane domain in the protein. The Rv2430c, originally considered a hypothetical protein, but based on results presented in this thesis, could be ascribed a possible role in acting as an immunodominant antigen. On the question of localization of this protein, *in silico* analysis points to the unlikely possibility of Rv2430c being a transmembrane protein. With a strong B cell response it is highly probable that Rv2430c is a secretory protein.