Summary and Conclusions
The completion of multiple mycobacterial genome sequencing projects has initiated a new approach towards the study of these diseases in order to comprehend the mechanisms underlying their pathogenesis. Using a genome-wide transposon insertional mutant library of *M. tuberculosis* along with *in vitro* and *in vivo* screens, several genes essential for the growth of the bacteria under different growth conditions have been identified (Lamichhane et al., 2003; Sassetti et al., 2001; Sassetti and Rubin, 2003). Interestingly, a large number of attenuating mutations identified through these strategies are found to be unique to mycobacteria and other closely related species. Such global studies generate invaluable information for classifying essential and non-essential genes, however, identification of the exact biological function of these gene products and their involvement in specific pathway requires systematic investigations. The genome projects have enormously changed the way research is performed in the area of metabolic biosynthesis. In the past, knowledge of metabolic pathways and enzymatic reactions were accumulated from numerous isolated biochemical experiments from different organisms. The identification of new putative genes in the present genomic era has been reduced to mere clicks on the computer.

In this thesis, the biochemical pathway leading to assembly of PDIMs has been elucidated. PDIM esters are surface-exposed lipids unique to the virulent strains of mycobacteria (Brennan and Nikaido, 1995; Daffé and Draper, 1998). In order to unambiguously delineate biosynthetic pathway, the precise roles of several proteins have been established with demonstrable functions. Cell-free reconstitutions of more than 22 catalytic steps demonstrate that proteins from *pps*
cluster can synthesize PDIMs by condensing long-chain fatty acids with simple dicarboxylic acyl-CoA units.

Genetic studies earlier had suggested an array of genes involved in the biosynthesis of DIM esters, which had complicated our understanding of the enzymology and assembly of these complex virulent lipids. The *pks7*, *pks10*, *pks12*, *ppsA-B* and *msl7* mutant strains of *M. tuberculosis* were shown to be deficient in phthiocerol derivatives, and had yielded strains with attenuated growth in the murine model (Azad et al., 1997; Kolattukudy et al., 1997; Rousseau et al., 2003b; Sirakova et al., 2003a; Sirakova et al., 2003b). Most of these genes could be classified in two large clusters. The *pps* locus consists of 35 open reading frames and encompasses ~70kbp of the genome. The other gene cluster contains six PKS genes - *pks10*, *pks7*, *pks8*, *pks17*, *pks9* and *pks11* and is flanked on either side by type III PKS genes. Since gene disruption studies result in loss of function, the involvement of all these genes in biosynthesis of PDIMs was inferred from their absence from mycobacterial cell envelope. Preliminary analysis of the PKS10, PKS7 and PKS9 proteins in our laboratory have suggested that these proteins synthesize metabolic intermediates that are chemically incompatible with the structure of PDIM and therefore cannot generate precursors leading to the synthesis of phthiocerol derivatives. It is possible that the proposed involvement of other PKS genes might be an experimental artifact arising from the disruption in lipid transport. Indeed in a recent report, a fraction of mycobacterial mutants of the unrelated loci were found to be deficient in PDIM esters (Domenech et al., 2004).
The biosynthetic pathway in this thesis has been established by demonstrating functions for purified isolated proteins. During the biochemical characterization of FadD proteins (also known as acyl-CoA synthetases), it was observed that some of the proteins present adjacent to PKS and FAS gene clusters were not able to synthesize acyl-CoA units. Instead, these proteins activated fatty acids as acyl-adenylates and transferred long-chain fatty acids on to the adjacent PKS proteins and were annotated as fatty acyl-AMP ligases (FAALs) (Trivedi et al., 2004). The mechanism of fatty acid activation is analogous to the activation of amino acids by adenylation domains of non-ribosomal peptide synthetases (NRPSs). It is worthwhile to note that the involvement of FadD proteins in biosynthesis of complex lipids is contrary to their previously established function of FadD proteins in fatty acid degradation.

Based on the retro-biosynthetic approach, the PDIM biosynthetic machinery has been deconvoluted in four steps. The proposed mechanism of biosynthesis is shown in Figure 1. The first step in phthiocerol synthesis was established by using FadD26 protein to transfer C\textsubscript{12} to C\textsubscript{18} fatty acids on to the phosphopantetheine thiol of purified PpsA protein. The synthesis of diol moiety is performed by two modular PKSs, which extend the lauroyl chain by 4 carbons by using 2 units of malonyl-CoA. The final step of phthiocerol was catalysed by PpsE protein. The acyl transferase domain of PpsE protein was shown to incorporate both malonyl as well methyl malonyl-CoA efficiently. Interestingly, the diesters of phthiocerol and related compounds are mixtures of long-chain β-diols. These esters possess heterogeneity at the priming precursor (R\textsubscript{1}) position as well as at terminal methyl group position (R\textsubscript{3}) of the phthiocerol moiety (Figure 1).
Figure 1: Mechanism of the biosynthesis of the virulent cell wall lipid PDIM. The *pps* cluster consists of 15 open reading frames involved in the synthesis of the core of PDIM. The intermediates synthesized during the assembly are covalently bound as acyl-thioesters on to the ACP of the respective proteins. The domains present in all the proteins are shown. The FadD26 transfers the fatty acyl adenylate onto the PCP of the PpsA protein. The Pps proteins elongate the acyl chains to form the phthiocerol. The Mas catalyses the formation of the mycocerosic acids. The PapA5 protein catalyses the transesterification of the mycocerosic acids onto the phthiocerol to form the PDIM.

catalytic flexibility of FadD26 and PpsE proteins readily explains the molecular basis of this chemical diversity. The Mas protein was shown to synthesize multi-methylated fatty acids. Interestingly, this protein also showed broad substrate specificity. Mas protein could utilize C₆ to C₂₀ fatty acids and extend it by 2, 3 or 4 iterations by using one molecule of methyl malonyl-CoA during each cycle of condensation and associated reductions. This enzyme was analogous to the previously characterised Mas protein from *M. bovis* (Rainwater and Kolattukudy, 1983; Rainwater and Kolattukudy, 1985). During the course of this work, the requirement of PapA5 (polyketide associated protein) in PDIM biosynthesis and its function using surrogate substrates was confirmed, although the mechanism of
PapA5 in PDIM formation was not demonstrated (Onwueme et al., 2004). In this study, the PapA5 was demonstrated to transacylate covalently bound mycocerosic acids on to long-chain alcohols. This mode of hand-to-hand transfer may be the most efficient method to channelize intermediates, since free long-chain methylated mycocerosic acids or its CoA-esters can interfere with lipid metabolism. The other Pap homologues in *M. tuberculosis* are also present adjacent to genes involved in lipid metabolism. It is plausible that these Pap proteins would probably condense individual components through a similar mechanism.

It is interesting to note that the complete biosynthesis proceeds in the cytoplasm by using soluble enzymes, although most of the enzymatic intermediates are covalently sequestered on to the protein. The PDIM is supposedly transported to the outermost portion of the cell surface by the transporter proteins such as the MmpL7 and the DrrABC (Camacho et al., 2001; Cox et al., 1999). However, the exact mechanism is still unknown. It is interesting to note that the Mas protein is identified at the membrane by using antibody localization as well during fractionation process (Gu et al., 2003). It is feasible that some of these proteins of the biosynthetic machinery are present at the cytosolic face of the plasma membrane, which then chaperone these lipids to the transporters. Interestingly, the inactivation of MmpL8, which is proposed in sulpholipid transport, result is in accumulation of intermediate lipid (Converse et al., 2003; Domenech et al., 2004). The reconstitution of the biosynthetic machinery for PDIM provides tools for understanding the mechanisms associated with transport of molecules across the
membrane. The secretion of lipids and proteins is a fascinating unexplored area in mycobacteriology.

The enzymes involved in the biosynthesis of several lipids are already prime targets for some of the antituberculosis drugs. The chemotherapeutic drug, isoniazid targets an enzyme involved in the synthesis of long chain α-alkyl, β-hydroxy mycolic acids (Banerjee et al., 1994; Quemard et al., 1995). Ethambutol inhibits the biosynthesis of arabinogalactan by interfering with the polymerisation of both arabinan and galactan segments (Deng et al., 1995). Pyrazinamide inhibits the eukaryotic FAS I (Zimhony et al., 2000). The characterisation of unique essential PDIM lipid biosynthetic pathway makes it attractive targets for developing new antimycobacterial drugs.

The in vitro synthesis of the components of PDIM biosynthetic machinery described in this study would facilitate investigation of the mechanistic and structural basis of the programming of these complex lipids. The results presented here unravel molecular mechanisms in which mycobacteria synthesizes complex lipids from ubiquitous n-long-chain fatty acids. This first report of characterization of PKSs in generating complex lipids provides an exciting opportunity to engineer new lipid metabolites by altering the specificity of various domains in these proteins. For example, the synthesis of linear-chain fatty acids instead of the hallmark branched-chain fatty acid could be explored by merely altering the specificity AT domain from methyl malonyl-CoA to malonyl-CoA. This tantalizing opportunity to perform subtle chemical changes at the mycobacterial cell envelope using this novel approach would provide new ways to explore host-pathogen interactions.