Chapter 3
Cell-free synthesis of mycocerosic acids and investigation of the role of type II thioesterase
3.1 INTRODUCTION

The mycobacterial cell wall plays an important role in the survival of the bacteria within the host macrophages. Apart from providing an impregnable outer coat, some of the molecules present on the cell wall are also proposed to alter the immune responses of the hosts (Brennan, 2003; Brennan and Nikaido, 1995). The unusual methylated fatty acids present as components of glycolipids and lipooligosaccharides form a highly compact structure that hinders diffusion of molecules through the cell wall. The mycobacteria contain unique tetramethyl branched long chain fatty acids which are esterified to phthiocerol (Daffe and Draper, 1998). Fatty acids are synthesized by elongation of acetyl-CoA primers by decarboxylative condensation with malonyl-CoA units. However, the closely related PKSs generate methyl-substituted secondary metabolites by incorporation of methyl malonyl-CoA (Hopwood and Sherman, 1990). Structural and functional homology between PKS and FAS systems suggest that the long chain multimethyl branched fatty acids could be synthesized by extension of fatty acids by methyl malonyl-CoA. The vertebrate FAS have been shown to incorporate methyl malonyl-CoA into fatty acids at very low rates although malonyl-CoA is an overwhelmingly favoured substrate (Buckner and Kolattukudy, 1976; Buckner et al., 1978). The FASs possess a thioesterase domain at the C-terminal end, which is involved in the active release of the fatty acids. Similar C-terminal thioesterase domains (type I TE) are present in PKS systems and are involved in the release and cyclization of the covalently bound final product (Chiu et al., 2001; Gehring et al., 1998c; Gokhale et al., 1999a). More recently, independent open reading frames
bearing low sequence homology to the thioesterase domains have been identified in the PKS and NRPS gene clusters (Linne et al., 2004; Xue et al., 1998). The functional significance of these proteins has been difficult to rationalise and in several cases, the disruption of such genes have led to dramatic decrease in product formation (Butler et al., 1999). These proteins are able to release stalled covalently bound chains from number of multienzymatic proteins and are therefore implicated in the proofreading activity during the biosynthesis of the polyketide or non-ribosomal peptides. Three dimensional crystal structures of type I and type II thioesterases have revealed different structural folds. The TE I proteins have the α/β hydrolase fold while the TE II proteins possess a new tertiary fold with twelve antiparallel β-sheets.

In this study, we have characterised an iterative FAS, which is involved in the biosynthesis of methylated mycocerosic acids. This Mas protein does not contain a C-terminal TE domain. An independent type II thioesterase annotated as tesA is present at the 5' end of the pps gene cluster. Here, we have characterised the biochemical function of TesA and its importance in the biosynthesis of phthiocerol dimycocerosate esters.

3.2 MATERIALS AND METHODS

3.2.1 Materials

Superior quality analytical grade reagents were used in this study. The detail of the materials used is described earlier in Chapter 2.
3.2.2 Methods

3.2.2.1 Cloning of the tesA genes

The tesA gene and the TE domain from pks6 gene (referred to as PKS6-TE) from *M. tuberculosis* were cloned with the help of Uttara by using standard molecular biology protocols as described in Chapter 2. The mas gene was cloned in our laboratory by Rashmi. The C-terminal TE domain from module 6 of DEBS3 (referred to as DEBS-TE) was previously cloned (Gokhale et al., 1999a).

3.2.2.2 Expression and purification of mycocerosic acid synthase

The Mas protein was expressed in the soluble form by growing the cultures (BAP1 strain of *E. coli* carrying the plasmid pRT19) at 30°C till A600 reaches 0.6 and then inducing for 6-8 hrs at 22°C with 0.5 mM IPTG. The cells were harvested and the media was removed by washing with 50 mM Tris-HCl (pH: 8.0) buffer. For purification, the cells were disrupted by French Press in 100 mM phosphate buffer (pH: 7.0) containing 10% glycerol and 150 mM NaCl. After removal of cell debris, the lysate was loaded on to a Ni^{2+}-NTA column. The unbound proteins were removed by washing with wash buffer. The Mas protein was eluted with increasing concentration of imidazole in buffer. The protein containing fractions were pooled and loaded on to a ResourceQ column. The Mas protein was eluted from the anion exchange resin with increasing concentration of NaCl in buffer. The purification was monitored at 280 nm and the protein fractions were analysed on SDS-PAGE gel. Protein containing fractions were stored at -80°C till further use.
3.2.2.3 Peptide analysis by Mass Spectrometry

The in-gel trypsin digestion and peptide analysis has been described previously. 5 μg of the purified Mas protein was used for in-gel trypsin digestion. The peptides were purified by using POROS R2 resin using 25% ACN/0.2% formic acid and was directly used for nanospray mass spectrometric analysis.

3.2.2.4 Labelling of the Mas protein with starter and extender units

The specificity of the Mas protein for the extender unit was probed by incubating the purified protein with the radio-labelled dicarboxylic acids-\(^{14}\)C-malonyl-CoA and \(^{14}\)C-methyl malonyl-CoA at 30°C for 15 min. The protein was resolved on a 7% SDS-PAGE gel and the radiolabel associated with the protein was monitored by phosphoimager. The loading of the \(^{14}\)C-lauroyl-CoA starter unit was also monitored in a similar manner.

3.2.2.5 In vitro reconstitution of Mas activity

The purified recombinant protein was used for in vitro reactions. Reaction mixtures contained 100 mM phosphate buffer (pH: 7.0), 200 μM starter acyl-CoA, 200 μM \(^{14}\)C-methyl malonyl-CoA, 4 mM NADPH, 10% glycerol, 2 mM TCEP/DTT and 2.5 μM Mas protein in a 100 μl volume. Acyl-CoAs of varying chain lengths (C\(_{6}\)-C\(_{20}\)) were used in the reaction. Reactions were incubated at 30°C for 1 hr and subsequently quenched by adding 50 μl of 45% potassium hydroxide (KOH). The covalently bound products were released by heating at 100°C for 15 min in the presence of alkali. The reaction mixture was then acidified by addition of 100 μl of 50% hydrochloric acid (HCl) and products were
extracted using hexanes, twice. The solvent was dried using vacuum concentrator and the concentrated products were spotted on silica gel TLC plate. The TLC plate was developed using hexanes/diethyl ether/glacial acetic acid (50:50:1) and visualized by phosphorimager.

**3.2.2.6 Mass spectrometric analysis of products**

The radioactive Mas products were scraped from TLC and extracted twice with hexanes. After evaporating the solvent, the residue was dissolved in 90% ethanol and analysed by nanospray ESI-MS/MS in the negative ion mode.

**3.2.2.7 Purification of the TE proteins**

The following methodology was used for purification of the TesA, PKS6-TE and DEBS-TE proteins. The cells were disrupted in 50 mM phosphate buffer (pH: 7.4) containing 150 mM NaCl. The cell debris was removed by centrifugation at 50,000 g for 30 min at 4°C. The crude lysates was loaded on to Ni$^{2+}$-NTA agarose column. The resin was washed with wash buffer (50 mM phosphate pH: 7.0) till the contaminating proteins were removed. The protein was eluted using elution buffers containing increasing concentration of imidazole. 2 mM EDTA was added to the fractions containing the proteins of interest. The pooled fractions were loaded onto a ResourceQ (anion exchange) column. The proteins were eluted using a linear gradient of 0-1 M NaCl in buffer. The protein of interest was analysed using SDS-PAGE gel and was concentrated after pooling the fractions using a centriprep filter (10 kDa MWCO). The proteins were stored at -80°C for further use after adding 2 mM TCEP. For the spectrophotometric assays, the proteins were stored without addition of any reducing agent.
3.2.2.8 Spectrophotometric determination of the thioesterase activity

The thioesterase activity of TesA, PKS6-TE and DEBS-TE was measured spectrophotometrically at 412 nm with 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB) using acyl-NAC thioester substrates. Each reaction mixture contained 50 mM Tris-HCl (pH: 8.4), 2 mM acyl-NACs, 6.45 μM of TE proteins and 5 μl of a saturated solution of DTNB in 50 mM Tris-HCl (pH: 8.4) in a final volume of 600 μl. The reactions were initiated by addition of the protein and the rate of thiol formation was monitored for 15 min. The rates of hydrolysis were calculated from the initial linear portion of the curves. Control reactions without protein were set up in all the cases.

3.2.2.9 TesA mediated release of acyl chains

The Mas and the PpsA assays were carried out as described in 3.2.2.5 and 2.2.2.25 respectively using 14C-lauroyl-CoA as starter unit. After incubation at 30°C for 8 hrs, the TesA protein was added to the reaction in a ratio of 1:2 and further incubated at 37°C for 1 hr. For kinetic measurements, the reactions were incubated for various time intervals (5-180 min). Controls for non-catalytic hydrolysis were monitored by setting up reactions in the absence of the TE proteins. The reactions were acidified by adding 7 μl of 50% HCl and extracted with hexanes. The products were resolved on TLC plates using hexanes/diethyl ether/glacial acetic acid (50:50:1) as solvent. For PpsA reactions, the released products were extracted with ethyl acetate and the TLC plate developed using hexanes/ethyl acetate/glacial acetic acid (40:60:5) as solvent. Assays with DEBS-TE and PKS6-TE were also carried out in an analogous manner.
3.2.2.10 TesA catalyzed deacylation of Mas and PpsA proteins

The Mas and the PpsA proteins were incubated with 50 μM of $^{14}$C-lauroyl-CoA in 100 mM phosphate buffer (pH: 7.0), 10% glycerol at 30°C for 15 min. The radio-labelled proteins were then incubated with TesA at 30°C for 1 hr. The reaction was then loaded onto a 7% SDS-PAGE gel. After brief staining and destaining, the gel was treated with Amplify solution for 15 min. The gel was dried and the radioactivity was analysed by autoradiography using phosphoimager. Control reactions were carried out without TesA protein.

3.3 RESULTS

3.3.1 Expression and purification of mycocerosic acid synthase

The mas gene had been previously cloned in pET vector system for expression as a C-terminal hexahistidine-tagged protein in our laboratory. The plasmids were transformed in BAP1 strain of E. coli for expression in the holo form. The 232 kDa protein was expressed in soluble form by inducing the cultures with 0.5 mM IPTG at 22°C for 6 hr. The protein was purified by using Ni$^{2+}$-NTA affinity chromatography. The majority of the Mas protein was eluted in 20 and 50 mM imidazole containing buffer [100 mM phosphate (pH: 7.0) + 10% glycerol] as shown in Figure 3.1a (lanes 12-22). These fractions had several contaminating proteins of lower molecular weight. These were then pooled and further purified by anion exchange chromatography with increasing concentration of NaCl in buffer. The protein eluted in 40-50% B (400-500 mM NaCl) fractions (Figure 3.1b, lanes 9-18). The Mas protein was purified to ~95% homogeneity.
Figure 3.1: Purification of Mas protein
(a) SDS-PAGE analysis of affinity purification of Mas
1. Supernatant 7-11. 10 mM imidazole fractions
2. Pellet 12-16. 20 mM imidazole fractions
3. Flow through 17. Protein Marker
4. Protein Marker 18-21. 50 mM imidazole fractions
5-6. Wash fraction 22-24. 100 mM imidazole fractions

Figure 3.1: Purification of Mas protein
(b) Analysis of FPLC purification of Mas
1. Load 9. 45% B Fr 12
2. Flow through 10. 45% B Fr 14
3. Wash 11. 45% B Fr 16
4. 40% B Fr 6 12. 47% B Fr 18
5. 40% B Fr 6 13. 50% B Fr 21
6. Protein Marker 14. 50% B Fr 25
7. 42% B Fr 9 15. 60% B Fr 30
8. 45% B Fr 11
3.3.2 Peptide analysis by Mass Spectrometry

The Mas protein was excised from the gel and in-gel tryptic digestion was carried out to obtain the peptides for mass spectrometric analysis. Peptide mass fingerprinting was carried out to confirm the identity of the protein using the non-redundant database. Based on their masses, 19 peptides could be identified which covered 27% (570 out of 2110 amino acids) of the Mas protein. The most probable protein predicted was the Mas protein from *M. tuberculosis* (Figure 3.2).

![Figure 3.2: The amino acid sequence of the Mas protein.](image)

The protein was identified as the most probable protein using the peptides in the TOF-MS spectrum. A total of 570 amino acids out of 2110 were matched which covered 27.0% of the protein. The peptides identified in the tryptic digests are underlined.

3.3.3 Labelling of Mas protein

The specificity of the AT domain of Mas protein for the extender units was probed by incubating the protein with $^{14}$C-malonyl-CoA and $^{14}$C-methyl malonyl-CoA. The radiolabel was associated with the protein only in the case of $^{14}$C-
methyl malonyl-CoA but was not observed for $^{14}$C-malonyl-CoA (Figure 3.3, lanes 2 and 3).

![Figure 3.3: Labeling of Mas protein with $^{14}$C-lauroyl-CoA, $^{14}$C-malonyl-CoA and $^{14}$C-methyl malonyl-CoA. The radiolabel is associated with the protein when incubated with $^{14}$C-lauroyl-CoA or $^{14}$C-methyl malonyl-CoA (lanes 1, 2) and not with $^{14}$C-malonyl-CoA (lane 3).](image)

In order to probe whether lauroyl-CoA can be accepted as a starter unit, the protein was incubated with $^{14}$C-lauroyl-CoA at 30°C for 15 min and then analysed on SDS-PAGE gel. The radioactivity was associated with the protein (Figure 3.3, lane 1) which indicated that the lauroyl-CoA could be used as a starter unit.

### 3.3.4 Catalytic activity of the Mas protein

Previous studies have characterised Mas protein to be an iterative fatty acid synthase, which is involved in the synthesis of long chain methylated mycocerosic acids in *M. bovis* BCG (Rainwater and Kolattukudy, 1983; Rainwater and Kolattukudy, 1985) (Figure 3.4). The catalytic activity of the *M. tuberculosis* Mas protein was analysed by incubating with various starter (C₆-C₂₀) acyl-CoAs, radio-

![Figure 3.4: Reaction catalysed by Mas protein. Synthesis of mycocerosic acids involves four rounds of iterative condensation of methyl malonyl-CoA on to the starter acyl-CoA substrates. The Mas protein elongates a range of C₆-C₂₀ starter CoAs to synthesize methylated fatty acids.](image)
labelled methyl malonyl-CoA, NADPH and DTT as reducing agent. Since this protein lacked a thioesterase domain, the acyl chains could not be actively released. Therefore, the covalently bound products were chemically hydrolysed by alkali and analysed by radio-TLC. The radio-TLC for the extension of C₆ and C₁₂ starter units is shown in Figure 3.5a. It was interesting to note that three radioactive bands could be observed in each lane and the first assumption was that these bands correspond to the different cycles of iterations. Since KOH hydrolysis would release all the intermediates, it was likely that the products released by chemical hydrolysis would consist of heterogeneous mixtures. However, in the same solvent system, the linear chain fatty acids, lauric acid (C₁₂), myristic acid (C₁₄), palmitic acid (C₁₆) and stearic acid (C₁₈) could not be resolved. Since DTT is a thiol based reducing agent, it is possible that DTT can also be involved in the thioester exchange reactions with the products bound to the protein. TCEP, a non-thiol based reducing agent, which is active in a wider

![Figure 3.5: Synthesis of mycocerosic acids by Mas using different starter units (a) Mas products formed using hexanoyl-CoA and lauroyl-CoA as starter units. DTT was used as a reducing agent in the reaction. (b) The autoradiograph shows the Mas products formed by using C₆, C₁₂ and C₁₆ acyl-CoAs with TCEP as reducing agent](image)
range of pH was therefore substituted for DTT. Surprisingly, the enzymatic assays with Mas protein resulted in a single band on the radio-TLC. Small changes in the $R_f$ of the products were obtained for different starter acyl-CoA substrates (Figure 3.5b). The DTT used in previous assays formed adducts with the CoA analogues and had resulted in multiple radioactive bands.

Our experiment suggested that the Mas protein had relaxed specificity for incorporating the starter unit, which in turn could result in heterogeneity in the acyl chain lengths of mycocerosic acids, which have been identified from mycobacteria (Daffe et al., 1988).

### 3.3.5 Identification of the Mas products by mass spectrometry

Since the radio-TLC analysis cannot directly provide information regarding the number of iterative cycles, the heterogeneous mixture of Mas products was extracted from the TLC plates and characterised by ESI-MS in the negative ion mode. The molecular weight of the expected products along with the chemical structures obtained from one to four iterations are shown in Figure 3.6.

1. 200.17 Da
2. 242.21 Da
3. 284.26 Da
4. 326.31 Da
5. 368.35 Da

Figure 3.6: The chemical structures and the corresponding molecular weights of the Mas products. No. 1 is the lauric acid ($C_{12}$) which is the starter acyl chain. The products of each round of extension are shown as No. 2-5 along with the molecular weight.
Preliminary MS analysis showed peaks corresponding to 241.21 and 283.26 Da, which corresponded to the products of the first and the second iterative cycles catalysed by Mas protein. The background mass spectrum from the organic solvents interestingly showed all these peaks. Careful analysis of a number of organic solvents in the negative ion mode suggested contamination from lauric acid, myristic acid, palmitic acid and stearic acid. The molecular weight of these fatty acids overlaps with the molecular weight of the expected Mas catalysed products. Our best efforts to purify these solvents through multiple steps of distillation could not remove these minor contaminations. Interestingly, all the MS grade organic solvents available with various companies also showed similar contaminating peaks in the mass spectra. Therefore, we decided to overcome this problem by using $^{14}$C-radio-labelled starter unit. The TOF MS spectrum of the Mas catalysed extension of lauroyl-CoA starter unit primarily showed a molecular ion peak of $[\text{M-H}]^{-1}$ 325.28 Da (Figure 3.7a). This corresponded to the product obtained by the third iterative cycle catalysed by Mas product. Small amount of product for the fourth cycle could also be detected while the product of the second iterative cycle was completely masked by the contaminating stearic acid peak (283.21). It can be observed that peak corresponding to the $^{14}$C-isotope (326.28 Da) was considerably more populated than the $^{13}$C-isotopic peak (327.28 Da). The MS/MS fragmentation pattern of the molecular ion peak of 325.28 and 327.28 is shown in Figure 3.7b and c. The difference of 2 mass units was observed for the fragment ions. These peaks have a difference of 14 mass units that correspond to the $-\text{CH}_2$ units. These studies thereby characterised the Mas product suggesting that the lauroyl starter unit generates a product from three iterative cycles.
Trimethylated mycocerosic acids esterified to the phthiocerol have been previously identified in *M. tuberculosis* cell wall (Daffe et al., 1988).

![Figure 3.7: ESI-MS analysis of the Mas product in the negative ion mode. (a) TOF MS spectrum of the Mas product indicating the region between 322-329 Da. The molecular ions of the product formed by three iterative cycles along with the isotopic peaks are indicated in the spectrum. The $^{14}$C-isotopic peak is marked with an asterisk. (b) MS/MS profile of the molecular ion peak of 325.28 (c) MS/MS fragmentation profile of the molecular ion of the isotopic peak 327.28. A difference of 2 Da is observed in the fragments obtained in (b) and (c). The difference of 14 Da is due to the $-\text{CH}_2$ units of the fatty acyl chain.](image-url)
3.3.6 Kinetic analysis of the Mas protein

Since Mas protein does not possess an active mechanism of release of products, the rate constants estimated here provide an apparent kinetic characterization. Moreover, the Mas protein performs six different catalytic steps for each round of extension by 2C units, which further complicated the overall kinetics. In order to calculate the kinetic rate constant, the amount of product formed by the Mas protein was monitored over a period of 3 hrs using $^{14}$C-lauroyl-CoA as starter unit (Figure 3.8). The apparent $k_{cat}$ obtained from the initial linear region of the curve was estimated to be $0.08 \pm 0.01\text{ min}^{-1}$. The $K_m$ for the lauroyl-CoA and methyl malonyl-CoA could not be estimated due to poor saturation kinetics.

![Figure 3.8: Kinetic analysis of Mas protein. Amount of product formed by Mas protein monitored at different time points over a period of three hrs.](image)

3.3.7 Sequence analysis of various TE proteins

The thioesterase (TE) proteins are associated with PKS/FAS clusters and are important in releasing the final products. Most of these proteins are fused at the C-terminal end of such PKS/FAS modules and are referred to as type I TEs (TE I). In certain cases, standalone TE genes are present proximal to these FAS/PKS gene clusters. Such proteins are designated as type II TEs (TE II). Mycobacterial
The *pp* cluster contains *tesA* (TE II) gene and its function in the PDIM has not been investigated.

The *tesA* gene was present at the 5' end of the *pp* cluster. It was likely that the *tesA* product is involved in the biosynthetic pathway of PDIM esters. In order to understand the role of this protein we probed the activity of the TesA protein along with the TE domain of the PKS6. The PKS6-TE is present as a C-terminal domain in the PKS6 protein analogous to the chain releasing TE domains of polyketide synthases while the *tesA* is an independent ORF comparable to the type II TEs present in the NRPS/PKS gene clusters. The TesA protein sequence was compared with various Type II TE sequences from the NRPS or PKS clusters and our analysis showed that the TesA had high sequence homology with the other TE II proteins. This suggested that the TesA protein may belong to the α/β hydrolase family of proteins having the GHSXG active site motif (Holmquist, 2000) (Figure 3.9). In order to characterise its biochemical function, the catalytic activity of the TesA protein was compared with two other TE I (C-terminal domains of the DEBS and the PKS6) proteins.

![Figure 3.9: Comparison of the sequences of various TE II domains.](image)

The conserved active site residues are shown in blue. The conserved residues of the TE II are shown in green.
3.3.8 Cloning, expression and purification of recombinant TE proteins

3.3.8.1 TesA

The \textit{tesA} (Rv2928) gene along with an unknown ORF (Rv2929) was PCR amplified from the BAC genomic library of \textit{M. tuberculosis} H37Rv. \textit{Nde I} restriction sites were engineered at both the 5' and the 3' end and the amplicon was cloned in the \textit{EcoR V} sites of pBluescript-SK+ vector (pCR53). This \textit{Nde I} fragment was then cloned in the pET28c vector for expression as N-terminal hexa his-tagged protein (Strategy XI, Appendix A). Since a natural stop codon was present at the end of the \textit{tesA} gene, the induced expression provided only the TesA protein.

![Figure 3.10: SDS-PAGE gel for Ni\textsuperscript{2+}-NTA purification of TesA protein](image)

1. Pellet 6-8. 5 mM imidazole wash fractions
2. Supernatant 9-10. 15 mM imidazole wash fractions
3. Flow through 11-13. 50 mM imidazole elution fractions
4. Wash 14. Protein Marker
5. Protein Marker 15-20. 100 mM imidazole elution fractions

The TesA protein was expressed using the BL21-DE3 strain of \textit{E. coli} in the soluble form by inducing the culture with 0.5 mM IPTG at 27°C for 6-8 hrs. The protein was purified using Ni\textsuperscript{2+}-NTA affinity chromatography [50 mM
phosphate buffer (pH: 7.4)]. The TesA protein eluted at 50 and the 100 mM imidazole concentration (Figure 3.10, lanes 11-13 and 15-18). Since the protein was >90% pure, the imidazole was removed during concentration and the protein was used for further analysis.

3.3.8.2 PKS6-TE

The PCR amplified TE domain of the pks6 gene was cloned in the pBluescript-SK+ vector by blunt-ended ligation (pCR99). Spe I and EcoR I restriction sites were engineered at the 5' and the 3' ends. This fragment was then cloned in pET21c expression vector in the Nhe I and EcoR I restriction sites. The PKS6-TE protein was resilient to expression in the soluble form under most of the conditions. This protein could be obtained in the soluble form by growing uninduced cultures at 18°C for 24 hours. The protein was purified using affinity chromatography and eluted in buffer containing 100 mM imidazole (Figure 3.11, lanes 19-21). The fractions were concentrated and stored at -80°C till further use.

Figure 3.11: SDS-PAGE analysis of the affinity purification of PKS6-TE protein
1. Pellet 9-10. 10 mM imidazole wash fractions
2. Supernatant 11-13. 25 mM imidazole elution fractions
3. Flow through 14. Protein Marker
4. Wash 15-17. 50 mM imidazole elution fractions
5-7. 5 mM imidazole wash fractions 18-20. 100 mM imidazole elution fractions
8. Protein Marker
3.3.8.3 DEBS-TE

The DEBS-TE had been previously cloned and expressed in *E. coli* with a C-terminal his-tag (Gokhale et al., 1999a). The protein was expressed in the soluble form by inducing cultures at 30°C with 0.5 mM IPTG for 8 hrs. This protein was again purified by affinity chromatography and the ~ 33 kDa protein eluted in 50 and 100 mM imidazole containing fractions (Figure 3.12, lanes 14-24). The protein was concentrated for further characterization.

![Figure 3.12: Purification of the DEBS-TE protein by Ni²⁺-NTA chromatography](image)

1. Pellet 7-9. 5 mM imidazole wash fractions
2. Supernatant 10-11. 10 mM imidazole wash fractions
3. Supernatant 12-13. 25 mM imidazole elution fractions
4. Flow through 14-17. 50 mM imidazole elution fractions
5. Wash 18. Protein Marker
6. Protein Marker 19-24. 100 mM imidazole elution fractions

3.3.9 Catalytic activity of the various TE proteins

The catalytic activity of three TE proteins was monitored by using a spectrophotometric assay. The NAC-thioesters, which mimic the phosphopantetheine arm of the ACP domains, have been previously used as surrogate substrates to measure the activity of TE proteins (Gokhale et al., 1999a). The hydrolysis of the NAC-thioesters was coupled to Ellman’s reagent (DTNB) to detect the release of free thiols. DTNB reacts at neutral pH with sulfhydryl groups
to liberate the yellow compound 5-thio-2-nitrobenzoic acid which can be monitored by UV spectroscopy ($\lambda_{\text{max}}=412$ nm, $\varepsilon=13600$ M$^{-1}$cm$^{-1}$). The reaction was initiated by addition of the substrate and the reaction was monitored at 412 nm for 15 min. To decrease the background absorbance, the assays were carried out in absence of thiol reducing agents. The TE proteins used in this assays were purified in the absence of thiol reducing agents. These TE proteins have free thiols whose modification by DTNB did not affect the catalytic activity. Most of the NAC-thioesters used in this study were synthesized by Archana (Figure 3.13). The diketide-NAC (DK-NAC) was kindly provided by Prof. Chaitan Khosla.

![Chemical structures](image)

**Figure 3.13: Structures of various NAC-thioester substrates**

These NAC-thioesters have different functional groups in order to test the broad range of specificity of the TE proteins. The DK-NAC is the natural diketide synthesized by DEBS module 1 using a molecule of propionyl-CoA and methyl malonyl-CoA. The hexanoyl-NAC (H-NAC) and the 9-hydroxy decanoyl-NAC (9HD-NAC) had varying number of carbon chains whereas the benzoyl-NAC (B-NAC) contained an aromatic moiety. The H-NAC and 9HD-NAC represented the acyl-chains bound to the PKS and Mas proteins in *M. tuberculosis* since these
proteins have been shown to be involved in complex lipid biosynthesis. Because of the poor solubility of the decanoyl-NAC, lauroyl-NAC and palmitoyl-NAC, these substrates could only be used at very low concentrations and therefore resulted in poor catalytic activity. The hydrolytic activities of the three TE proteins are shown in Figure 3.14. The comparison of the rates of hydrolysis of various NAC-thioesters catalyzed by DEBS-TE is shown in Figure 3.14a. DEBS-TE showed maximum activity with 9HD-NAC with a \( k_{\text{cat}} \) value of 74.3 pmol mg\(^{-1}\) min\(^{-1}\) and suggested significant preference for the longer chain substrates. This Figure 3.14: Spectrophotometric analysis of the TE activity. The hydrolysis of different NAC thioesters catalysed by TE proteins monitored at 412 nm. (a) DEBS-TE; (b) PKS6-TE; (c) TesA; (d) Comparison of the hydrolytic activity of the three TE proteins
result is analogous to the previously reported studies with DEBS-TE (Gokhale et al., 1999a). Figure 3.14b indicated that the PKS6-TE showed similar substrate preference as DEBS-TE albeit with significantly higher rate of catalysis. The $k_{cat}$ values obtained for H-NAC and 9HD-NAC were 146.5 and 272.1 pmol mg$^{-1}$ min$^{-1}$ for PKS6-TE. The TesA protein showed much weaker activity in hydrolysis of these NAC-thioester substrates in comparison to the DEBS-TE and PKS6-TE (Figure 3.14c). H-NAC was the most preferred substrate for TesA protein ($k_{cat}$=28.33 pmol mg$^{-1}$ min$^{-1}$). The comparative chart for the three TE proteins for the four substrates tested here is shown in Figure 3.14d. These studies clearly showed that the TE proteins were catalytically competent to hydrolyse thioester substrates.

**3.3.10 Editing activity of TesA**

**3.3.10.1 Release of acyl chains bound to the Mas protein**

Since the pps cluster did not contain any TE I, we wanted to examine if the TesA protein could be involved in the release of covalently bound products. Towards this goal, the mycocerosic acids bound to the Mas protein were incubated with all the three TE proteins separately. The enzymatically hydrolysed products were extracted with hexanes and analysed by radio-TLC (Figure 3.15). The background hydrolysis is shown in lane 2, which did not contain any protein, and as a positive control, the chemical hydrolysis was performed with KOH (lane 1). It could be observed Figure 3.15 that whereas the PKS6-TE and DEBS-TE did not release any methylated fatty acids; TesA was able to release ~ 25% of these acyl chains from the Mas protein. The hydrolysis observed with PKS6-TE and the
DEBS-TE was comparable to the non-catalytic background hydrolysis (Figure 3.15, lanes 2, 4 and 5). These studies clearly demonstrated that the DEBS-TE and PKS-TE are incapable of hydrolysing protein bound intermediate. It is possible that TesA may be important in releasing mycocerosic acids from the Mas protein. Since we cannot discriminate between different iterative products of the Mas protein, it was not clear whether the TesA selectively releases the final products. In order to optimize the hydrolysis of mycocerosic acids, four different molar ratios of the Mas-TesA were used. Two-fold and 10-fold excess of TesA did not dramatically enhance the TesA catalysed hydrolytic activity.

![Figure 3.15: Release of Mas products by TesA, PKS6-TE and DEBS-TE](image)

1. Mas with KOH hydrolysis
2. Mas without KOH hydrolysis
3. Mas incubated with TesA
4. Mas incubated with PKS6-TE
5. Mas incubated with DEBS-TE

We therefore followed the time kinetics of TesA mediated release of the acyl chains over a longer period of time. Under our experimental condition, at best one-third of the acyl chains could be released from the Mas protein (Figure 3.16). The maximum release of acyl chains was obtained with alkali hydrolysis. Y-axis
in Figure 3.16 denotes the percentage of TesA catalysed release of products in comparison with the chemical hydrolysis.

![Graph showing release of Mas product by TesA as a function of time. The amount of Mas product released by TesA is plotted as percentage of the total acyl chains bound to the Mas protein (calculated by radio-labeled product released by alkaline hydrolysis) The ratio of TesA:Mas protein was 2:1.]

3.3.10.2 Release of the acyl chains from PpsA

TE II proteins from other PKS/NRPS systems are proposed to release misprimed and stalled acyl chains from different proteins. To analyse whether the TesA would be able to release the acyl chains from PpsA as well, the TesA protein was added to the PpsA reaction mix. The PpsA reaction was carried out using $^{14}$C-lauroyl-CoA and malonyl-CoA as starter and extender units. PpsA catalyses the formation of the 3-hydroxy tetradecanoic acid, which is an intermediate in the synthesis of phthiocerol. Analysis of the ethyl acetate extracts on the TLC indicated that the TesA also able to catalyse the release of the bound acyl chain from PpsA. In Figure 3.17, the radioactive signals corresponding to $^{14}$C-lauric acid and the PpsA product are indicated. The top band above the PpsA product observed in the lanes 2 and 3 arises due to spontaneous hydrolysis of the lauroyl-CoA during the incubation. Although the formation of 3-hydroxy tetradecanoic acid is the final product of PpsA protein, this would be an intermediate in the
biosynthesis of PDIM. This data supports the hypothesis that the \textit{tesA} gene product might play an editing role and remove the aberrant intermediates blocking the synthesis.

![Figure 3.17: Monitoring the release of PpsA product by TesA.](image)

1. PpsA product released by KOH hydrolysis
2. PpsA without KOH hydrolysis
3. PpsA incubated with TesA

### 3.3.10.3 Thioesterase activity of TesA on acylated Mas and PpsA proteins

In order to probe whether the TesA is able to release the stalled starter units from the proteins, the lauroyl acyl chains were loaded onto the KS of the Mas and PpsA proteins using $^{14}$C-lauroyl-CoA. In the absence of the extender units in the reaction mixture, these acyl chains would not be transferred between the catalytic domains. This covalently acylated Mas and PpsA proteins were then incubated with or without TesA at 30°C for 1 hr. The reaction was then analysed by SDS-PAGE and autoradiography. The reaction without TesA showed a radioactive protein band associated with the Mas and the PpsA proteins, whereas the radioactive signal in the reactions containing TesA was dramatically reduced (Figure 3.18). These studies therefore clearly demonstrated that even the KS
bound intermediates could be released by TesA. The hydrolysis of the stalled lauroyl chain from both the proteins provided evidence of the thioesterase activity of TesA on the acylated Mas and PpsA proteins. Overall, our studies with TesA protein suggest a similar editing function, as observed previously for other TE II proteins.

![Figure 3.18: Release of the stalled acyl chains from Mas and PpsA proteins.](image)

**Figure 3.18:** Release of the stalled acyl chains from Mas and PpsA proteins. The Mas and the PpsA proteins were acylated at the KS domain with the lauroyl unit using $^{14}$C-lauroyl-CoA. The radioactivity associated with the Mas and PpsA proteins is drastically reduced on incubation with TesA.

### 3.4 DISCUSSION

Stereochemical studies with the cell wall glycolipids from various mycobacterial strains suggested differences in the fatty acid composition. *M. marinum* and *M. ulcerans* produce dextrorotatory methylated fatty acids called the phthioceranic acids while the *M. tuberculosis*, *M. leprae* and *M. bovis* possess laevorotatory mycocerosic acids (Daffe and Laneelle, 1988). The significance of the differences in the stereochemistry of these methyl branched chiral centres in the fatty acids is not clear.

In this study, we have reconstituted the cell-free synthesis of methylated fatty acids catalysed by the Mas protein. Since Mas protein is an iterative protein
and does not contain a thioesterase domain to release products, these metabolites were released by chemical hydrolysis. The Mas protein is believed to carry out four steps of iterations. However, the mass spectrometric analysis revealed products with various degree of chain extension. The major product consisted of 2,4,6-trimethylhexadecanoic acid which would be formed by extension of the lauroyl starter unit. The analysis of these methylated fatty acids from the phthiocerol esters of *M. tuberculosis* and *M. bovis* has been reported earlier. The hydrolysis of the phthiocerol esters produced 2D,4D,6D-trimethyltetraicosanoic acid and 2D,4D,6D-trimethylhexacosanoic acid which would be synthesized from three iterations of C18 and C20 fatty acids with methyl malonyl-CoA (Daffe et al., 1988). The *in vitro* synthesis of the polymethylated fatty acids by *M. tuberculosis* Mas protein would provide a good opportunity to dissect the stereochemical preferences of the Mas protein.

The *mas* gene is present in the *pps* cluster, which also contains a thioesterase like gene (*tesA*). Since none of the multienzymatic proteins in the *pps* cluster contain a TE domain, we decided to probe whether the TesA would be able to specifically interact with Pps and Mas proteins to release the phthiocerol and mycocerosic acids respectively. Our studies showed that TesA protein released the acyl chains from the Mas and PpsA proteins from the KS domain as well as the P-pant arm of the ACP domains. Since the final products are acylated at the P-pant arm of the ACP domain, the hydrolytic release of the chains bound to the Cys of the KS domain would be a futile process. Interestingly, previous studies have demonstrated that TE II proteins in the PKS and NRPS gene clusters may serve as an ‘editing’ enzyme and reactivate these modules by removing the acyl moieties.
attached to ACP/PCP domains (Butler et al., 1999; Schneider and Marahiel, 1998). More recently, *in vitro* studies have shown that the TE II can regenerate misacylated proteins generated by erroneous priming or stalled enzymatic machinery resulting from aberrant intermediates produced during the synthesis (Kim et al., 2002; Schwarzer et al., 2002; Yeh et al., 2004). Gene inactivation studies of type II TEs demonstrated that the loss of the TE II function resulted in a drastic reduction in the production of the final polyketides or the non-ribosomal peptides (Butler et al., 1999; Doi-Katayama et al., 2000; Reimmann et al., 2004; Schneider and Marahiel, 1998). The co-expression of the *ery-ORF5* TE II gene along with the DEBS genes in heterologous hosts, *S. coelicolor* and *E. coli* doubled the production of the 6-dEB and decreased the formation of 15-nor-dEB synthesized due to incorporation of the acetyl starter unit (Hu et al., 2003), providing support for the editing mechanism. In collaboration with Dr. J. S. Cox’s group, the inactivation of *tesA* gene in *M. tuberculosis* resulted in a dramatic decrease in the PDIM present in the cell wall of the mutants. The complementation of the *tesA* gene *in trans* restored the synthesis of PDIM demonstrating the importance of the type II thioesterase in the PDIM biosynthesis. The two TE I proteins (DEBS-TE and PKS-TE) did not show any hydrolytic activity with multifunctional proteins. All the three proteins were able to hydrolyse surrogate NAC-thioester substrates. Therefore, we believe that the TesA protein in this cluster may be performing similar editing function as the previously characterised TE II proteins. We thereby demonstrate that the TesA plays a significant role in releasing the aberrant chains and is not involved in the direct release of the final products in the synthesis of PDIM.