THE CATABOLISM OF THE LAST TWO RINGS OF CHOLESTEROL BY *MYCOBACTERIUM TUBERCULOSIS*

by

Adam Michael Crowe

B.Sc., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2018

© Adam Michael Crowe, 2018

Abstract

Cholesterol, a four-ringed steroid with an alkyl side chain, is an important growth substrate for *Mycobacterium tuberculosis* (*Mtb*) during infection. Many aspects of this catabolism remain unknown although steroid catabolism is a defining feature of mycobacteria and the related rhodococci. Using a variety of approaches, I elucidated key aspects of cholesterol catabolism in *Mtb* and other bacteria, particularly with respect to 3a*α*-*H*-4*α*(3'-propanoate)-7a*β*-methylhexahydro-1,5-indane-dione (HIP), a metabolite that contains the last two steroid rings (C/D). Chapter 2 demonstrates that the first two steroid rings (A/B) are degraded prior to the side chain in mycobacteria and rhodococci. This was established by targeting HsaD, the final ings A/B-degrading enzyme. Thus, a Δ*hsaD* mutant of *Rhodococcus jostii* RHA1 accumulated cholesterol-derived catabolites with partially degraded side-chains. Moreover, HsaD from *Mtb* had 100-fold higher specificity (k_{cat}/K_M) for a metabolite with a partially-degraded side chain. Chapter 3 presents a mechanism for KstR2, a TetR family transcriptional repressor that regulates the HIP catabolic genes, including *ipdABCF* and *echA20*. The KstR2 dimer bound two equivalents of HIP-CoA with high affinity (K_D = 80±10 nM). Crystallographic analyses revealed that HIP-CoA binding induces conformational changes in the dimer that preclude DNA-binding. Mutagenesis substantiated the roles of Arg162 and Trp166 in HIP-CoA binding. In Chapter 4, key HIP catabolic steps are elucidated. Two previously undescribed metabolites, 3a*α*-*H*-4*α*(carboxyl-CoA)-5-hydroxy-7a*β*-methylhexahydro-1-indanone (5-OH-HIC-CoA) and (*R*)-2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA (COCHEA-CoA) were identified using deletion mutants of *ipdC* and *ipdAB*, respectively, combined with novel metabolomics approaches. Together, purified IpdC, IpdF and EchA20 transformed 5-OH-HIC-CoA to COCHEA-CoA. These data, along with those from additional mutants, were used to formulate a

HIP catabolic pathway and to predict that cholesterol catabolism yields four propionyl-CoA, four acetyl-CoA, one pyruvate, and one succinate. Chapter 5 establishes that IpdAB catalyzes a retro-Claisen-like ring-opening of COCHEA-CoA (k_{cat}/K_M = 2±0.7 × 10^5 $M^{-1}s^{-1}$) despite structural similarity with Class I CoA transferases. Based on crystal structures of IpdAB and biochemical data, a mechanism for ring-cleavage is proposed in which conserved Glu105 acts as a catalytic base. Overall, this work significantly advances our understanding of bacterial steroid catabolism and facilitates the development of novel therapeutics to treat TB.

Lay Summary

The ability of *Mycobacterium tuberculosis* (*Mtb*) to cause tuberculosis depends on the pathogen's ability to degrade host-derived cholesterol, a 4-ringed steroid with a side chain. I used a variety of techniques to provide novel insights into the degradation of cholesterol in *Mtb* and related bacteria. Firstly, I demonstrated that the first two cholesterol rings are degraded prior to the side chain in *Mtb* and related bacteria. Secondly, I described how a two-ringed cholesterol metabolite regulates cholesterol degradation genes. Thirdly, I elucidated the pathway by which the last two cholesterol rings are degraded in *Mtb* and related bacteria. Finally, I established that IpdAB, a virulence factor in Mtb, cleaves open the last cholesterol ring and proposed a mechanism for how the enzyme does this. This dissertation provides insight into bacterial steroid degradation and facilitates the development of novel therapeutics to treat TB.

Preface

Chapter 1: Figure 1.5 and Figure 1.6 were adapted from Capyk *et al.* (2012) and Griffin *et al.* (2012), respectively, with permission from appropriate sources. Portions of the introductory text are used with permission from Crowe *et al.* (2015) of which I am first author. I created Table 1.1, modifying it to summarize data from Capyk *et al.* (2012), Griffin *et al.* (2012), Nesbitt *et al.* (2010), and Pandey *et al.* (2008).

Chapter 2: The contents of this chapter were prepared for submission in 2018 as: 'Crowe, A., Brown, K., Casabon, I., Kulkarani, J., Yam, K., and Eltis, L.D. The unusual convergence of steroid catabolic pathways in *Mycobacterium abscessus*'.

Sections, 2.2.4, and 2.3.3 were performed and written by KB. NMR structure elucidation of DHSBNC was performed by KY. Production of CoA thioesterified substrates was aided by IC. All other sections were performed by myself and written by LE and myself. All work presented in Chapter 2 was performed at The University of British Columbia, Vancouver, Canada.

Chapter 3: A version of this material has been published as: 'Crowe, A., Stogios, P., Casabon, I., Evdokimova, E., Savchenko, A., and Eltis, L.D. (2015). Structural and functional characterization of a ketosteroid transcriptional regulator of *Mycobacterium tuberculosis. Journal of Biological Chemistry.* 290(2): 872-882'. Experiments relating to the crystallization of KstR2 and structure refinement were performed by PS and EE at the University of Toronto, Toronto, Canada. HIP-CoA for crystallography was produced by IC. Sections 3.2.5 and 3.3.2 were performed and written by PS and EE. All other experiments were performed by myself and all other sections were written by LE and myself.

Chapter 4: A version of this material has been published as: 'Crowe, A., Casabon, I., Brown, K., Liu, J., Lian, J., Rogalski, J., Hurst, T., Snieckus, V., and Eltis, L.D. (2017).

Catabolism of the last two steroid rings in *Mycobacterium tuberculosis* and other bacteria. *mBio* 8(2) 00321-17'. Design and optimization of extraction of CoA thioester metabolites and much of the LC-MS analysis were performed by IC with help from JR (The University of British Columbia, Vancouver, Canada). Gene deletion and growth of *Mtb* was performed by KB and JL (The University of British Columbia, Vancouver, Canada). Macrophage infection studies were performed by KB. 5α-OH-HIC and 5β-OH-HIC were chemically synthesized by TH and VS at Queen's University, Kingston, Canada. All other experiments were performed by myself. Section 4.2.4 was written by KB. All other sections were written by LE and myself.

Chapter 5: The contents of this chapter were prepared for submission in 2018 as: 'Crowe, A., Workman, S., Worrall, L., Watanabe, N., Casabon, I, Strynadka, N., and Eltis L.D. IpdAB, a virulence factor in *Mycobacterium tuberculosis*, is a cholesterol ring-cleaving hydrolase'. Experiments and figures relating to the refinement of X-ray crystal structures were performed by SW, LW, and NW at The University of British Columbia, Vancouver, Canada. Intact protein mass spectrometry was performed at The Center for High-Throughput Biology, The University of British Columbia. SW, LW, and NS were the primary authors of Sections 5.2.7 and 5.2.8 and made significant contributions to the writing and analysis described in Sections 5.3.4, 5.3.5, 5.3.6, and 6.4.1. All other work was performed by myself and all other sections were written by LE and myself.

Chapter 6: This material was adapted from the publications listed above for Chapters 2-5. All sections were written by LE and myself.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4-AD	4-androstene-3,17-dione
5OH-HIC	3a*α*-*H*-4*α*(carboxyl-CoA)-5-hydroxy-7a*β*-methylhexahydro-1-indanone
ADD	1,4-androstadiene-3,17-dione
BLAST	Basic Local Alignment Search Tool
CoA	Coenzyme A
COCHEA-CoA	(*R*)-2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA
DHSA	3,4-Dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione
DHSBNC	3,4-dihydroxy- 17-isopropionoyl- 9,10-seconandrost-1,3,5(10)-triene-9-one
DSHA	4,5–9,10-diseco-3-hydroxy-5,9,17-tri-oxoandrosta-1(10),2-diene-4-oic acid
DSHBNC	4,5-9,10-diseco-3-hydroxy-5,9-dioxo-23,24-bisnorchola-1(10),2-dien-4,22-dioate
FMN	Flavin Mononucleotide
FPLC	Fast Protein Liquid Chromatography
GC-MS	Gas Chromatography coupled Mass Spectrometry
HIP	3a*α*-*H*-4*α*(3'-propanoate)-7a*β*-methylhexahydro-1,5-indanedione
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny Broth
LC-MS	Liquid Chromatography coupled Mass Spectrometry
M9G	M9 basal salts with Ca/Mg/Thiamine and Goodies Mix Media
MOODA	4-methyl-5-oxo-octanedioc acid
Mtb	*Mycobacterium tuberculosis*
Mab	*Mycobacterium abscessus*
NAD^+	Nicotinamide adenine dinucleotide (oxidized)

NADH	Nicotinamide Adenine Dinucleotide (reduced)
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
OD_{600}	Optical density at 600 nm
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
r.m.s.d	Root Mean Square Distance
RHA1	*Rhodococcus jostii* RHA1
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Acknowledgements

I would like to thank my supervisor, Prof. Lindsay D. Eltis for his patience and insights into the work presented herein. I am ingratiated for the research opportunities and learning experiences that you have provided me with.

Both my committee members, Prof. Natalie Strynadka and Prof. Martin Tanner have been an invaluable source of wisdom throughout my research.

I would like to thank all of the current and previous members of the Eltis Lab: Dr. Jenna Capyk, Dr. Israel Casabon, Dr. Nicolas Seghezzi, Dr. Hiroshi Otani, Dr. Rahul Singh, Dr. Antonio Ruzzini, Dr. Morgan Fetherolf, Jonathan Penfield, Carlos Diaz-Salazar, Eugene Kuatsjah, Keith Story, Jie Liu, and Jennifer Lian, for their expertise and advice. I would like to especially thank Kirstin Brown, James Round, and Raphael Roccor for their insights and friendship throughout my years in the Eltis Lab.

I have been lucky to collaborate with many individuals. I would like to thank Dr. Leonard Foster (UBC), Dr. William Mohn (UBC), Dr. Victor Snieckus (Queen's University, Kingston, ON), Dr. Alexei Savchenko (The University of Toronto, ON), Dr. Timothy Hurst (Queen's University), Dr. Peter Stogios (The University of Toronto), Dr. Liam Worrall (UBC), Dr. Nobu Watanabe, Dr. Sean Lott (University of Auckland, Auckland, New Zealand), and Sean Workman (UBC). I would especially like to thank Mark Okon (UBC) and Jason Rogalski (UBC) for their aid with NMR and LC-MS, respectively.

I would like to acknowledge the financial support from the Canadian Institute of Health Research (CIHR), The BC Lung Association, and The University of British Columbia.

Lastly, I would like to thank my loving wife, Amanda Deacon, for her patience over the years. Her continued support has eased the trials and tribulations of my PhD. I could not have succeeded without you.

Chapter 1: **Introduction**

1.1 Steroids

Steroids are polycyclic organic molecules with diverse functions across eukaryotes and prokaryotes [1]. The core steroid structure consists of three cyclohexane rings (Rings A, B, and C) and a single cyclopentane ring (Ring D). As exemplified in Figure 1.1, this core domain can be modified to include a branched alkyl side chain at carbon 17 (C-17), as with cholesterol, β-sitosterol, and lanosterol; a hydroxylated alkyl side chain at C-17, as with cholic acid and progesterones; or only hydroxylation at C-17, as with the androgens [1]. Similarly, the cycloalkyl rings can display multiple hydroxylations, as with cholic acids; a single hydroxylation at C-3, as with cholesterol, lanosterol, and β-sitosterol; or a C-3 oxo, as with the androgens and progesterones [1]. Additional chemical modifications include aromatization of Ring A, as with the estrogens, and desaturations, as found in many steroids [1].

Animals, plants, and fungi use steroids for a number of vital functions. In animals, cholesterol can be enzymatically modified to yield a broad range of steroid hormones. In humans, these include the sex hormones, testosterone, estradiol and progesterone, as well as the mineralocorticoids and glucocorticoids. Bile acids, derived from cholic acid, are utilized in the emulsification of fats within the small intestine. Across all animals, cholesterol is employed to maintain membrane fluidity. In fungi, membrane fluidity is maintained by ergosterol [2]. Plant steroids, referred to as phytosterols, similarly act as signaling molecules and maintain membrane fluidity [3]. Due to their ubiquity in nature, the decomposition of biomass and excreta results in an abundance of environmental steroids. This bioavailability of steroids permits their utilization as growth substrates by microorganisms.

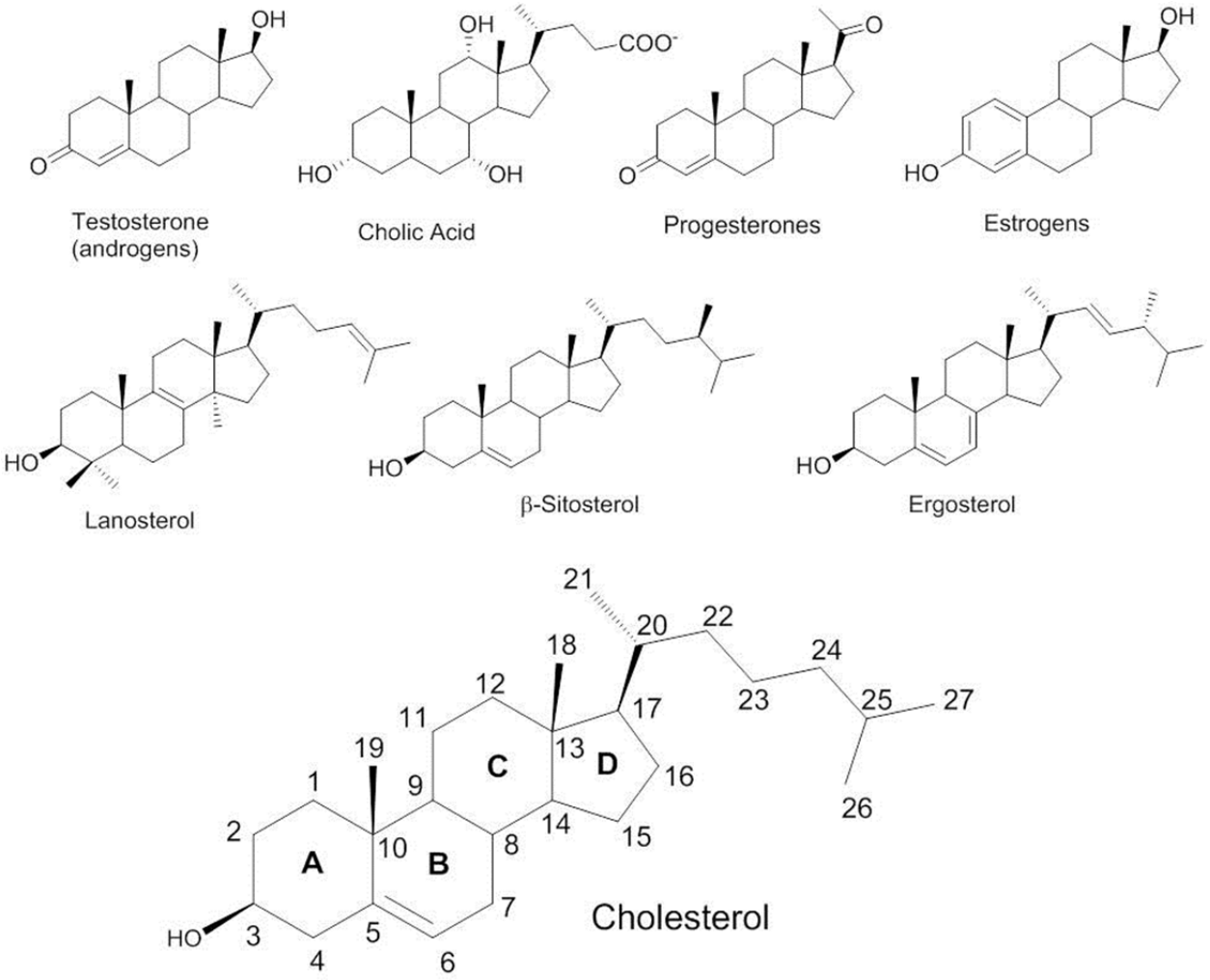


Figure 1.1. Chemical structures of representative steroids.
Cholesterol is displayed at the bottom with steroidal Rings A to D identified as well as the numbering of carbon atoms employed in this thesis when referring to cholesterol.

1.1.1 Cholesterol

The steroid, cholesterol, is characterized by an eight carbon alkyl side chain at C-17, an (S)-hydroxyl at C-3, and a carbon-carbon double bond between C-5 and C-6 (Figure 1.1). First isolated in 1815 by the French chemist M.E. Chevreul from human gall stones, cholesterol consists of up to 45 mol% of total lipids within mammalian cell membranes, making it the most abundant steroid in mammals. Up to 90% of cellular cholesterol is located within cell

membranes in mammals [4]. The C-3 hydroxyl of cholesterol makes the molecule slightly amphipathic permitting polar Van der Waals interactions with the phosphatidyl head groups of membrane lipids, thereby embedding the cholesterol molecule in an ordered orientation within cell membranes. This packing within membrane lipids results in increased membrane fluidity. However, cholesterol distribution within cell membranes is not uniformly distributed, but rather coalesces into regions high in sphingolipids and lipoproteins known as lipid rafts. Consisting of more than 100 lipoproteins, lipid rafts are thought to bring receptors and secondary- messenger molecules into close proximity and have been implicated in T-cell activation [5], cell adhesion [6], and calcium uptake [6]. Depletion of cholesterol from lipid rafts results in the dissociation of lipoproteins and has been shown to disrupt signaling pathways in T lymphocytes [7], enhance Influenza A release from infected cells [8], and decrease fertilization rates in oocytes [9].

De novo steroid biosynthesis (Figure 1.2) is ubiquitous across animals, fungi, and plants and utilizes a similar anabolic pathway [10]. In mammals, for example, cholesterol synthesis starts with the production of isoprenoid units from acetyl-Coenzyme A (AcCoA). These isoprenoid units are converted to isopentyl-5-pyrophosphate and dimethylallylpyrophosphate via the mevalonate pathway, which are condensed to form farnesyl pyrophosphate, geranyl phosphate, and finally, the 30-carbon molecule squalene. Squalene oxygenase incorporates one atom from dioxygen into squalene to produce an epoxide, (3*S*)-2,3-oxidosqualene. Cyclization of squalene epoxide by oxidosqualene cyclase produces lanosterol which is modified by a series of oxygenases to form cholesterol. In fungi, lanosterol is converted into ergosterol. Phytosterols are synthesized from cycloartenol in plants, which is also generated from squalene epoxide.

Although *de novo* cholesterol biosynthesis and dietary sources lead to blood cholesterol levels of 5-6 mM in healthy human adults [11], humans, like all animals, are unable to degrade

cholesterol and other steroids. Instead, human livers convert up to 1 g of cholesterol to bile acids per day which are then excreted via the small intestine [11].

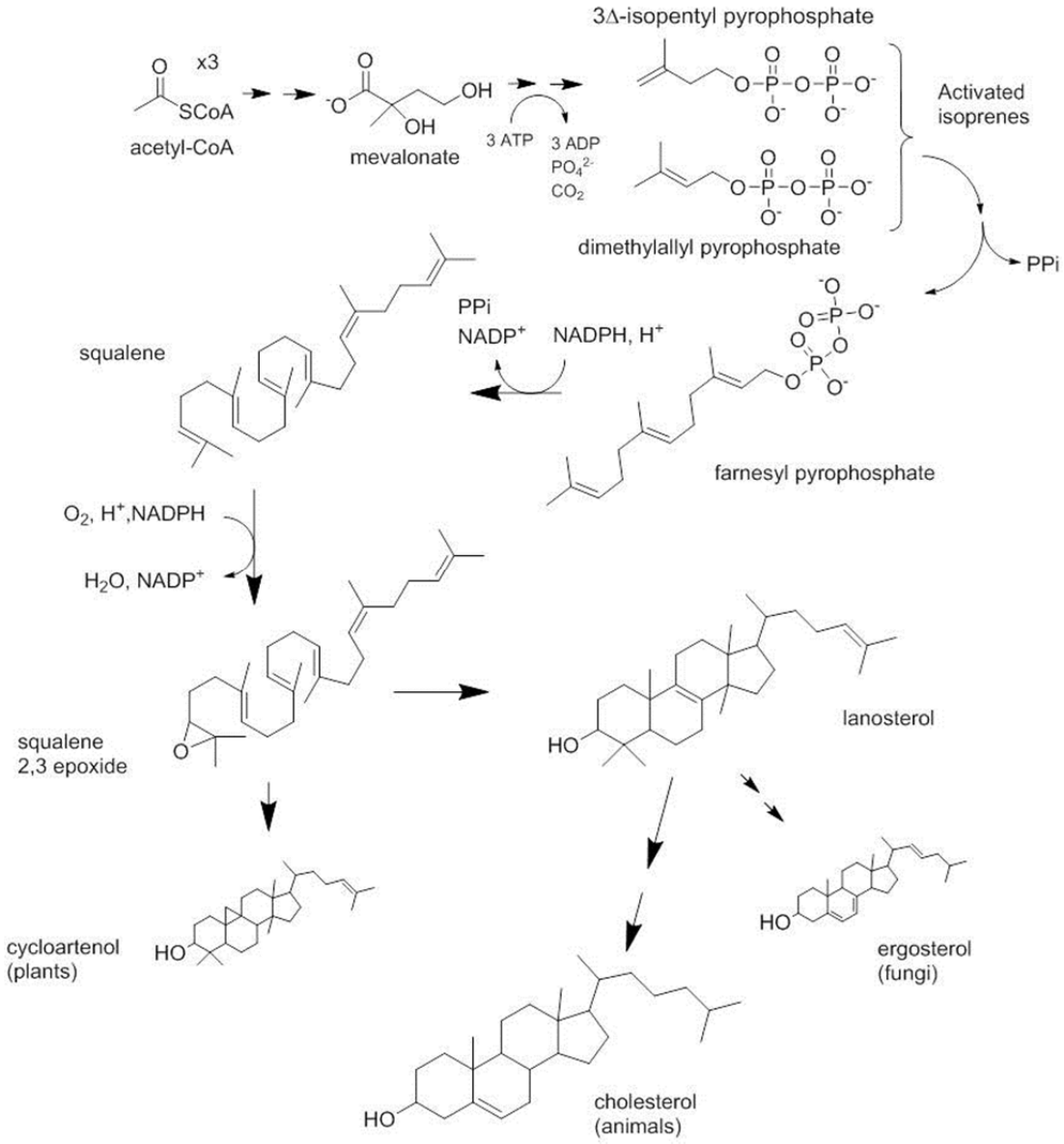


Figure 1.2. The cholesterol biosynthetic pathway in animals.

1.2 Bacterial metabolism of steroids

Despite the ubiquity of steroid biosynthesis across animals, plants, and fungi, only a handful of prokaryotes have been demonstrated to undergo *de novo* steroid synthesis. Prokaryotic steroid biosynthesizers include *Gemmata obsuriglobus* [12], *Methylococcus capsulatus* [13], and some myxobacteria [14]. However, advances in metagenomic analyses have led to the identification of oxidosqualene cyclase genes in 34 bacterial genomes from 5 phyla including Bacteroidetes, Cyanobacteria, Planctomycetes, Proteobacteria, and Verrucomicrobia, suggesting that bacterial steroid biosynthesis may be more wide spread than initially thought [15]. Steroid biosynthesis has not been observed in archaea [16].

In contrast to steroid biosynthesis, steroid catabolism is wide spread across at least two bacterial phyla: Actinobacteria and Proteobacteria [17]. Notable steroid-degrading bacteria include *Commamonas testosteroni* TA441 [18], *Nocardia restrictius* [19], *Steroidobacter denitrificans* [20], *Pseudomonas putida* DOC21 [21], and multiple species of rhodococci [22-24] and mycobacteria [25-27]. Figure 1.3 shows the evolutionary relationship between notable steroid-degrading bacteria as characterized by their 16S rRNA sequences.

Bacterial steroid catabolism was first extensively studied in *Nocardia* and *Mycobacterium* sp. by Sih and Lee between 1960 and 1970 [28, 29]. Early observations noted that *Nocardia* were able to grow on cholesterol as a sole carbon source and, in the process, excrete cholesterol derived metabolites into the culture media [30]. Since the 1970s, studies into bacterial steroid catabolism have focused on *Comamonas testosteroni* [18, 31], *Rhodococcus jostii* RHA1 [22, 25], and *Mycobacterium smegmatis* MC^2155 [26, 32], due, in part, to their potential biotechnological applications of steroid modification [26]. However, in 2007 following the publication of the genome from RHA1, an homologous cholesterol catabolic gene cluster to

that of RHA1 was identified in the human pathogen, *Mycobacterium tuberculosis* (*Mtb*) [25]. Since 2007, interest into steroid catabolism has emerged due to its correlation with *Mtb* pathogenesis, as described in Section 1.3.3 [33].

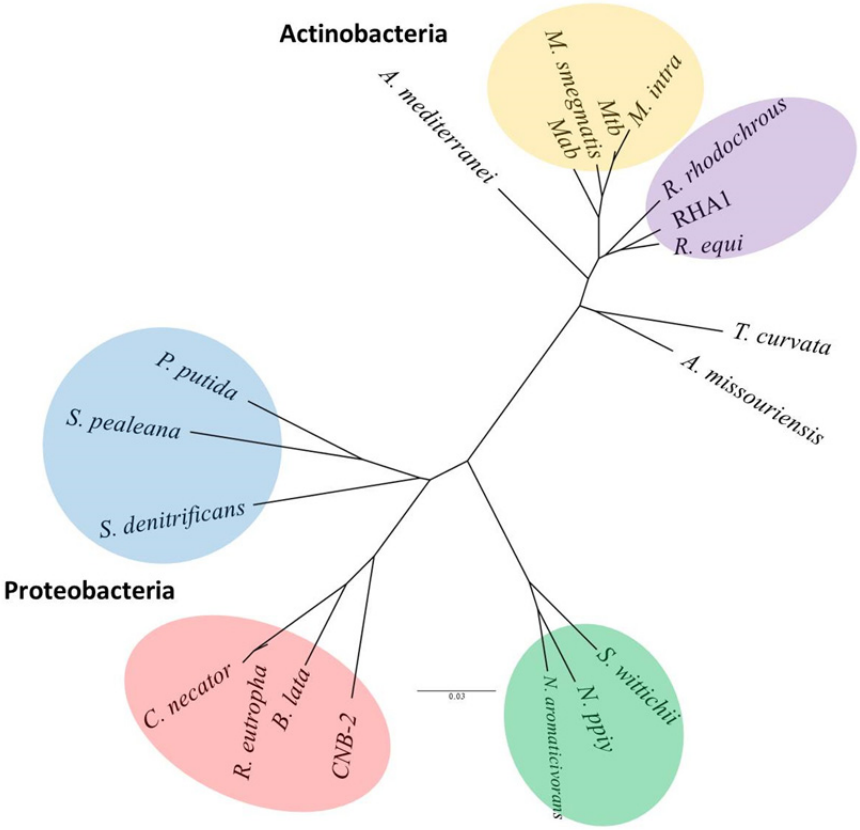


Figure 1.3. Evolutionary relationship between steroid degrading bacteria.
Phylogeny was analyzed from an alignment of 16S rRNA sequences. Yellow, purple, blue, red, and green highlighted regions indicate bacteria belonging to Myobacteria, Rhodococci, Gamma-proteobacteria, Beta-proteobacteria, and Alpha-proteobacteria, respectively. Species are identified as follows: *Amycolatopsis mediterranei* RB (A_mediterranei), *Mycobacterium abscessus* ATCC19977 (*Mab*), *Mycobacteria smegmatis* mc^{2}155 (M_smegmatis), *Mycobacterium tuberculosis* H37Rv (Mtb), *Mycobacterium intracellulare* 1956 (M_intra), *Rhodococcus rhodochrous* (R_rhodochrous), *Rhodococcus jostii* RHA1 (RHA1), *Rhodococcus equi* (R_equi), *Thermomonospora curvata* DSM 43183 (T_curvata), *Actinoplanes missouriensis* 431 (A_missouriensis), *Sphingomonas wittichii* RW (S_wittichii), *Novosphingobium* sp. PP1Y (N_PP1Y), *Novosphingobium aromaticivorans* DSM 12444 (N_aromaticivorans), *Comamonas tesosteroni* TA441 (CNB-2), *Burkholderia lata* (B_lata), *Ralstonia eutropha* H16 (R_eutropha), *Cupriavidus necator* N-1 (C_necator), *Steroidobacter denitrificans* DSM18526 (S_denitrificans), *Shewanella pealeana* ATCC 700345 (S_pealeana), and *Pseudomonas putida* DOC21 (P_putida). Figure adapted from Bergstand *et al.* (2016) [17].

Both aerobic and anaerobic steroid catabolism have been described. Anaerobic catabolism is best studied in *Steroidobacter denitrificans* DSMZ18526 [20, 34, 35]. Anaerobic testosterone catabolism occurs in a primarily β-oxidative manner in which steroid Rings A and B are opened by a series of hydroxylations and retro-Claisen like ring opening reactions [34]. 17-Hydroxy-1-oxo-2,3-*seco*-androstand-3-oic acid (2,3-SAOA) is diagnostic of the anaerobic pathway as it is not observed in the aerobic pathway [36].

In contrast, aerobic steroid catabolism is well characterized in both proteobacteria and actinobacteria. All aerobic pathways characterized to date are homologous, although specific pathways have divergently evolved to accommodate particular types of steroids. In each case, the catabolism involves up to three distinct processes: β-oxidation of the alkyl side chain, steroid Rings A and B opening by oxygenases, and catabolism of steroid Rings C and D. Each of these processes will be described at length in Sections 1.2.3 – 1.2.5, respectively. Steroid catabolic pathways have been characterized for cholate [18, 22, 31], androgens [18, 37, 38], and cholesterol [25, 39, 40]. Proteobacteria, such as *C. testosteroni*, are able to grow on testosterone and cholate, although typically not on cholesterol [18]. Actinobacteria, in contrast, often contain multiple steroid catabolic gene clusters. For example, RHA1 contains four such gene clusters [22], encoding for the degradation of cholesterol, cholate, androgens (predicted), and an undetermined steroid. Similarly, *R. rhodochrous* contains up to five homologs of steroid catabolic genes [41, 42], and *M. smegmatis* possess at least two [26]. Additionally, gene redundancy in steroid catabolic pathways permits access of steroid-degrading bacteria to modified steroids, as exemplified by *Mtb*'s ability to oxidize both cholesterol and cholesterol esters [43, 44].

1.2.1 Cholesterol catabolic pathway

The catabolism of cholesterol is the best described steroidal catabolic pathway in bacteria due in large part, to its involvement in *Mtb* pathogenesis. Although many of the ~80 genes involved in cholesterol degradation and their operonic structure are conserved across actinobacteria [17], herein, the cholesterol catabolic gene cluster in *Mtb* will be discussed at length due to: (A) its ubiquity in the literature and (B) the subsequent focus of the research presented herein. Variations between cholesterol catabolism in other bacteria as well as in the catabolism of different steroids will be addressed in Sections 1.2.7 and 1.2.8.

The *Mtb* cholesterol catabolic genes are, for the most part, clustered in the genome between *rv3494c* and *rv3574* [25]. Cholesterol catabolic genes have been identified via a combination of transposon mutagenesis looking for insertions in genes that abrogate growth of *Mtb* on cholesterol [40] and transcriptomics studies looking at the up-regulation of genes during growth on cholesterol [45-48]. Table 1.1 summarizes the data for the cholesterol catabolic gene cluster as well as additional genes involved in cholesterol catabolism. Although the functions of many of these genes remain unknown, parallels between β-oxidative and aromatic degradation pathways have provided the frame work for the catabolic pathway for cholesterol.

Table 1.1. Cholesterol catabolic genes in *Mtb*

Locus	Synonym	Annotation	Essential for growth on cholesterol[40]	Up regulated on cholesterol[47]	Essential for virulence[A]
rv1106c	*3β-HSD*	3β hydroxysteroid dehydrogenase			
rv1130	*prpD*	methylcitrate dehydratase	✓		✓[49]
rv1131	*prpC*	methylcitrate synthase	✓		✓[49]
rv1143	*MCR*	Methyl Acyl-CoA racemase			
rv3409c	*choD*	cholesterol oxidase			✓[50]
rv3492c	-	MCE associated protein	✓		
rv3493c	-	MCE associated protein	✓		
rv3494c	*mce4F*	MCE family protein	✓		
rv3495c	*lprN*	MCE family protein	✓		
rv3496c	*mce4D*	MCE family protein	✓		
rv3497c	*mce4C*	MCE family protein	✓		
rv3498c	*mce4B*	MCE family protein	✓		
rv3499c	*mce4A*	MCE family protein	✓		
rv3500c	*yrbE4B*	MCE-associated protein	✓		
rv3501c	*yrbE4A*	MCE-associated protein	✓		✓[33]
rv3502c	*hsd4A*	17-hydroxysteroid dehydrogenase	✓	✓	
rv3503c	*fdxD*	Ferredoxin		✓	
rv3504	*chsE4*	acyl-CoA dehydrogenase		✓	
rv3505	*chsE5*	acyl-CoA dehydrogenase		✓	
rv3506	*fadD17*	acyl-CoA synthetase			
rv3515c	*fadD19*	acyl-CoA synthetase	✓	✓	
rv3516	*echA19*	enoyl-CoA hydratase		✓	
rv3518c	*cyp142*	Cytochrome P450		✓	
rv3520c		Reductase		✓	
rv3521		hypothetical protein		✓	
rv3522	*ltp4*	ketoacyl-CoA thiolase		✓	
rv3523	*ltp3*	ketoacyl-CoA thiolase		✓	
rv3526	*kshA*	ketosteroid monooxygenase	✓	✓	✓[51]
rv3534c	*hsaF*	4-hydroxy-2-oxovalerate aldolase	✓	✓	
rv3535c	*hsaG*	acetaldehyde dehyrogenase		✓	
rv3536c	*hsaE*	2-hydroxypentadienoate hydratase	✓	✓	
rv3537	*kstD*	ketosteroid dehydrogenase	✓	✓	✓[52]
rv3538	*hsd4B*	2-enoyl acyl-CoA hydratase		✓	
rv3540c	*ltp2*	ketoacyl-CoA thiolase	✓	✓	
rv3541c	*chsH1*	Mao-C like enoyl-CoA hydratase			
rv3542c	*chsH2*	Mao-C like enoyl-CoA hydratase	✓	✓	✓[52]
rv3543c	*chsE2*	acyl-CoA dehydrogenase	✓	✓	
rv3544c	*chsE1*	acyl-CoA dehydrogenase	✓	✓	✓[52]
rv3545c	*cyp125*	Cytochrome P450	✓	✓	

Locus	Synonym	Annotation	Essential for growth on cholesterol[40]	Up regulated on cholesterol[47]	Essential for virulence[A]
rv3546	*fadA5*	ketoacyl-CoA thiolase	✓	✓	✓[47]
rv3548c	-	short chain dehydrogenase	✓	✓	
rv3549c	-	short chain dehydrogenase	✓	✓	
rv3550	*echA20*	enoyl-CoA hydratase		✓	
rv3551	*ipdA*	CoA transferase subunit A	✓	✓	✓[52]
rv3552	*ipdB*	CoA transferase subunit B		✓	✓[52]
rv3553	*ipdC*	acyl-CoA reductase	✓	✓	
rv3556c	*fadA6*	ketoacyl-CoA thiolase		✓	✓[52]
rv3557c	*kstR2*	TetR type repressor		✓	
rv3559c	*ipdF*	ketoacyl-CoA reductase	✓	✓	
rv3560c	*fadE30*	acyl-CoA dehydrogenase	✓	✓	
rv3561	*fadD3*	HIP-CoA synthetase	✓	✓	
rv3562	*fadE31*	acyl-CoA dehydrogenase		✓	
rv3563	*fadE32*	acyl-CoA dehydrogenase	✓	✓	✓[52]
rv3564	*fadE33*	acyl-CoA dehydrogenase	✓	✓	
rv3565	*aspB*	aminotransferase		✓	
rv3567c	*hsaB*	monooxygenase subunit B		✓	
rv3568c	*hsaC*	DHSA dioxygenase	✓	✓	✓[39]
rv3569c	*hsaD*	DSHA hydrolase	✓	✓	✓[53]
rv3570c	*hsaA*	3-HSA monooxygenase	✓	✓	✓[52]
rv3571	*kshB*	Rieske Oxygenase reductase	✓	✓	✓[51]
rv3573c	*cshE3*	acyl-CoA dehydrogenase	✓	✓	
rv3574	*kstR*	TetR-type repressor		✓	

[A]Respective gene deletion or transposon mutant in *Mtb* in referenced study displayed reduced virulence in macrophage and/or animal studies

1.2.2 Cholesterol import

Cholesterol is transported across the bacterial cellular envelope by a set of proteins encoded by the *mce4* (mammalian cell entry) locus in actinobacteria [54]. In *Mtb* these genes correspond to the *rv3492c* to *rv3501c* locus [33]. The *mce4A-mce4F* genes are predicted to code for proteins that assemble into a hetero-hexameric ring with a helical, needle-like C-terminal domain that spans the periplasm, connecting the outer and inner membranes [55]. The MCE4 transporter is specific for steroids with long hydrophobic side chains [54, 56]. Although cholesterol import is predicted to be ATP dependent, disruption of the putative ATPase, MceG, or permease subunits of the ATP-binding cassette (ABC) transporter, YrbE4A, did not

completely prevent growth of *Mtb* on cholesterol [33]. Similarly, deletion of the *mce4* operon in *M. smegmatis* did not completely prevent cholesterol uptake, suggesting that other mechanisms of cholesterol uptake may exist [56]. In fact, the hydrophobic nature of cholesterol may permit passive diffusion through the inner membrane into the bacterial cell [33].

1.2.3 Side chain degradation

Degradation of the alkyl side chain of cholesterol starts with the oxidation of the terminal carbon at C-26 or C-27 to a carboxylic acid (Figure 1.4). In *Mtb*, two cytochromes P450 (Cyp) are able to catalyze this step [57, 58]. Cyp125 (Rv3545c) uses dioxygen to oxidize C-26 of cholesterol to generate 5-cholestene-26-oate [58]. The physiological reductase that transfers reducing equivalents from NADH to the Cyp appears to be KshB [58, 59]. Deletion of *cyp125* in *Mtb* CDC1551 leads to the accumulation of cholest-4-ene-3-one upon incubation with cholesterol, suggesting that this metabolite may be the physiological substrate of Cyp125 [60]. Interestingly, in *Mtb* H37Rv deletion of *cyp125* does not disrupt growth on cholesterol due to the presence of a functional Cyp142 which cansubstitutefor Cyp125 in the C-26 hydroxylation of cholesterol [57, 61]. Following formation of 5-cholestene-26-oate, the carboxylate is thioesterified using ATP and CoASH, to 5-cholestene-26-oyl-CoA by the acyl-CoA synthetase, FadD19 permitting the start of β-oxidation of the side chain [62].

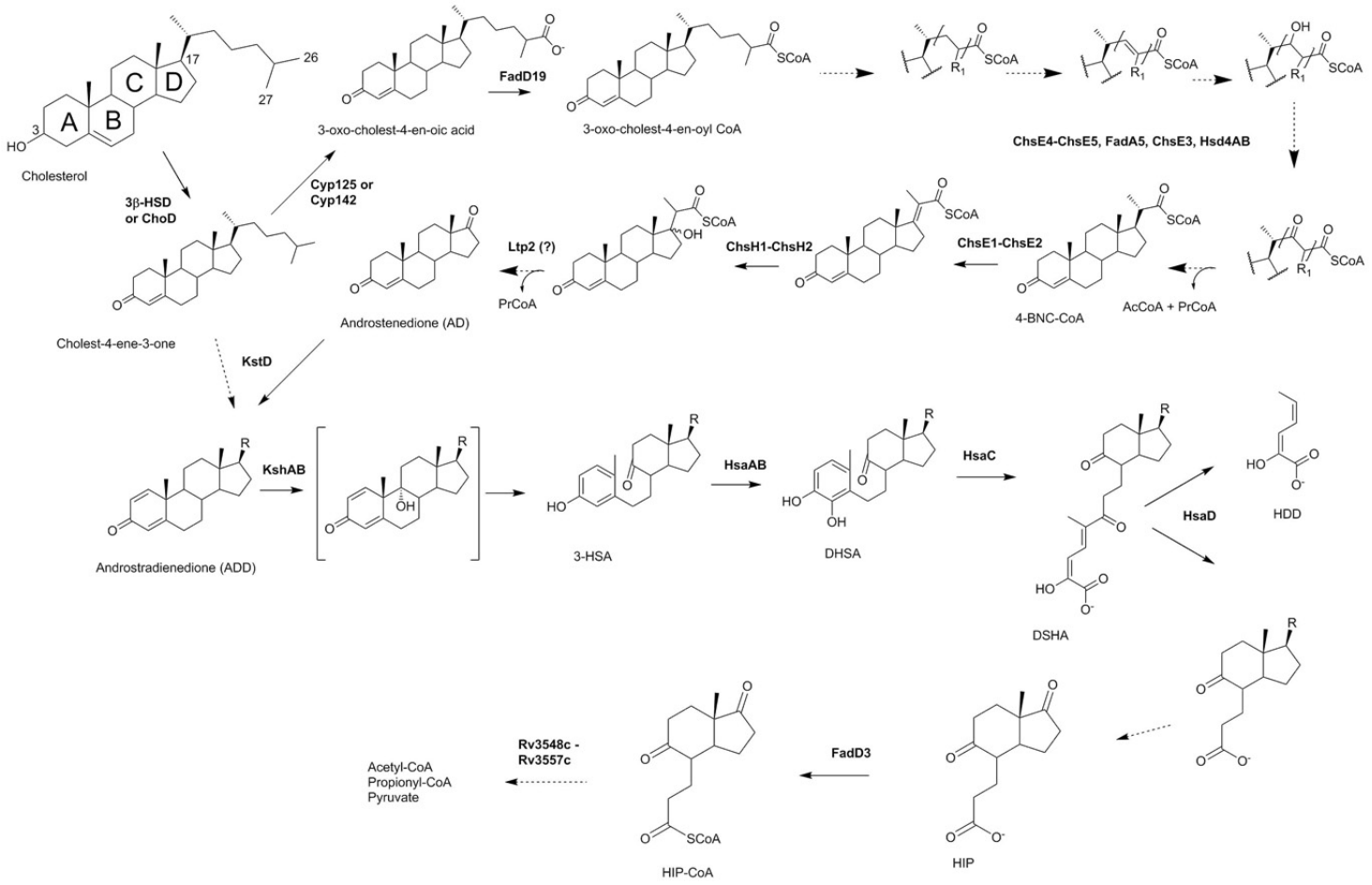


Figure 1.4 Cholesterol catabolic pathway in *Mtb.*
The cholesterol catabolic pathway as currently known for *Mtb*. Enzymes are shown in bold. R_1 indicates H or $–CH_3$ for an eight or five carbon side chain, respectively; R_2 resprents an unknown side chain or C-17 oxo. Reactions catalyzed by multiple enzymes are displayed as dashed lines. Proposed enzyme reactions are depicted in grey. Metabolite naming for ADD, 3-HSA, DHSA, and DSHA, all represent the C-17 oxo analogs of the depicted structures.

β-oxidation of the alkyl side chain involves four repeating steps: desaturation of C_α and C_β adjacent to a coenzyme A (CoA thioester) by an acyl-CoA dehydrogenase (FadE), hydroxylation of the double bond at C_β by an enoyl-CoA hydratase (EchA), oxidation of the C_β hydroxyl to a ketone by a 3-hydroxyl-acyl-CoA dehydrogenase (FadB), and thiolysis of the β-keto-acyl-CoA, releasing an acetyl-CoA or propionyl-CoA group by a β-keto-acyl-CoA thiolase (FadA). In *Mtb,* cholesterol metabolites with eight carbon CoA side chains, such as 3-oxo-chol-4-en-26-oyl-CoA are preferentially desaturated by the FadE complex of ChsE4-ChsE5 (Rv3504-Rv3505), whereas the five carbon side chain is desaturated by ChsE3 (Rv3573c) [63]. The *echA19* gene, located adjacent to *fadD19*, is up-regulated in the presence of cholesterol and likely encodes the EchA responsible for hydroxylation of eight and five carbon side chains due to the absence of another homolog [25, 47]. Although there are no FadB homologs encoded within the cholesterol catabolic gene cluster, Hsd4A (Rv3502c), annotated as a 17β-hydroxysteroid dehydrogenase, is essential for growth on cholesterol and predicted to dispense for a FadB type enzyme [40, 64]. FadA5 is responsible for the cleavage of the five carbon β-keto-acyl-CoA cholesterol metabolite and may act on the eight carbon side chain as well [65].

Two rounds of β-oxidation, yields the metabolite 3-oxo-23,24-bisnorchol-4-en-22-oyl-CoA (4-BNC-CoA) which contains an isopropionyl-CoA side chain at C-17. Genes involved in the removal of the isopropionyl-CoA side chain are clustered together between *rv3540c* and *rv3545c* in the *igr* operon (intracellular growth) which is essential for growth on cholesterol and bacterial persistence in macrophages [66]. β-Oxidation cannot be employed to remove the propionyl-CoA moiety due to the presence of a tertiary carbon at C-17. Removal of the propionyl-CoA moiety involves desaturation by ChsE1-ChsE2 (Rv3544c-Rv3543c) [63] then

hydroxylation at C-17 by a unique MaoC-like (*R*) – enoyl-CoA hydratase, ChsH1-ChsH2 (Rv3541c-Rv3542c) [67]. Retro-aldol C-17/C-22 bond cleavage has been predicted to be catalyzed by an aldolase from the lipid transfer protein (Ltp) superfamily of which 3 are encoded in the *Mtb* cholesterol catabolic gene cluster: Ltp2, Ltp3, and Ltp4 [25, 68]. Ltp3 and 4 are likely involved in the aldolytic cleavage of C-24 branches in steroids such as β-sitosterol and campesterol, but not cholesterol [68]. Although Ltp2 (Rv3540c) has been implicated in catalyzing this step as *ltp2* is adjacent to *chsH1*, alignments of Ltp2 with other aldolases suggest that Ltp2 may be lacking catalytically relevant residues suggesting that it may display a different role in cholesterol catabolism (Vanderven, 2017; unpublished).

Overall, complete side chain degradation yields two propionyl-CoA and one acetyl-CoA per molecule of cholesterol and generates androstenedione (AD), a metabolite with four steroidal rings and an oxo group at C-17. The previous paradigm stated that side chain degradation completes prior to the start of Rings A and B degradation. Therefore, side chain degradation enzymes were predicted to act on substrates with intact steroidal Rings A and B and conversely, substrates for Rings A and B degradation enzymes were not predicted to possess a CoA thioester. However, a growing body of evidence refutes this paradigm [41, 69], as discussed in Section 1.2.6.

1.2.4 Cholesterol Rings A and B degradation

Rings A and B degradation commences with the conversion of cholesterol to cholest-4-ene-3-one [50, 70]. Both 3β-hydroxysteroid dehydrogenase (3β-HSD, Rv1106c) and cholesterol oxidase (ChoD, Rv3409c) have been implicated in catalyzing this reaction based on the function of homologs [71]. Of the two, only 3β-HSD has been demonstrated to catalyze this reaction *in vitro* [50, 70]. However, both *choD* and *rv1106c* are dispensable for growth of *Mtb* on

cholesterol [72] and deletion of 3β-*hsd* in *Mtb* did not reduce growth in guinea pigs [70]. Overall, it is possible that the enzyme responsible for this reaction under physiological conditions has not yet been identified. One possibility is that *rv0139,* which encodes a putative hydroxysteroid dehydrogenase, may be involved in cholesterol catabolism as it is up regulated in the presence of cholesterol [32].

The catabolism of cholesterol Rings A and B, as well as enzyme and metabolite names are summarized in Figure 1.4. Androstenedione (AD) is converted to androsta-1,4-diene-3,17-dione (ADD) via the 3-ketosteroid Δ^1-dehydrogenase, KstD [73, 74]. ADD is then hydroxylated at C-9 by the 3-ketosteroid-9α-hydroxylase, KshAB, resulting in an unstable intermediate which readily aromatizes Ring A and undergoes cleavage of Ring B to form 3-hydroxy-9,10-secondrost-1,3,5(10)-triene-9,17-dione (3-HSA) [29, 59, 75]. 3-HSA is hydroxylated at C-4 by the flavin-dependent monooxygenase, HsaAB, generating a catechol, 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (DHSA) [19, 76]. The extradiol dioxygenase, HsaC, uses dioxygen to catalyze the *meta*-cleavage of DHSA resulting in a *meta*-cleavage product (MCP), 4,5-9,10-diseco- α3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid (DSHA) [39]. Finally, the MCP hydrolase, HsaD, cleaves the C-5/C-10 bond of DSHA (using steroid carbon numbering), to generate 2-hydroxy-hexa-2,4-dienoic acid (HHD) and 3aα-H-4α(3′-propanoyl-CoA)-7aβ-methylhexa-hydro-1,5-indanedione (HIP) [77]. HHD is degraded by HsaEFG [25, 78] and HIP feeds into Rings C and D degradation [79].

Rings A and B degradation results in one pyruvate and propionyl-CoA per molecule of cholesterol from the catabolism of HHD. The oxidative manner by which cholesterol Rings A and B are degraded is analogous to the degradation of aromatic compounds such as biphenyl [80]. Interestingly, the identification of the cholesterol catabolic gene cluster in *Mtb* was due in

large part to the observation that *Mtb* encoded homologs to the extradiol dioxygenase, BphC, and MCP hydrolase, BphD, involved in biphenyl degradation [81, 82]. Homologs of KshA, HsaA, HsaC, and HsaD occur in all known aerobic steroid catabolic pathways [17].

1.2.5 Cholesterol Rings C and D degradation

In contrast to side chain and Ring A and B degradation, the catabolism of cholesterol Rings C and D (HIP) is poorly understood. In the 1960s, Lee and Sih proposed that Rings C and D are degraded either oxidatively, via a Bayer-Villager Monooxygenase (BVMO) and subsequent hydrolysis of the rings, or in a β-oxidative-like manner involving retro- Claisen ring opening reactions [28]. Failure to identify a candidate BVMO in the *Mtb* genome has implicated the latter pathway proposed by Lee and Sih. Interestingly, *Mtb* and RHA1 are able to grow on HIP as a sole carbon source suggesting that their respective genomes encode for HIP catabolic genes [79].

The acyl-CoA synthetase, FadD3, was the first characterized HIP catabolic enzyme. FadD3 catalyzes the ATP- and CoASH-dependent thioesterification of HIP to HIP-CoA [79]. Further, deletion of *fadD3* in RHA1 caused the extracellular accumulation of HIP in the presence of cholesterol [79]. Based on the formation of HIP-CoA, Casabon *et al*. proposed that the catabolism of HIP would be β-oxidative in nature involving CoA thioester metabolites [79]. Although FadD3 is the only HIP catabolic enzyme characterized to date, deletion studies in other bacteria have implicated additional genes. *R. equi* deficient in *fadE30* or *ipdAB* accumulated extracellular 5OH-HIP or HIP, respectively, upon incubation with cholesterol [23]. Therefore, van der Geize *et al.* proposed that IpdAB, a predicted CoA transferase, would be involved in HIP-CoA synthesis and FadE30 would desaturate HIP-CoA [23]. Interestingly, *fadD3*, *fadE30*, and *ipdAB* are all part of the KstR2 regulon, a differentially regulated set of genes in the

cholesterol catabolic gene cluster [83], as discussed in Section 1.2.7. Lastly, in 1977, Hashimoto and Harakawa observed the accumulation of 5-methyl-4-oxo-octane-1,8-dioate (MOODA) when HIP was incubated with *Streptomycetes rubescens* [84].

1.2.6 Concurrent side chain and Rings A and B degradation

In 2011, Capyk *et al.* reported that KshAB, responsible for the aromatization of Ring A, had a higher substrate specificity (k_{cat}/K_M) for steroidal substrates that possessed a CoA thioesterified side chain [69]. This suggested that side chain degradation and Rings A/B degradation occur concurrently to some extent. The extent of this concurrent catabolism is unclear. However, it is noted that the substrate specificities of KshAB, HsaAB, and HsaD for substrates with completely degraded side chains are 2-3 $\log_{10}$ orders lower than reported for homologous enzymes and their physiological substrates [59, 76, 77]. The implication is that perhaps these enzymes preferentially act on substrates with partially degraded side chains. Interestingly, FadD3 does not ligate CoASH onto 1β(2′-propanoate)-3aα-H-4aα(3”-propanoate)-7aβ-methylhexahydro-5-indanone (HIP with an isopropionyl side chain at C-17) implying that side chain degradation completes prior to the degradation of cholesterol Rings C and D [79].

1.2.7 Organization and regulation of steroid catabolic gene clusters

The cholesterol catabolic gene cluster in *Mtb* is primarily located between *rv3409c* and *rv3574* [25]. Figure 1.5 depicts the organization of the cholesterol catabolic gene cluster in *Mtb*. In general, the cholesterol catabolic gene cluster is separated into regions encoding for enzymes that catalyze cholesterol import (*rv3492 – rv3501c*) or side chain (*rv3504 – rv3546*), Rings A and B (*rv3567c – rv3574*), and Rings C and D (*rv3548c – rv3564*) degradation [25]. Similarly, many of these aforementioned regions are organized into operons which are reflected in their co-regulation on cholesterol [47].

The cholesterol catabolic gene cluster is regulated by at least two TetR family transcriptional repressors, KstR (Rv3574) and KstR2 (Rv3557c) [32, 83], described in Section 1.6. KstR and KstR2 bind DNA at conserved 14 bp palindromic sequences located at the -10 to -35 bp regions of their regulated genes [32].

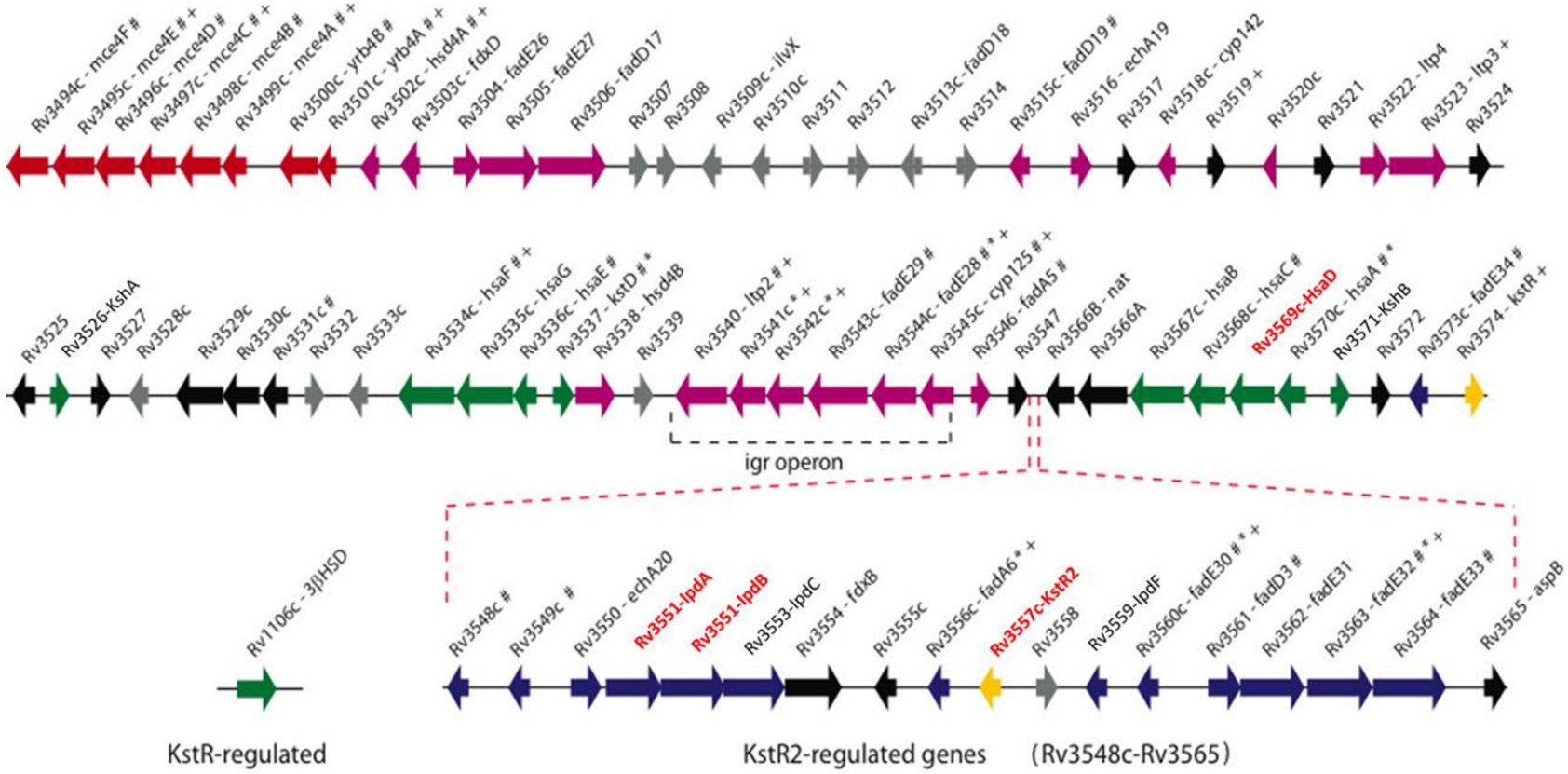


Figure 1.5. Organization of the cholesterol catabolic gene cluster in *Mtb*.
Arrows indicate genes involved in cholesterol import (red), side-chain degradation (purple), Rings A and B degradation (green), Rings C and D degradation (blue), unassigned function (black), regulation (yellow), and genes not involved in cholesterol catabolism (grey). Genes described in length herein are identified in red text. Figure adapted from Capyk, J. (2012) [85].

First characterized by Kendall *et al.* in 2007 in *M. smegmatis [32]*, the KstR regulon is predicted to comprise at least 74 genes in *Mtb* based on the occurrence of 16 KstR binding motifs in the genome, of which 13 occur within the cholesterol catabolic gene cluster [32]. Deletion of *kstR* in *M. smegmatis* caused the up-regulation of many of the cholesterol catabolic genes however yielded only a slight growth defect [32]. The inducer of KstR was identified to be 3-oxo-cholest-4-en-26-oyl-CoA [86], a cholesterol metabolite generated in side chain degradation (Section 1.2.3). Binding of 3-oxo-cholest-4-en-26-oyl-CoA to KstR results in a conformational change that causes the release of KstR from the DNA binding motif to permit transcription [86]. The first CoA thioester formed in a catabolic pathway is often the effector of that pathway. For example, the effectors of PaaR and FadR from *Thermus thermophilus* HB8, involved in the catabolism of phenylacetate (PAA) and fatty acids, are PAA-CoA and lauroyl-CoA, respectively [87-89], while the effector of CouR of RHA1, a MarR-type transcriptional repressor, is *p*-coumaroyl-CoA [90]

The KstR2 regulon is much smaller, comprising only ~14 genes. In Mtb, these genes are clustered between *rv3548c* and *rv3564* in the cholesterol catabolic gene cluster [83]. These genes are predicted to be involved in the catabolism of cholesterol Rings C and D [79]. In 2013, Casabon *et al.* identified HIP-CoA, the product of the FadD3 reaction, as the effector of KstR2 from *Mtb* [91]. In RHA1 grown on HIP, genes of the KstR2 regulon were up-regulated, but not those of the KstR regulon [91]. Intriguingly, the KstR2 regulon is highly conserved across steroid degrading actinobacteria and proteobacteria [17, 20, 22]. However, actinobacteria with multiple steroid catabolic clusters only have a single KstR2 regulon, suggesting that the catabolism of all steroids feeds into a shared HIP catabolic pathway [22]. Although genes of the

KstR2 regulon are typically conserved between actinobacteria and proteobacteria, they appear to be regulated by a LuxR family transcriptional repressor in proteobacteria [92].

1.2.8 Cholate and testosterone catabolism

Cholate and testosterone catabolism is similar to that described for cholesterol, however with a few differences. Testosterone, and more generally androgens, lack a C-17 alkyl side chain and therefore, their catabolic gene clusters only encode for enzymes responsible for the degradation of steroidal rings [18]. Cholate possesses a five carbon side chain at C-17 with a terminal carboxylic acid. Therefore, a cytochrome P450 is not required for activation of the side chain prior to catabolism [18]. As with cholesterol, cholate is thioesterified by an acyl-CoA synthetase, CasG in RHA1, enabling the β-oxidation of the alkyl side chain [62]. Interestingly, catabolism of the cholate alkyl side chain does not involve a FadA5-like β-keto-acyl-CoA thiolase in *Pseudomonas putida* Chol1, put proceeds through a unique aldehyde intermediate presumably generated by the aldolase, Skt [93].

1.3 *Mycobacterium tuberculosis* and tuberculosis

Mtb, the causative agent of tuberculosis (TB), is responsible for ~1.8 million deaths annually [94]. It is proposed that up to one third of the world's population are either infected with *Mtb* and/or will test positive for *Mtb* exposure, however less than 10% of these individuals will develop TB [94]. Over 60% of the world's TB cases occur within India, China, South Africa and Russia, with Sub-Saharan Africa having the highest per capita rates of TB [94].

Mtb is an intracellular human pathogen transmitted between hosts via aerosolized droplets resulting from coughing by infected individuals. Once inhaled into the lungs, *Mtb* is phagocytosed by alveolar macrophages. Using mechanisms yet to be fully understood, *Mtb* is able to prevent fusion of the phagosome with the lysosome within macrophages, thereby

preventing acidification and maintaining an environment permissive for bacterial replication [95]. *Mtb* is then sequestered by phagocytes into granulomas or tubercles which can occlude the bacteria from the rest of the lung for decades [95]. Active TB results from the breakdown of these granulomas releasing *Mtb* into the lungs [95]. The exact mechanisms that lead to the development of TB remain unclear, but the immune status of the infected individual plays an important role as the disease is common in the immunocompromised as evidenced by the fact that nearly a quarter of the deaths caused by TB in 2016 were HIV-associated [94].

Streptomycin, the first antibiotic used to treat TB, was first isolated in 1943 by Selman Waksman and his graduate student Albert Schatz [96]. Although initially effective, streptomycin resistant strains of *Mtb* were observed shortly after its introduction due to its administration as a monotherapy [97]. Thereafter, anti-tuberculosis therapeutics have been prescribed as combination therapies in an attempt to limit the spread of multi drug resistance (MDR). Currently, standard *Mtb* treatment employs four front line anti-tuberculosis therapeutics, rifampicin (RIF), isoniazid (INH), ethambutol (EMB), and pyrazinamide (PZA), in which all four are administered daily for two months followed by four months of INH and RIF only [97, 98]. Largely due to the negative side effects of the current drugs and distribution issues within developing nations, MDR-TB has emerged in response to low compliance in completing the standard treatment. Therefore, second line treatments such as the fluoroquinolones and aminoglycosides are increasingly prescribed for the treatment of TB, which has in turn led to widening the spectrum of antibiotic resistance within *Mtb* [98]. Extensively drug resistant strains, referred to as XDR-TB, are resistant to both isoniazid and rifampicin as well as to at least one of the fluoroquinolones and injectable second line drugs. XDR-TB are estimated to make up almost 10% of MDR-TB cases world-wide [94]. In March 2012, a totally drug resistant *Mtb* (TDR-TB)

was reportedly isolated in India that displayed resistance to all first and second line drugs [99]. Although questions have arisen regarding the validity of this claim by the World Health Organization (WHO), this exemplifies the urgent need for novel therapeutics to obvert further antibiotic resistance in *Mtb* [97].

1.3.1 Mycolata

Mycobacteria and rhodococci belong to a bacterial taxon known as mycolata. Mycolata are characterized by an outer membrane that contains mycolic acids. Mycolic acids are 2-alkyl, 3-hydroxy long-chain fatty acids consisting of 20 – 100 carbon atoms with various genera and/or species specific desaturations and/or hydroxylations [100]. Mycolata cell walls consist of a thick mycomembrane, composed of a parallel arrangement of mycolic acids linked to arabinogalactan, which is covalently attached to a peptidoglycan layer [100]. The highly hydrophobic nature of the mycomembrane creates a barrier to hydrophilic molecules, giving Mycolata a Gram-negative like characteristic and limiting the effectiveness of many antibiotics [101].

The majority of Mycolata, are non-pathogenic saprophytes often regarded as important organisms due to their vast catabolic capabilities which are employed in bioremediation and biocatalytic applications [80, 102]. For example, the steroid degrading capabilities of Mycolata are being increasingly employed as biocatalysts to produce 17-ketosterols, AD, 9α-OH-AD, and ADD [26, 103-105]. Steroid degradation is ubiquitous across Mycolata [17]. In a recent analysis of 8,200 bacterial genomes, 212 putative steroid-degrading *Actinobacteria* spp. (including the Mycolata) from 16 genera were identified [17]. Of all mycobacterial and rhodococcal genomes characterized, only *M. leprae* and *Rhodococcus* sp. strain AW2509 did not encode a steroid catabolic pathway [17].

1.3.1.1 *Mycobacterium abscessus*

Mycobacterium abscessus (*Mab*) is a rapidly growing mycobacterium that is involved in nontuberculous pulmonary infections typically within immunocompromised or Cystic Fibrosis (CF) patients [106, 107]. Increased concern has emerged as the global incidence of *M. abscessus* infections has risen [107]. Due to its intrinsic multi-drug resistance [108], the recent identification of person-to-person transmission of *M. abscessus* further increases the need for novel therapeutic options [107]. *Mab* was recently identified as containing a steroid catabolic gene cluster within its genome [17], although the role of steroid catabolism in infection as well as the characterization of said pathway remains unstudied.

1.3.1.2 *Mycobacterium smegmatis*

Mycobacterium smegmatis MC^2155 is a largely non-pathogenic mycobacterium that is often used as a model organism in studying *Mtb* physiology. Compared to *Mtb, M. smegmatis* grows rapidly and is more genetically tractable permitting wider laboratory use [109]. *M. smegmatis* grows on cholesterol and contains a similarly organized cholesterol catabolic gene cluster to that of *Mtb* [17, 26]. A second steroid catabolic gene cluster was recently identified in *M. smegmatis* that was proposed to be involved in the degradation of androgens as it is not up regulated in the presence of cholesterol, however additional studies are required [27].

1.3.1.3 *Rhodococcus jostii* RHA1

Rhodococcus jostii RHA1 is a non-pathogenic, fast-growing bacterium that is genetically tractable. First isolated from lindane-contaminated soils, RHA1 has garnered much interest due to its ability to degrade a wide range of aromatic compounds [110]. The RHA1 genome is over 9.7 Mbp and is one of the largest bacterial genomes sequenced to date [111]. RHA1 contains three complete steroid catabolic pathways, two of which are responsible for the degradation of

cholate and cholesterol, respectively [22]. RHA1 has been used as a model organism to investigate cholesterol catabolism [39, 76, 79]. Unlike *Mtb*, the cholesterol catabolic gene cluster is split into three regions in RHA1 separated by two ~70 kbp segments [111].

1.3.2 Lipid metabolism in *Mtb*

Lipid metabolism in *Mtb* has emerged as a key area of research due to its crucial role in virulence. The current paradigm states that once inside the phagosome, *Mtb* is restricted to a limited number of host-derived nutrients of which lipids and cholesterol make up the majority [112]. Therefore, lipids are a primary source of carbon for *Mtb* during infection. This is reflected in the observation by Cole *et al.* that the *Mtb* genome is "overrepresented" in the number of genes predicted to be involved in the catabolism of fatty acids [113]. In fact, disruption of multiple genes involved in lipid metabolism within *Mtb* are predicted to result in either attenuation or avirulence in a mouse model [114, 115].

The utilization of lipids as a primary nutritional source requires specific metabolic adaptations by *Mtb* [49]. Catabolism of odd chain fatty acids yield propionyl-CoA which is toxic at high intracellular concentrations [49, 112, 116]. Propionyl-CoA is metabolized by the methyl citrate cycle (MCC), a central metabolic pathway in *Mtb* that involves the condensation of propionyl-CoA with oxaloacetate by methylcitrate synthase, PrpC, to make methylcitrate which is fed into TCA by way of isocitrate lyase (Icl) (Figure 1.6) [49]. Disruption of *prpC* or *icl* as well as other genes of the MCC prevent growth of *Mtb* in macrophages [49, 52, 117, 118]. However, a *prpDC* mutant had no effect on tissue pathology [119]. Similarly, propionyl-CoA is anabolized to virulence-associated polyketide lipids phiocerol-dimycoseroic acid (PDIM), poly-acylated trehaloses (PATS), sulfolipid (SL), and mycolic acids [120]. PDIM, for example, is important in host-pathogen interactions [112]. Lastly, availability of vitamin B12 permits

propionyl-CoA metabolism via the methylmalonyl pathway, which can compensate for disruption of the MCC [49, 116].

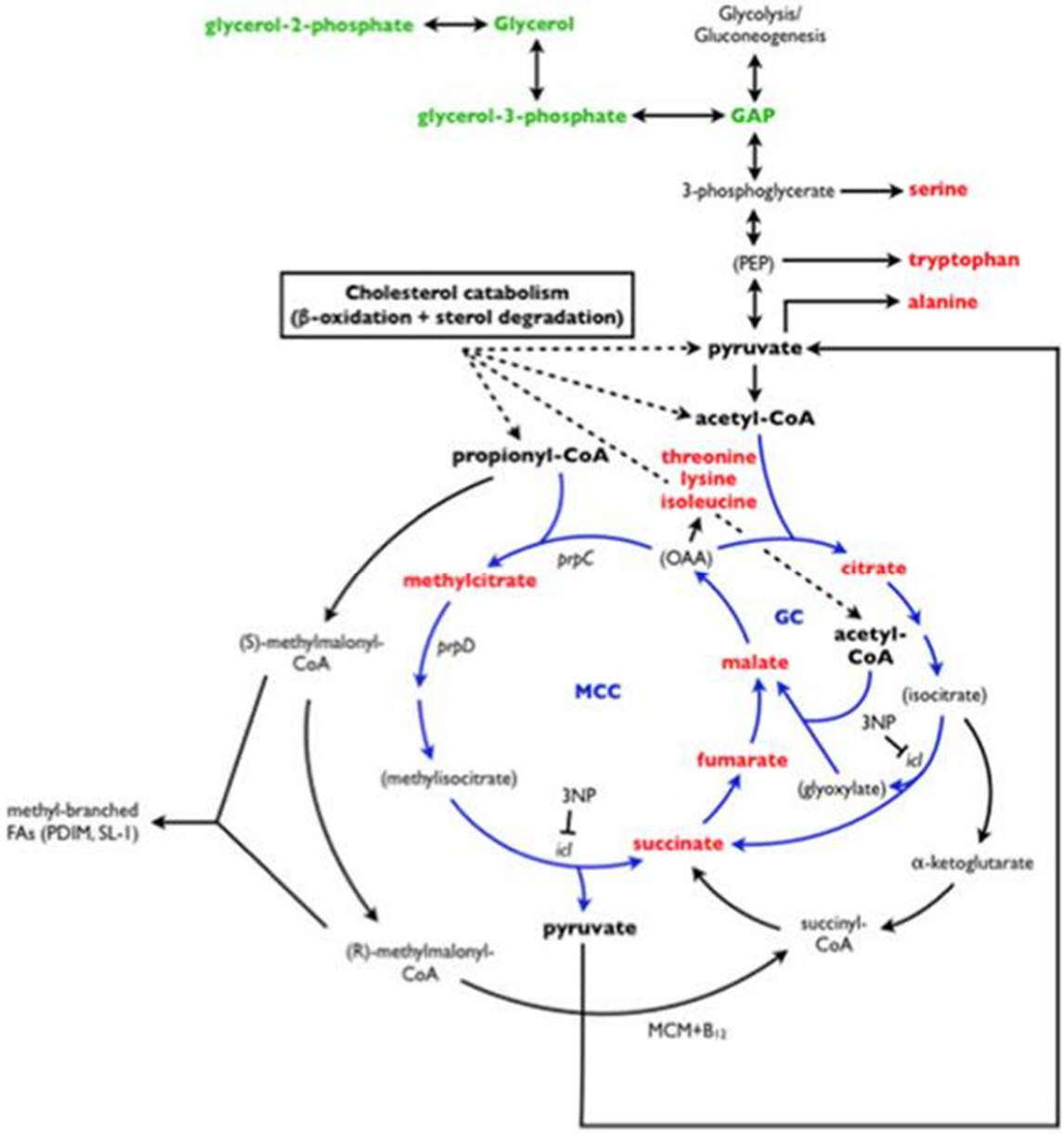


Figure 1.6. Central metabolic pathways of *Mtb* involved in fatty acid and cholesterol catabolism.
Diagram displaying how acetyl-CoA, propionyl-CoA, and pyruvate generated from lipid and cholesterol catabolism in *Mtb* feeds into the methylcitrate cycle (MCC), glyoxylate cycle (GC), and methylmalonyl-CoA pathway. Other acronyms are: FA (fatty acids), MCM (methylmalonyl-CoA mutase), PEP (phosphoenolpyruvate), and 3NP (3-nitropropionate). Figure adapted from Griffin *et al.* (2012) and used with permission from *Elsevier.*

1.3.3 Role of cholesterol catabolism in *Mtb* pathogenesis

Shortly after the identification of the cholesterol catabolic gene cluster in 2007, it was noted that deletion of many of the genes involved in cholesterol catabolism had been previously identified as essential for growth in macrophages [25, 52]. *Mtb* strains deficient in cholesterol import due to disruption of genes within the *mce4* locus demonstrated severe growth attenuation within mouse lungs [33]. Deletion of the *igr* locus, involved in the removal of the final propionyl-CoA from the cholesterol side chain, attenuates *Mtb* growth in macrophages and mouse models [66]. *Mtb* deficient in *hsaC* is unable to catabolize cholesterol, displays reduced virulence in mouse models, and causes significantly fewer pulmonary lesions within guinea pigs [39]. This severe attenuation of *ΔhsaC Mtb* is proposed to result from the accumulation of toxic catechols resulting from the absence of HsaC [39]. Similarly, investigations into *Mtb* deficient in *hsaD, fadA5, kshA, kshB* or *choD* demonstrate growth attenuation in macrophages and/or virulence in mouse models [47, 50, 51, 53]. *Mtb* deficient in *cyp125* display increased sensitivity to azoles, however are not attenuated in macrophages [121]. Furthermore, deletion of *ipdAB* in the related horse pathogen, *R. equi*, is patented as a live vaccine [23]. Overall, deletion of 17 of the ~80 genes involved in cholesterol catabolism have demonstrated some attenuation in infection models, strongly implicating a role of cholesterol catabolism in *Mtb* virulence (Table 1.1). Unexpectedly, many of the aforementioned deletion strains in *Mtb* are unable to grow in the presence of cholesterol and a second carbon source suggesting the accumulation of toxic cholesterol metabolites.

Cholesterol is readily available to *Mtb* during infection. Granulomas are rich in cholesterol and cholesterol esters [122]. During infection, *Mtb* extracts host-derived cholesterol from the phagosomal membrane [123]. Interestingly, individuals that display lower blood

cholesterol levels are less likely to develop TB [124, 125]. Additionally, patients receiving statins to lower blood cholesterol levels display a lower risk for TB [126, 127].

Although *Mtb* grows on cholesterol as a sole carbon source *in vitro* and disruption of cholesterol catabolism impairs pathogenicity, the exact function of cholesterol catabolism in virulence is poorly understood. Unexpectedly, growth of *Mtb* on ^{14}[C]-4 or ^{14}[C]-26 – cholesterol yielded radioactivity in primarily CO_2 and membrane lipids, respectively [33, 49]. This suggests that production of virulence lipids, such as PDIM, may depend on the cholesterol side chain.

In addition to its use as an intracellular carbon source, there are two other primary hypotheses concerning the role of cholesterol catabolism in *Mtb.* Firstly, it is possible that cholesterol is required for *Mtb* uptake by macrophages. Chemical depletion of cholesterol from lipid rafts in macrophages prevent *Mtb* uptake [128]. Additionally, cholesterol in the phagosomal membrane is essential in preventing the phagosome-lysosome fusion in mice [129]. Secondly, cholesterol catabolites may play a role in modulating host-pathogen interactions. For example, catabolism of the alkyl side chain of cholesterol by *Mtb* involves the formation of various hydroxylated cholesterol intermediates, some of which are structural analogs to immune-regulating steroids [59, 130]. Furthermore, degradation of cholesterol Rings A and B by *Mtb* generates structural analogs of cholecalciferol, or Vitamin-D_3. Vitamin-D_3 has broad immunological regulatory roles during *Mtb* infection and modulates host lipid metabolism [131]. Although it is unknown whether *Mtb* is able to excrete these catabolites to modulate host response, precedence exists for such activity within mycobacteria.

Mycobacterium leprae is an intracellular pathogen that is unable to utilize cholesterol as a nutrient source [132]. However, *M. leprae* converts host derived cholesterol to cholestenone during infection, which is excreted into the macrophage [132]. Marques *et al.* propose that

cholestenone, which is the preferred substrate for human Cyp27A1 [133], is readily converted to 27-hydroxycholestenone in the macrophages, which modulates immune response [132].

1.3.3.1 Cholesterol catabolism as a target for novel therapeutics

Due to the central role of cholesterol catabolism in the virulence of *Mtb*, and the potential for toxic cholesterol metabolites, numerous studies have attempted to identify small molecules that inhibit the cholesterol catabolic pathway directly [53, 134-136]. In their preeminent study, VanderVen *et al*. screened a chemical library for compounds that specifically inhibit *Mtb* growth in macrophages identifying a disproportionally high number of hits that target cholesterol catabolism [134]. A total of 41 compounds inhibited intracellular growth of *Mtb* and prevented *in vitro* growth of *Mtb* only in media containing cholesterol [134]. Of the characterized compounds, one targeted PrpC, involved in the MCC, one acted on the adenylate cyclase, Rv1625c, whose role in cholesterol catabolism is unclear, and two inhibited the HsaAB reaction with IC_{50} values of <10 μM [134]. HsaD has also seen much attention as a target for anti-tuberculosis compounds due to its essentiality in virulence, absence of homologs in humans, and ease of use in high throughput screening [53, 77, 135]. Recently, 7 compounds were identified that both inhibited HsaD *in vitro* and prevent growth of *Mtb* in media containing cholesterol [53]. These inhibitors were shown to bind within the HsaD active site in x-ray crystal structures [53]. Although reported IC_{50} values were relatively high (>100 μM), this study indicates that screening for inhibitors of specific cholesterol catabolic enzymes can be effective in identifying compounds with anti-tuberculosis activity *in vivo* [53].

1.4 Coenzyme A (CoA)

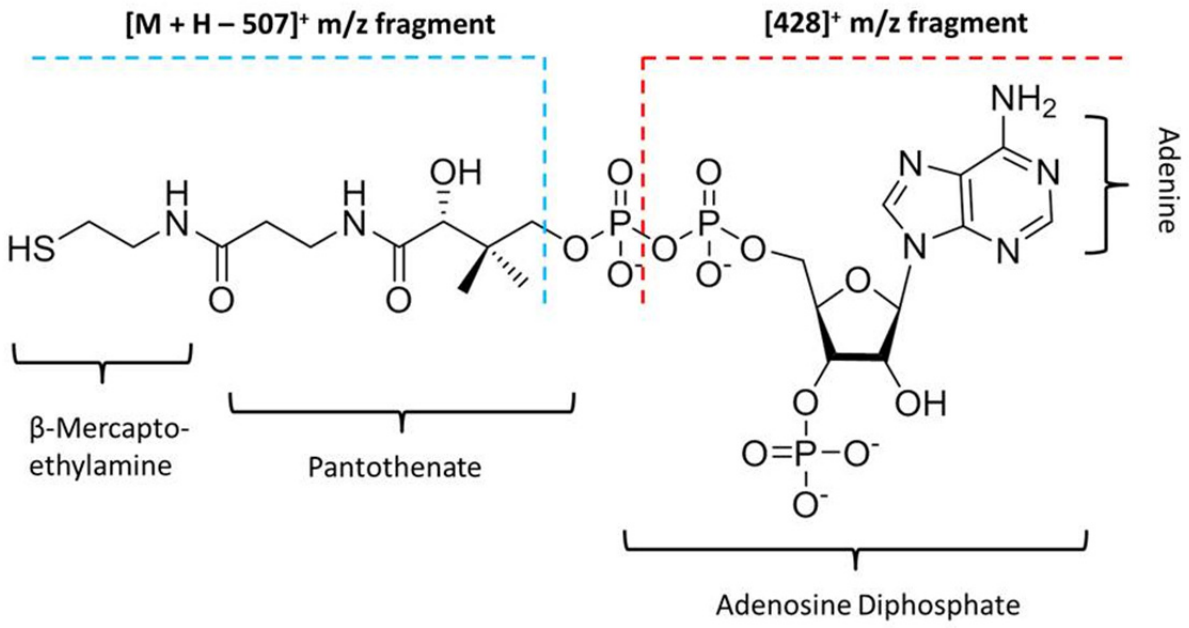


Figure 1.7 Structure of Coenzyme-A.
Specific moieties are identified on the bottom and right side. Blue and red dotted lines indicate fragments that generate the indicated molecular weights upon in source fragmentation during LC-MS/MS.

During the initial characterization of the cholesterol catabolic pathway, the role of coenzyme A (CoA) thioesters was thought to be limited to the catabolism of the alkyl side chain via β-oxidation. However subsequent studies involving KshAB, FadD3, KstR, and KstR2 have suggested that CoA thioesterified steroidal substrates may play a larger role in cholesterol catabolism than initially thought [69, 79, 86]. The body of work presented herein provides a much expanded role for CoA thioesters in cholesterol catabolism. In brief, CoA thioesters are shown to (A) determine substrate specificities of cholesterol catabolic enzymes, (B) regulate cholesterol catabolism, and (C) make up the majority of metabolites in the catabolism of Rings C and D.

CoA, or CoASH in its non-thioesterified form, is an essential cofactor involved in a number of cellular processes in all known organisms [137]. It is estimated that up to 4% of all

known enzymes require CoA as a cofactor [137]. CoA consists of three moieties: β-mercaptoethylamine, pantothenate, and adenosine diphosphate (ADP) (Figure 1.7). CoA is synthesized in five steps in *Mtb*: 1) phosphorylation of pantothenate (Vitamin B_5) to 4'-phosophopantothenate by pantothenate kinase (PanK or CoaA); 2) condensation with a cysteine residue by phosphopantothenoylcysteine synthetase (CoaB); 3) decarboxylation by phosphopantethenoylcysteine decarboxylase (CoaC); 4) adenylation by phosphopantetheine adenylyltransferase (CoaD) to form dephosopho-CoA; and 5) phosphorylation by dephosphocoenzyme A kinase (CoaE) to CoA [138]. CoaB and CoaC are fused as a bifunctional enzyme in *Mtb* [139]. CoaA, CoaBC, and CoaE are essential for *Mtb* survival [140, 141].

CoA has two primary functions: acting as an acyl group carrier and activating carbonyl groups. CoA thioesterification sequesters metabolites within the cell due to their inability to cross the hydrophobic cell membranes. Similarly, CoA increases the solubility of hydrophobic molecules. The CoA moiety is used to target specific acyl groups to their cognate enzymes in numerous anabolic and catabolic processes [142]. CoA functionality is dependent upon its free thiol group. Deprotonation of the thiol generates a strong nucleophile that is able to attack the carbonyl carbon of carboxylates or ketones resulting in a CoA thioester. Due to their relatively high standard free energies of hydrolysis, CoA thioesters have a high acyl group transfer potential [143]. Thus, the acyl group attached to CoA is described as 'activated'. For example, the synthesis of citrate from acetyl-CoA and oxaloacetate by citrate synthase is thermodynamically driven by the hydrolysis of acetyl-CoA [142]. Similarly, the formation of a CoA thioester results in increased acidity of adjacent protons, as compared to esters or carboxylic acids, permitting aldol or Claisen condensation reactions [144]. Finally, some enzymes are known to utilize the binding energy from noncovalent interactions with CoA to

decrease activation energy of energetically unfavourable reactions, thereby increasing the rate of catalysis [145-147].

Due to their ubiquity in metabolic processes, methods to identify and quantitate CoA thioesters have become paramount in lipidomic and metabolic research. High performance liquid chromatography (HPLC) of CoA thioesters has been broadly employed due to their characteristic absorbance at 258 nm (ε_{258} = 11.9 mM^{-1} cm^{-1}) from the adenine moiety [79, 148]. Chromatography is obtained in reverse phase typically with C18 or pentafluoridated phenyl (PFP) resins using an ammonium acetate or ammonium formate counter ion on the mobile phase due to the hydrophilicity of the CoA moiety [69, 79, 149, 150]. Mass spectrometry (MS) of CoA thioesters is typically performed using electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) due to low volatility and instability of the CoA moiety [149]. Recently, multiple reaction monitoring (MRM) has become prevalent in LC-MS analysis of CoA thioesters due to their characteristic fragmentation [149, 151]. Voltage induced in-source fragmentation of CoA thioesters results in a $[428]^+$ m/z fragment as well as a loss of 507 fragment $[M+H-507]^+$ *m/z* fragment (Figure 1.7) [151]. Therefore, highly sensitive and precise targeted analysis of CoA thioesters is obtained by monitoring the $[M+H]^+ \rightarrow [428]^+$ and $[M+H]^+ \rightarrow [M+H-507]^+$ transitions. Using MRM, CoA thioesters have been reportedly detected in the sub nanomolar range from biological samples, although sensitivity is highly sample and instrument dependent [149, 151].

1.5 *Meta*-cleavage product (MCP) hydrolases

HsaD, the cholesterol catabolic enzyme responsible for the hydrolysis of DSHA to HIP and HHD, is a *meta*-cleavage product (MCP) hydrolase. MCP hydrolases hydrolyze carbon-carbon bonds in vinylogous-1,5-diketones [152] and, with the exception of LigY [153], belong to

the α/β hydrolase superfamily. Genes encoding MCP hydrolases are prevalent in aromatic degrading bacteria, and tend to be organized into an operon with extradiol dioxygenase genes [154, 155]. MCP hydrolases have been separated into groups on the basis of their substrate specificities. Type I, II, and III MCP hydrolases act on the cleavage products of bicyclic, monocyclic, and heteroaromatics, respectively. To date, the best characterized MCP hydrolases are BphD from *Burkholderia xenovorans* LB400 and MhpC from *Escherichia coli*, which transform 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate (HOPDA) and 2-hydroxy-6-oxo-nona-2,4-diene- 1,9-dioate, respectively [81, 156-159].

1.5.1 Structure of MCP hydrolases

MCP hydrolases contain a conserved Ser-His-Asp catalytic triad [160] located adjacent to an oxyanion hole, characteristic of other hydrolytic enzymes [156, 161]**.** The α/β hydrolase fold consists of seven parallel and one antiparallel β-strand, twisted into a β-sheet surrounded by six alpha helices. An α-helical 'lid' domain is inserted between β6 and α6 of the core domain and contributes to substrate specificity [162]. The active site is an elongated tunnel located within the core domain and underneath the lid domain containing two subsites: a polar one binds the dienoate moiety of the MCP and a hydrophobic one binds the remainder of the MCP [77, 81, 162]. In HsaD, the entrance to the active site pocket is enlarged compared to $BphD_{LB400}$ and MhpC, with a single polar region located within 10 Å of the active site whose function is currently unknown [77].

1.5.2 Mechanism of serine-dependent MCP hydrolases

Early data had been interpreted to favour an acid/base mechanism of C-C bond cleavage in Ser-dependent MCP hydrolases [157]. However, recent evidence indicates that these enzymes employ covalent catalysis [156, 159]. Thus, when $BphD_{LB400}$ was rapidly quenched in the

presence of HOPDA, a benzoyl-Ser^{112} acyl enzyme intermediate was observed by ESI-MS [158]. The acylated species was also observed in the H265Q variant of BphD *in cristallo* [158]. Deacylation of the benzyl-Ser^{112} intermediate incorporated a single ^{18}O equivalent (from $H_2{}^{18}O$) into the product, benzoate, and proceeded at a rate consistent with HOPDA turnover [158]. This observation contrasted work done on MhpC in which two ^{18}O were observed in reaction products, supporting the reversible formation of a *gem*-diol intermediate [157]. Interestingly, the catalytic His^{265} in $BphD_{LB400}$ is not required for acylation of Ser^{112} but is essential for deacylation, further supporting covalent catalysis [158]. Consequently, deprotonation of the catalytic serine (Ser^{112} in $BphD_{LB400}$ or Ser^{114} in $HsaD_{Mtb}$) appears to be substrate-assisted and facilitated by localization of an electron lone pair adjacent to the C-6 carbonyl (Figure 1.8) [159]. In summary (Figure 1.8), the alkoxide-like catalytic serine, generated by substrate-assisted deprotonation, acts as a nucleophile to attack the C-6 carbonyl and generated a tetrahedral intermediate. This intermediate collapses, causing C-C bond cleavage, release of HHD and an acyl-enzyme intermediate. The latter undergoes deacylation via hydrolysis to release a carboxylate: HIP and benzoate in the case of HsaD and $BphD_{LB400}$, respectively.

Mechanistic studies of MCP hydrolases have been facilitated by the characteristic yellow colouration ($\lambda_{max,\ HOPDA}$ = 405 nm; $\lambda_{max,\ DSHA}$ = 392 nm) of the MCP and the use of spectrophotometric techniques [39, 77, 159]. Interestingly, mutation of the catalytic serine to alanine in MCP hydrolases results in a orange-coloured intermediate (ES^{red}) which, in the HsaD and DSHA system, absorbs at 459 nm [77]. Ruzzini *et al.* proposed that ES^{red} develops from torsion of the MCP π orbital system coupled with the lone pair localization at C-5. However, the exact nature of ES^{red} is unknown [159]. Titration of these catalytically inactive MCP hydrolases

into MCPs permits the measurement of dissociation constants (K_D) by following the increase in ES^{red} formation [77].

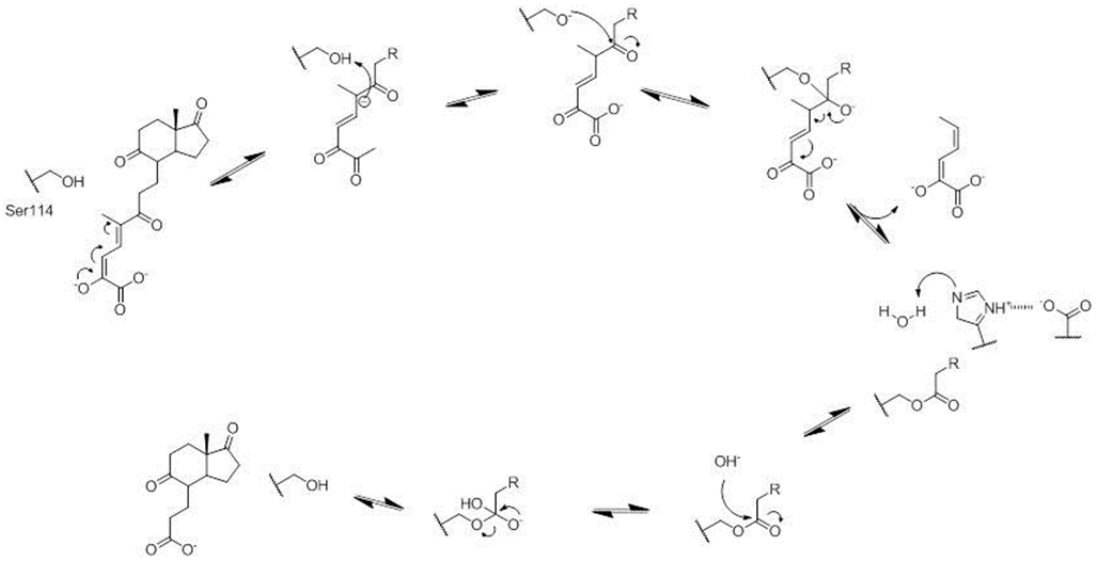


Figure 1.8 Mechanism of HsaD and MCP hydrolases.
R represents the Rings C and D moiety in DSHA for the HsaD catalyzed reaction or could represent a phenyl group in reactions catalyzed by BphD or DxnB2.

1.6 TetR family of transcriptional repressors

The cholesterol catabolic gene cluster is regulated by two TetR family repressors (TFR): KstR and KstR2 [83]. TFRs are named after a founding member involved in regulating tetracycline resistance [163] and is one of the most widely distributed families of transcriptional regulators in bacteria [164-166]. Despite significant sequence variation, the proteins form a conserved L-shaped α-helical structure featuring an N-terminal DNA-binding domain (DBD) and a larger C-terminal effector-binding domain (EBD) [164]. The DBD contains a helix-turn-helix motif involved in binding to operator DNA. TetR protomers associate into homodimers or higher oligomers that recognize palindromic sequences in the operator DNA. More specifically, an α-helix of the TetR DBD called the recognition helix forms specific electrostatic and aromatic contacts in the major groove of the DNA of the operator sequence [164]. The vast majority of

TFRs are repressors in the absence of their effector. Binding of the effector triggers a conformational change that shifts the position of the recognition helix, resulting in release of operator DNA by the regulator. Despite a generally conserved structure and mechanism of action, the specific position of the ligand-binding pocket in TFRs and its chemical composition vary dramatically, resulting in specific responses to a vast assortment of small molecules. This variation also makes it difficult to predict the chemical nature of the cognate ligand, necessitating the characterization of individual family members.

1.7 Coenzyme A transferases

1.7.1 The *ipdAB* genes encode a predicted CoA transferase

IpdAB, a predicted Class I CoA transferase (CoT), has recently garnered much attention due to its involvement in *Mtb* pathogenesis. Based on transposon mapping studies, *ipdA* and *ipdB* were predicted to be essential for the growth of *Mtb* in macrophages [52]. Interestingly, an *ipdAB* mutant in the related horse pathogen, *Rhodococcus equi* has been patented as a live vaccine for use in foals [23]. The *ipdAB* genes are also predicted to be essential for *in vitro* growth of *Mtb* on cholesterol based on Tn-seq studies [40].

The *ipdAB* genes are located within the KstR2 regulon and are strongly implicated in the catabolism of HIP [23, 91]. Both genes are up-regulated in the presence of HIP [91] and incubation of *R. equi* deficient in *ipdAB* accumulated 5OH-HIP in the supernatant causing the authors to predict that IpdAB would be involved in the production of 5OH-HIP-CoA via transferring a CoA moiety from a CoA donor such as acetyl-CoA or succinyl-CoA [23]. However, with the discovery that CoA thioesterification of HIP is catalyzed by FadD3 [79], it is unclear what role a CoA transferase would have in the catabolism of HIP.

1.7.2 Classes of CoA transferases

CoTs catalyze the reversible transfer of CoA from a thioester to a free carboxylate [143, 167]. Three classes of CoTs have been identified, the first two of which belong to the same superfamily. In Class I CoTs, CoA is transferred from the acyl group of the donor substrate to a free carboxylate of an acceptor substrate, typically a small organic acid (see Section 1.7.3) [168-170]. Class II CoTs are subunits of the citrate and citramalate lyase complexes, and catalyze the hydrolysis of small CoA thioesters using citrate or citramalate in a partial mechanism analogous Class I CoTs [171, 172]. Although Class III CoTs belong to a different superfamily, they appear to utilize a mechanism similar to that of Class II CoTs [173, 174].

Class I CoT make up the most diverse class of CoT as they are found in a wide range of catabolic and anabolic pathways [167]. These enzymes are widely promiscuous, displaying CoA transferase activity between an array of CoA donors, such as acetyl-CoA, propionyl-CoA or succinyl-CoA, and small acids, such as acetate, propionate, or succinate [143, 169, 175]. Although countless studies have involved expressing and purifying Class I CoT from across all kingdoms of life, three have been characterized in detail and are representative of the family: Succinyl-CoA: 3-oxoacid Coenzyme A transferase (SCOT) from pig heart is involved in the metabolism of ketone bodies and forms acetoacetyl-CoA from succinyl-CoA [143]; the β-ketoadipate:succinyl-CoA transferase from *Pseudomonas putida* (PcaIJ), involved in aromatic degradation; and the glutaconate CoA transferase from *Acidaminococcus fermentans* (GCT).

1.7.3 Structure of Class I Coenzyme A transferases

Class I CoTs are formed from two subunits typically assembled as an $\alpha_2\beta_2$ heterotetramer [169]. However in some CoT, such as SCOT, the α and β subunits are fused into a single enzyme which displays an α_4 oligomeric state [147]. The α and β subunits appear to evolutionarily related

as they display similar, but not identical, folds, however they often share low amino acid identity. The α subunit consists of a seven stranded parallel β-sheet sandwiched between seven α-helices, whereas the β subunit contains a six stranded (5 parallel, one antiparallel) β-sheet sandwiched between six α-helices [147, 169]. The active site of CoT occurs between the α/β dimeric interface and is formed from catalytic residues from both subunits. The active site pocket is typically connected to a long channel following the contours of the dimeric interface to an adenine binding pocket located near the enzyme's surface [147]. X-ray crystal structures of substrate bound Class I CoT show the acyl moiety bound within the active site with a highly conserved catalytic glutamic reside located less than 3 Å from the thioester carbonyl and the CoA moiety binding along the channel [147].

1.7.4 Mechanism of Class I Coenzyme A transferases

Class I CoT utilize a well characterized and highly conserved ping-pong mechanism in the transfer of a CoA moiety from a CoA donor to a small acid involving a highly conserved catalytic glutamic acid residue summarized in Figure 1.9 [143, 169, 170]. In this reaction, the conserved glutamic acid acts as a nucleophile, subsequently forming a glutamyl-CoA-enzyme intermediate prior to CoA transfer to a small acid [143, 175]. Formation of the glutamyl-CoA intermediate causes the release of the acyl group from the CoA donor, prior to binding of the acceptor acid within the active site which deacylates the catalytic glutamate [175]. In all Class I CoT characterized to date, the conserved glutamate is located in the β subunit or domain of the $\alpha_2\beta_2$ and α_4 oligomers, respectively, and occurs at the back of the active site [167, 169, 175, 176]. The glutamyl-CoA intermediate has been trapped using sodium borohydride ($NaBH_4$) which reduces the thioester carbonyl of the glutamyl-CoA intermediate to a hemithioacetal [143]. Therefore, Class I CoA transferases are irreversibly inhibited by $NaBH_4$.

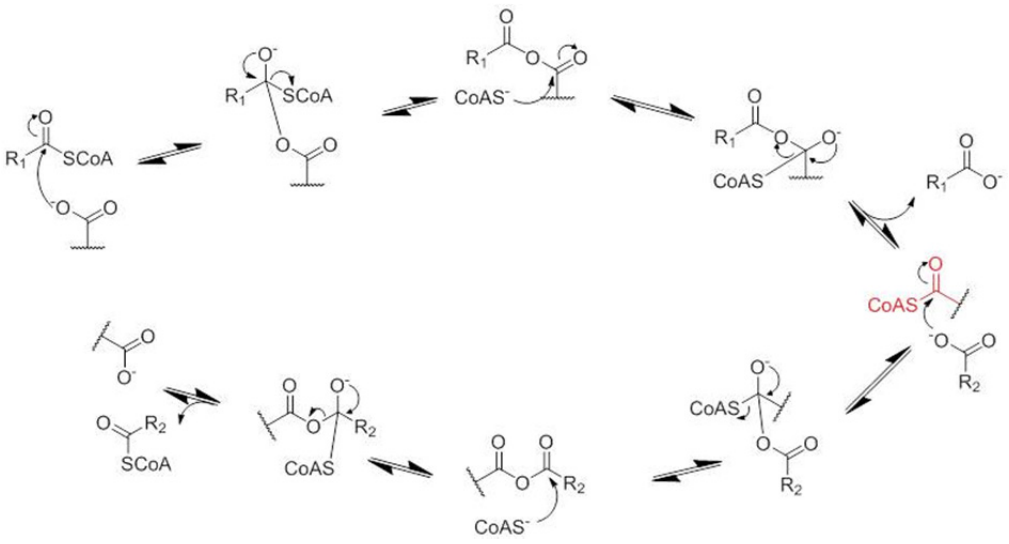


Figure 1.9 Mechanism of CoA transferases.
Red coloured species represents the glutamyl-CoA intermediate.

1.8 Aim of this study

The aim of this study is to elucidate metabolic and regulatory steps involved in the catabolism of cholesterol Rings C and D by *Mtb* and other actinomycetes. Chapters 2 – 5 describe four different studies related to this aim, and which together develop the central theme that CoA thioesters are essential in the catabolism of the last half of the steroid molecule.

Chapter 2 describes the role of CoA thioesters in determining the substrate specificity of the MCP hydrolase, HsaD. This work extends the observation by Capyk *et al.* (2011) that cholesterol side chain and Rings A and B degradation can occur concurrently to some extent in *Mtb* [69]. HsaD from *Mtb* and Mab_3810, the putative HsaD homolog, from *M. abscessus* were expressed heterologously and purified from *E. coli*. CoA thioesterified substrates were generated enzymatically from metabolites accumulating in a *hsaC* deficient strain of RHA1. Experiments relating to growth and gene expression of *Mab* were performed by Kirstin Brown. Overall, data for the completion of Rings A and B degradation prior to side chain degradation in at least some

mycobacteria is presented via the characterization of the recently identified steroid catabolic gene cluster in the related pathogen *M. abscessus*.

Chapter 3 describes the mechanism by which the cholesterol catabolite, HIP-CoA acts as an effector of KstR2. This work builds on the identification of HIP-CoA as the effector of KstR2 by Casabon *et al.* (2013) and is adapted from Crowe, *et al.* (2015) [91, 177]. More specifically, the strength of KstR2:HIP-CoA complex was determined using isothermal titration calorimetry (ITC) and a mechanism for HIP-CoA mediated regulation of KstR2 was proposed via the comparison of X-ray crystal structures of substrate-free KstR2 from RHA1 ($KstR2_{RHA1}$) and $KstR2_{Mtb}$·HIP-CoA. Experiments relating to the crystallization and structure determination of KstR2:HIP-CoA were performed by Peter Stogios.

Chapter 4 describes the elucidation of HIP catabolism using molecular genetics, metabolomics and biochemical approaches. KstR2 regulon genes were deleted in *Mtb*, RHA1, and/or *M. smegmatis*. Metabolites that accumulated in the presence of cholesterol were characterized using a combination of mass spectrometry, chemical synthesis, and nuclear magnetic resonance (NMR). This involved developing a method to purify and enrich CoA thioesters. Five KstR2 regulon enzymes were expressed heterologously, purified to homogeneity, and used to reconstitute the HIP catabolic pathway *in vitro*. Overall, the enzymes responsible for opening of cholesterol Rings C and D were identified and a near complete pathway for HIP catabolism is proposed. Much of the LC-MS experiments were performed by Dr. Israel Casabon and were aided by Jason Rogalski and Dr. Leonard Foster at The University of British Columbia. Experiments involving the culturing of *Mtb* were performed by Kirstin Brown and Jie Liu. The contents of this chapter were published in *mBio* in 2017 [178].

Chapter 5 provides the first biochemical and structural characterization of IpdAB. IpdAB, previously annotated as a CoA transferase, was identified as the enzyme responsible for opening cholesterol Ring C from the study described in Chapter 4 [178]. Further analysis of the enzyme demonstrated that IpdAB does not possess CoA transferase activity, but rather catalyzes the hydrolysis of cholesterol Ring C via a retro-Claisen like mechanism. A combination of X-ray crystal structures of substrate-free and substrate-bound IpdAB from RHA1 ($IpdAB_{RHA1}$), oligonucleotide-directed mutagenesis, stable isotopic labelling, and NMR was employed to characterize the mechanism by which IpdAB opens Ring C. The crystallization of IpdAB and structure determination was performed by Sean Workman, Liam Worrall, and Nobuhiko Watanabe from the Strynadka Lab at UBC.

Overall, this thesis describes a series of steps in the catabolism of cholesterol by *Mtb* beginning with the hydrolysis of the Ring A-opened MCP, DSHA, to the Rings C/D-opened metabolite MOODA-CoA. Similarly, the mechanism of regulation for these steps is described. The work included in this thesis contributes to the understanding of bacterial steroid degradation and provides a frame work for developing novel anti-tuberculosis therapeutics targeting cholesterol catabolism.

Chapter 2: **Steroid catabolism in *M. abscessus* provides insights into the order of cholesterol Rings A and B and sidechain degradation in actinobacteria**

2.1 Introduction

The classical paradigm of cholesterol catabolism states that the alkyl side chain is degraded prior to Rings A and B. However, work by Capyk *et al.* (2011) demonstrated that the Rings A and B degrading enzyme, KshAB, preferentially act on CoA thioesterified substrates, thereby indicating that side chain and Rings A and B degradation occur concurrently to some extent [69]. In contrast, Casabon *et al.* (2013) showed that side chain degradation must be completed before Rings C and D degradation occurs [79]. These studies established that the intervening enzymes, HsaAB, HsaC, and HsaD, could act on substrates with incompletely degraded side chains. However, the extent to which side chain and Rings A and B degradation occurs concurrently is unclear.

Herein, the order of cholesterol side chain and Rings A and B degradation was investigated in *Mtb*, the related pathogen, *M. abscessus* (*Mab*), and RHA1. The ability of *Mab* to grow on different steroids was determined. Comparative bioinformatic analyses of the steroid catabolic clusters in *Mab*, *Mtb* and *M. smegmatis* identified differences in orthologs of HsaD, the last enzyme involved in Rings A and B degradation, suggesting that the order of cholesterol side chain and Rings A and B degradation differs between mycobacterial species. Similarly, RHA1 deficient in *hsaC* and *hsaD* accumulated Rings A/B catabolites with partially degraded side chains. The substrate specificity of HsaD from *Mtb* ($HsaD_{Mtb}$) and *Mab* ($HsaD_{Mab}$), were compared for steroids with and without a completely degraded side chain. The results from steady-state kinetic analyses were validated by determining the dissociation constant (K_D) of a

catalytically inactive variant of $HsaD_{Mtb}$. Overall, we present the first *in vitro* evidence for steroid utilization by *Mab* and provide a new order of side chain and Rings A and B degradation of cholesterol in mycobacteria.

Experiments relating to the growth of *Mab* were performed by Kirstin Brown. All other experiments were performed by myself. Some of this chapter was prepared as a manuscript for submission in 2018 as:

Crowe, A. M, Brown, K., Kulkarani, J., Yam, K., and Eltis, L. D. (**2018**) The unusual convergence of steroid catabolic pathways in *Mycobacterium abscessus*.

2.2 Materials and methods

2.2.1 Chemicals and reagents

CoASH and ATP were purchased from Sigma-Aldrich (St. Louis, MO). All reagents used were of HPLC or analytical grade. Phusion High-Fidelity DNA polymerase, restriction enzymes, and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). CasI, an acyl-CoA ligase, was purified as previously described [179]. Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA). Water for buffers was purified using a Barnstead Nanopure Diamond^{TM} system (Dubuque, IA) to a resistivity of at least 18 megaohms.

2.2.2 DNA manipulation

DNA was propagated, purified, amplified and cloned using standard techniques [180]. Expression plasmids for $HsaD_{Mtb}$ and its S114A variant were described previously [77]. The *hsaD*$_{Mab}$ gene (*mab_3810*) was amplified from *M. abscessus* ATCC 19977 genomic DNA using primer sequences 5'-GAACCATATGGACGTGACATACGAGGGCACCA-3' and 5'-CCCAAGCTTTTAGCGCCGCAAAAAGTCGACG-3'. The amplicon was digested with NdeI and HindIII and ligated into pET41b(+) (EMD Millipore) generating pETHDMAB. DNA was

sequenced by Genewiz (South Plainfield, NJ). Plasmids were transformed into *Escherichia coli* Rosetta 2 (DE3) (EMD Millipore) by electroporation using a MicroPulser from Bio-Rad (Hercules, CA). The *hsaC* and *hsaD* (RHA1_RS22130) deficient strains of RHA1 were generated as previously described [25].

2.2.3 Phylogenetic analyses

Phylogenetic analyses were performed using the PhyML server available on the phylogeny.fr web service [181, 182] using structure-based alignments of *meta*-cleavage product (MCP) hydrolases generated by TCOFFEE ESPRESSO [183, 184]. Phylogenetic tree figures were prepared using FigTree v1.4.3.

2.2.4 Growth of bacterial strains

Mab ATCC19977 was grown on M9 minimal media + 0.5% tyloxapol supplemented with either 0.5 mM cholate, 0.5 mM 4-AD, 0.5 mM cholesterol, or 4.5 mM glycerol at 37^{o}C. Growth was measured by optical density at 600 nm and by CFU enumeration. RHA1 WT and mutants were grown on solid media containing M9G and biphenyl. Single colonies were used to inoculate 50 ml of M9G [185] containing 20 mM sodium pyruvate, and cultures were grown at 30^{o}C to turbidity (~3 d). For protein production, *E. coli* Rosetta 2(DE3) containing pT7HD1, pEMHSA, or pETHDMAB were grown in 1 L LB supplemented with 100 μg ml^{-1} ampicillin (pT7HD1 and pEMHSA) or 25 μg ml^{-1} kanamycin (pETHDMAB) and 34 μg ml^{-1} chloramphenicol at 37^{o}C and 200 rpm. At mid log (OD_{600} = 0.6), isopropyl β-D-1-thiogalactopyranoside was added to 0.5 mM to induce *hsaD* expression. Cells were immediately transferred to 30^{o}C, grown overnight, then harvested by centrifugation at 4^{o}C (4,000 g, 15 min) and frozen at -80^{o}C until use.

2.2.5 Metabolite analysis

Metabolites were analyzed in the culture supernatants of pyruvate-grown RHA1 cells incubated with cholesterol and of steroid-grown *Mab*. For RHA1, cells were grown to mid log (OD_{600} = 0.6) on 500 ml M9G containing pyruvate, harvested by centrifugation (4000 × *g*, 16°C, 25 min), suspended in 50 ml M9G containing 0.5 mM solid cholesterol and incubated at 30°C for 24 h in 250 ml baffled flasks. At regular time intervals, aliquots of 1 ml were withdrawn and cells were removed by centrifugation. To extract metabolites, 750 µl of the supernatant was acidified using 25 µl glacial acetic acid and was mixed vigorously with 750 µl of ethyl acetate. 5α-Cholestane was added to 26 µM as an internal standard. Supernatants of *Mab* cultures were handled similarly except that 0.3 ml aliquots were taken, 0.16 mM 5α-cholestane was added, samples were acidified with 10 µl glacial acetic acid, and metabolites were extracted with 0.3 ml ethyl acetate. Organic layers were dried under nitrogen, suspended in 50 µl pyridine and derivatized with trimethylsilane (TMS). Samples were analyzed by gas chromatography coupled-mass spectrometry (GC-MS) as previously described [76].

2.2.6 Purification of $HsaC_{Mtb}$, $HsaD_{Mtb}$, $HsaD_{Mtb}$ S114A, and $HsaD_{Mab}$

$HsaC_{Mtb}$ was purified as previously described [39]. $HsaD_{Mtb}$, $HsaD_{Mtb}$ S114A, and $HsaD_{Mab}$ were purified as previously described for $HsaD_{Mtb}$ using the following modifications [77]. Following purification using Source15 Q anionic exchange resin (GE Healthcare), HsaD-containing fractions were concentrated to ~4 mg ml^{-1}. $HsaD_{Mtb}$ (WT and S114A) were brought to 1 M ammonium sulphate in 20 mM HEPES, pH 7.5 then loaded onto 8 ml of Source15 Phenyl resin (GE Healthcare) and eluted over a gradient of 1 to 0 M ammonium sulphate in 20 mM HEPES, pH 7.5 over 20 column volumes. HsaD-containing fractions were pooled, concentrated to ~10 mg ml^{-1} using a stirred cell concentrator equipped with a 10K amicon ultracentrifugation

membrane (EMD Millipore) then flash frozen in liquid nitrogen and stored at -80^{o}C until use. $HsaD_{Mab}$ was brought to 1.3 M ammonium sulphate in 20 mM HEPES, pH 7.5 causing a white precipitate to form. After incubating for 1 h at 4^{o}C, the $HsaD_{Mab}$ precipitate was harvested by centrifugation (4,000 g; 10 min) and the pellet was solubilized in 20 mM HEPES, pH 7.5. $HsaD_{Mab}$ was exchanged into 20 mM HEPES, pH 7.5, then concentrated to 16 mg ml^{-1} using a 10K Amicon Ultracel^{R} unit (Merck-Millipore). All proteins were prepared to greater than 95% apparent homogeneity as assessed by SDS PAGE analysis. Protein concentrations were determined by Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) using the manufacturer's protocols.

2.2.7 Preparation of substrates

3,4-Dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (DHSA) and 3,4-dihydroxy- 17-isopropinoyl- 9,10-seconandrost-1,3,5(10)-triene-9-one (DHSBNC) were prepared from cholesterol using an *hsaC* mutant of RHA1 as a biocatalyst as described previously with the following revisions [39]. Supernatant extracts containing DHSA and DHSBNC were loaded onto a Hewlett Packard Series 1100 HPLC system equipped with a Luna 5u Silica (2) 100A HPLC column (Phenomenex) operated at 1 ml min^{-1}. DHSA and DHSBNC were separated using a 28 ml gradient of 100% hexane and 0.5% acetic acid into 28% ethyl acetate and 0.5 % acetic acid. DHSBNC and DHSA eluted at 22 and 26 min, respectively. The purity and identity of DHSA and DHSBNC were verified using GC-MS on an Agilent 6890 series GC equipped with an HP-5ms 30 m x 250 μm capillary column (Hewlett Packard) as described previously [39].

DHSBNC-CoA was produced by reacting 1 mM DHSBNC with 1.25 mM CoASH, 1.5 mM ATP, 5 mM $MgCl_2$, and 2 μM CasI in 20 mM HEPES, pH 7.5 at 22^{o}C for 1 hour. The

resulting DHSBNC-CoA was purified by HPLC using a Luna 3u PFP(2) 100A (50 x 4.6 mm) column in a gradient of 100 mM ammonium acetate, pH 4.5 into 100 mM ammonium acetate, pH 4.5, 90% methanol, over 20 min. DHSBNC-CoA eluted at 18 min. Methanol was removed from the purified DHSBNC-CoA under nitrogen and the sample was diluted 1:3 with water then desalted using a Strata-X 33u 30 mg solid phase extraction column (Phenomenex) and eluted in methanol using the manufacturer's protocols. The mole recovery of DHSBNC-CoA was ~80%. DSHA, 4,5-9,10-diseco-3-hydroxy-5,9-dioxo-23,24-bisnorchola-1(10),2-dien-4,22-dioate (DSHBNC), and DSHBNC-CoA were generated by incubating 1 mM of the corresponding catechol with 2 μM HsaC in I = 0.1 M KP_i, pH 7.5 for 10 min. DSHA, DSHBNC, and DSHBNC-CoA each had a molar absorptivity at 392 nm (ε_{392nm}) of 3.8 mM^{-1} cm^{-1}, determined using an Oxygraph oxygen electrode (Hansatech Instruments Ltd) as described previously [77].

2.2.8 Steady- state kinetic analyses

The HsaC and HsaD-catalyzed reactions were followed spectrophotometrically by measuring the increase or decrease in absorbance at 392 nm associated with the production or hydrolysis of the MCP, respectively, using a Varian Cary 5000 spectrophotometer equipped with a thermostatted cuvette holder (Varian Canada, Mississauga, Canada). Assays were performed in 250 μl KP_i, pH 7.5 (I = 0.1 M) at 25 ± 0.5 ^oC. Reactions were initiated by adding HsaD, whose final concentration was such that the progress curves were linear over at least 2 min. Initial velocities were determined from the progress curves using least fit squares as calculated by the kinetics module of the Cary WinUV software. Steady-state parameters were determined by fitting the Michaelis-Menten equation to the data using LEONORA [186]. Buffers were prepared on the day of use. MCPs were prepared and protein was thawed immediately prior to use. The

concentration of MCP stock was determined before and after the experiment to correct for non-enzymatic hydrolysis.

2.2.9 K_D determination

Dissociation constants (K_D) for complexes of $HsaD_{Mtb}$ S114A with each of DSHA, DSHBNC and DSHBNC-CoA were determined essentially as previously described for DSHA [77]. Briefly, the variant was titrated into 5 μM substrate in an initial volume (V_i) of 200 μl KP_i, pH 7.5 (I= 0.1 M) at 25°C and measuring the change in absorbance at 456 nm. The dissociation constant was calculated by plotting the data to the quadratic equation (Equation 1) using the nonlinear curve fitting function of R where A_o = initial absorbance 456 nm, A_{max} = maximum absorbance 456 nm, [E] = concentration $HsaD_{Mtb}$ S114A (μM), and [S] = substrate concentration (μM). Dilution of the substrate was accounted for using Equation 2 where S_o is the measured initial substrate concentration and V_T = the assay volume after each titrant addition. In fitting Equation 1, [S] and [E] were treated as independent variables, and A_o, A_{max}, and K_D, as dependent variables.

Equation 1
$$\Delta A_{456\,nm} = A_o + A_{max}\left(\frac{([E]+[S]+K_D)-\sqrt{([E]+[S]+K_D)^2-4[E][S]}}{2[E]}\right.$$

Equation 2
$$[S] = S_o\left(\frac{V_i}{V_T}\right)$$

2.3 Results

2.3.1 Bioinformatic analysis of the steroid catabolic gene cluster in *M. abscessus*

To better characterize the steroid catabolic genes of *Mab*, a Protein BLAST search was performed for each of 69 proteins from *Mtb* that have been implicated in cholesterol catabolism in various genomic studies as listed in (Table 2.1) [25, 32, 40, 47]. Orthologs of 65 of these

proteins were identified based on being reciprocal best hits, and shared 60-91% amino acid sequence identity with the corresponding protein in *Mtb* (Table 2.1). The four genes for which orthologs were not found were Rv3519, Rv3528c, FdxB and Rv1106c. The first three of these do not have an experimentally validated role in cholesterol catabolism and none have been predicted to be essential for growth on this sterol [40]. The role of the fourth, a 3β-hydroxysteroid dehydrogenase [70], has been debated due in part to functional redundancy with Rv3502c and Rv3409c, a cholesterol oxidase [72]. In addition, it is unclear whether *Mab* possesses an ortholog of Cyp142. The identified best hit, Mab_3825, has 29-34% amino acid sequence identity with four cytochromes P450 (Cyp) from *Mtb*: Cyp124, Cyp125, Cyp142, and Cyp130. However, Cyp142 appears to be dispensable for growth on cholesterol due to redundancy with Cyp125 [58].

The 65 predicted cholesterol catabolic genes of *Mab* share significant synteny to those in *Mtb*, *M. smegmatis* and RHA1. Moreover, the predicted operonic structure is largely conserved, with the side chain and Rings A/B degradation genes occurring downstream of KstR-binding motifs and the Rings C/D degradation genes occurring downstream of KstR2-binding motifs, as in related bacteria. Nevertheless, the clustering of these genes in *Mab* appears to be unusual in at least two respects. First, the cholesterol catabolic genes in *Mab* are arranged in two clusters separated by ~1.39 Mb: the first comprises genes encoding steroid uptake and side chain degradation (Table 2.1; yellow and green, respectively) and the second encoding Rings A/B and C/D degradation (Table 2.1;pink and blue, respectively). In comparison, most of the cholesterol catabolic genes are clustered within 105 kb in *Mtb* [25], 145 kb in *M. smegmatis* [187] and 183 kb in RHA1 [25]. A second unusual feature of the gene cluster in *Mab* concerns the *hsaACDB* operon, which encodes fragmentation of Rings A/B. Unlike in *Mtb*, *M. smegmatis* and RHA1,

the operon in *Mab* lacks *hsaD* (Figure 2.1A). This gene occurs in a predicted single gene operon, 1.8 Mbp from the *hsaACB* operon and encodes the only predicted HsaD ortholog in the *Mab* genome. Interestingly, a mycobacterial KstR-binding site as defined by Kendall *et al*. [32, 188] occurs 10 bp upstream of the *hsaD* start codon.

Searching the *Mab* genome for cholesterol catabolic genes revealed the presence of additional homologs of Rings A/B degradation genes that are not present in *Mtb* (Table 2.1; Figure 2.1A). Four of these occur in a cluster adjacent to *mab_0082c*, which encodes a predicted PadR-family transcriptional repressor. Further analyses revealed that these five genes are reciprocal best hits of homologs that specify the catabolism of 4-AD in *M. smegmatis* [27]. Again, in contrast to the 4-AD catabolic gene cluster of *M. smegmatis*, the cluster of *Mab* does not include a homolog of *hsaD.* However, the *Mab* genome harbours additional homologs *kshB2* and *kshA3* (Table 2.1). Finally, no KstR-binding motifs were identified upstream of any these additional steroid catabolic genes.

Table 2.1 The steroid and monoaromatic catabolic genes in *M. abscessus* ATCC 19977

Cholesterol[c,d]					
Mab gene	Gene name[a]	Gene product	H37Rv	*M. smegmatis*	Identity (%)[b]
MAB_0013c[e]	*nat*	N acetyl-transferase	*rv3566*	MSMEG_0306	32
MAB_0579c	*kstR*	TetR transcriptional repressor	*rv3574c*	MSMEG_6042	71
MAB_0580	*chsE3*	acyl-CoA dehydrogenase	*rv3573c*	MSMEG_6041	61
MAB_0582c	*rv3572*	hypothetical Protein	*rv3572*	MSMEG_6040	43
MAB_0583c	*kshB1*[c]	ketosteroid hydroxylase, reductase	*rv3571*	MSMEG_6039	67
MAB_0584	*hsaA1*	3-HSA hydroxylase	*rv3570c*	MSMEG_6038	79
MAB_0585	*hsaC1*	DHSA dioxygenase	*rv3568c*	MSMEG_6036	70
MAB_0586	*hsaB*	3-HSA hydroxylase, reductase	*rv3567c*	MSMEG_6035	77
MAB_0593c	*fadE33*	acyl-CoA dehydrogenase	*rv3564c*	MSMEG_6016	63
MAB_0594c	*fadE32*	acyl-CoA dehydrogenase	*rv3563c*	MSMEG_6015	65
MAB_0595c	*fadE31*	acyl-CoA dehydrogenase	*rv3562c*	MSMEG_6014	74
MAB_0596c	*fadD3*	HIP-CoA synthetase	*rv3561*	MSMEG_6013	65
MAB_0597	*fadE30*	acyl-CoA dehydrogenase	*rv3560c*	MSMEG_6012	77
MAB_0598	*ipdF*	short chain dehydrogenase	*rv3559c*	MSMEG_6011	82
MAB_0599	*kstR2*	TetR transcriptional repressor	*rv3557c*	MSMEG_6009	71
MAB_0600	*fadA6*	acyl-CoA thiolase	*rv3556c*	MSMEG_6008	79
MAB_0603c	*ipdC*	acyl-CoA reductase	*rv3553*	MSMEG_6004	81
MAB_0604c	*ipdB*	COCHEA-CoA hydrolase	*rv3552*	MSMEG_6003	74
MAB_0605c	*ipdA*	COCHEA-CoA hydrolase	*rv3551*	MSMEG_6002	78
MAB_0606c	*echA20*	HIEC-CoA hydrolase	*rv3550*	MSMEG_6001	78
MAB_0607	*rv3549c*	short chain dehydrogenase	*rv3549c*	MSMEG_6000	69
MAB_0608	*rv3548c*	short chain dehydrogenase	*rv3548c*	MSMEG_5999	78
MAB_0609c	*rv3547*	hypothetical Protein	*rv3547*	MSMEG_5998	52
MAB_0610		hypothetical Protein	-	MSMEG_5997	-
MAB_0612c	*fadA5*	acyl-CoA thiolase	*rv3546*	MSMEG_5996	76
MAB_0613	*cyp125*	27-cholesterol oxygenase	*rv3545c*	MSMEG_5995	77
MAB_0614	*chsE1*	acyl-CoA dehydrogenase	*rv3544c*	MSMEG_5994	57
MAB_0615	*chsE2*	acyl-CoA dehydrogenase	*rv3543c*	MSMEG_5993	75
MAB_0616	*chsH2*	MaoC-like enoyl-CoA hydratase	*rv3542c*	MSMEG_5992	71
MAB_0617	*chsH1*	MaoC-like enoyl-CoA hydratase	*rv3541c*	MSMEG_5991	75
MAB_0618	*ltp2*	lipid Transfer Protein	*rv3540c*	MSMEG_5990	83
MAB_0621c	*hsd4B*	hydroxyl steroid dehydrogenase	*rv3538*	MSMEG_5943	62
MAB_0622c	*kstD1*	ketosteroid dehydrogenase	*rv3537c*	MSMEG_5941	78
MAB_0624	*hsaE*	hydratase	*rv3536c*	MSMEG_5940	71
MAB_0625	*hsaG*	acetaldehyde dehydrogenase	*rv3535c*	MSMEG_5939	87
MAB_0626	*hsaF*	4-hydroxy-2-oxovalerate aldolase	*rv3534c*	MSMEG_5937	91
MAB_0685[e]	*aspB*	aminotransferase	*rv3565*	MSMEG_6017	59
MAB_0718c[e]	*rv0139*	oxidoreductase	*rv0139*	MSMEG_6474	29
MAB_3719c	*choD*	cholesterol oxidase	*rv3409c*	MSMEG_1604	79
MAB_3810[e]	***hsaD***	**DSHA hydrolase**	***rv3569c***	**MSMEG_6037**	**36**
-[f]	*cyp142*	cytochrome P450	*rv3518c*	MSMEG_5918	
MAB_4148c	*mce4F*	ABC transporter	*rv3494c*	MSMEG_5895	65
MAB_4149c	*mce4E*	ABC transporter	*rv3495c*	MSMEG_5896	57
MAB_4150c	*mce4D*	ABC transporter	*rv3496c*	MSMEG_5897	58
MAB_4151c	*mce4C*	ABC transporter	*rv3497c*	MSMEG_5898	63
MAB_4152c	*mce4B*	ABC transporter	*rv3498c*	MSMEG_5899	66
MAB_4153c	*mce4A*	ABC transporter	*rv3499c*	MSMEG_5900	58
MAB_4154c	*yrbE4B*	ABC transporter	*rv3500c*	MSMEG_5901	78
MAB_4155c	*yrbE4A*	ABC transporter	*rv3501c*	MSMEG_5902	87
MAB_4156c	*hsd4A*	hydroxyl steroid dehydrogenase	*rv3502c*	MSMEG_5903	72
MAB_4157c	*fdxD*	Ferredoxin	*rv3503c*	MSMEG_5904	70
MAB_4158	*chsE4*	acyl-CoA dehydrogenase	*rv3504*	MSMEG_5906	84
MAB_4159	*chsE5*	acyl-CoA dehydrogenase	*rv3505*	MSMEG_5907	65
MAB_4160	*fadD17*	acyl-CoA synthetase	*rv3506*	MSMEG_5908	60
MAB_4162c	MSMEG_5913	nitronate monooxygenase	-	MSMEG_5913	-
MAB_4163c	*fadD19*	27-cholestenoyl-CoA synthetase	*rv3515c*	MSMEG_5914	68
MAB_4164	*echA19*	enoyl-CoA hydratase	*rv3516*	MSMEG_5915	79
MAB_4166c	*Rv3520c*	luciferase-like Protein	*rv3520c*	MSMEG_5920	78
MAB_4167	*Rv3521*	hypothetical Protein	*rv3521*	MSMEG_5921	73
MAB_4168	*ltp4*	lipid Transfer Protein	*rv3522*	MSMEG_5922	77
MAB_4169	*ltp3*	lipid Transfer Protein	*rv3523*	MSMEG_5923	87

Cholesterol (cont.)					
Mab gene	Gene name[a]	Gene product	H37Rv	*M. smegmatis*	Identity (%)[b]
MAB_4173	*kshA1*	ketosteroid hydroxylase, oxygenase	*rv3526*	MSMEG_5925	75
MAB_4174	*rv3527*	hypothetical Protein	*rv3527*	MSMEG_5927	53
MAB_4177c	*rv3529c*	hypothetical Protein	*rv3529c*	MSMEG_5930	72
MAB_4178c	*rv3530c*	hypothetical Protein	*rv3530c*	MSMEG_5931	68
MAB_4179c	*rv3531c*	hypothetical Protein	*rv3531c*	MSMEG_5932	69
4-AD catabolism					
Mab gene	Gene name[a]	Gene product	H37Rv	*M. smegmatis*	Identity (%)[b]
-	*hsaD2*	DSHA hydrolase	-	MSMEG_2900	-
-	*kstD3*	ketosteroid dehydrogenase	-	MSMEG_2867	-
MAB_0075	*hsaA2*	3-HSA hydroxylase, oxygenase	-	MSMEG_2892	77
MAB_0076	*hsaC2*	DHSA dioxygenase	-	MSMEG_2891	75
MAB_0080c	*kshA2*	ketosteroid hydroxylase, oxygenase	-	MSMEG_2870	75
MAB_0081c	*kstD2*	ketosteroid dehydrogenase	-	MSMEG_2869	66
MAB_0082c	*MSMEG_2868*	PadR-like transcriptional repressor	-	MSMEG_2868	52
MAB_3627c	*kshA3*	ketosteroid hydroxylase, oxygenase	-	-	-
MAB_3918	*kshB2*	ketosteroid hydroxylase, reductase	-	MSMEG_2893	57
Isopropyl benzene					
Mab gene	Gene name	Gene product	*R. erythropolis* BD2		Identity (%)[g]
MAB_4358c	-	2,3-dihydroxybiphenyl 1,2 dioxygenase III	-		-
MAB_4361	*mphD*	2-keto-4-pentenoate hydratase	-		-
MAB_4365c	*ipbB*	IPB dihydrodiol dehydrogenase	PBD2.158		62
MAB_4366c	*ipbD*	HOMODA hydrolase	PBD2.174		50
MAB_4367c	*ipbA2*	IPB dioxygenase, small subunit	PBD2.154		60
MAB_4368c	*ipbA4*	IPB dioxygenase, ferredoxin reductase	PBD2.156		49
MAB_4369c	*ipbA3*	IPB dioxygenase, ferredoxin	PBD2.155		50
MAB_4370c	*ipbA1*	IPB dioxygenase, large subunit	PBD2.153		54
MAB_4372	-	3-(2,3-dihydroxyphenyl) propionic acid dioxygenase	-		-

[a]In the absence of a gene name, the H37Rv (cholesterol) or *M. smegmatis* (4-AD) gene ID is listed.
[b]For cholesterol catabolic gene products, the indicated amino acid sequence identity is between *M. abscessus* and *M. tuberculosis* proteins. For 4-AD catabolic gene products, the value is between *M. abscessus* and *M. smegmatis* proteins.
[c]Genes in purple, green, red and blue font are involved in cholesterol uptake, side chain degradation, Rings A/B degradation and Rings C/D degradation, respectively.
[d] No reciprocal best hit for *hsd* (*rv1106c*), *fadD18* (*rv3513c*), *rv3519, rv3528c,* and *fdxB* (*rv3554*), predicted to be involved in cholesterol catabolism in *M. tuberculosis,* were identified in *M. abscessus*
[e]Reciprocal best hit between *M. abscessus* ATCC 19977 and *M. tuberculosis* H37Rv
[f]Mab_3825 occurs in this cluster but is not the reciprocal best hit of Cyp142 from *M. tuberculosis*, sharing 34%, 33%, 33%, and 29% amino acid identity with *M. tuberculosis* Cyp124, Cyp125, Cyp142 and Cyp130, respectively.
[g]CLUSTAL alignment (BLOSM62) identity displayed.

2.3.2 *M. abscessus* possesses a HsaD ortholog dissimilar from other actinobacteria

In light of the finding that $hsaD_{\text{Mab}}$ does not cluster with known steroid catabolic genes, phylogenetic analyses were preformed to better understand the gene product's relationship with other HsaDs. Previous analyses had identified three Subfamilies (I to III) comprising enzymes involved in the catabolism of biphenyl/steroids, alkylated benzenes and dibenzofuran-like compounds, respectively [189]. By contrast, the updated phylogenetic analyses performed using a structure-based alignment and a larger number of sequences identified seven subfamilies of MCP hydrolases (Figure 2.1B). The enzymes appear to group largely according to their substrate

specificities, as seen previously [189]. However, the newly defined Subfamily I comprises only biphenyl-degrading homologs and three distinct subfamilies of steroid-degrading MCP hydrolases were identified: Subfamily IV, comprising cholesterol catabolic enzymes from Actinobacteria; Subfamily V, cholate and 4-AD catabolic enzymes from Actinobacteria; and Subfamily VI, cholate and testosterone catabolic enzymes from Gram negative strains. Subfamily VII includes enzymes involved in the degradation of 3-(hydroxyphenyl)propionate and related compounds.

$HsaD_{Mab}$ belonged to Subfamily V which includes the cholate-degrading HsaD3 from RHA1 [22, 93] as well as HsaD2 from RHA1 and HsaD2 from *M. smegmatis*, both of which are involved in 4-AD catabolism [22, 190]. These enzymes are not predicted to act on steroids with partially degraded side chains [93, 190], in contrast to the cholesterol-catabolising $HsaD_{RHA1}$ and $HsaD_{Mtb}$.

A second MCP hydrolase, Mab_4366c, is encoded within the *Mab* genome and displays 36% amino acid sequence identity (88% coverage) with $HsaD_{Mtb}$. Although the phylogenetic analysis indicated that it occurs within Subfamily IV (Figure 2.1), the bioinformatics evidence indicates that it is unlikely to act on a steroid-derived substrate. The *mab_4366c* gene does not occur adjacent to known steroid catabolic genes nor is a KstR motif located upstream. More specifically, *mab_4366c* occurs within a predicted six-gene operon, *mab_4365c* - *mab_4370c*, encoding enzymes predicted to be involved in isopropyl benzene (Ipb) degradation (Table 2.1) [191]. A similarly organized Ipb cluster was characterized in *Rhodococcus erythropolis* BD2 whose gene products are the reciprocal best hits of the *M. abscessus* genome are located within this cluster sharing 35 – 45% amino acid identity (Table 2.1) [192, 193]. Therefore, Mab_4366c

is predicted to be involved in isopropyl benzene catabolism and is provisionally annotated as IpdD.

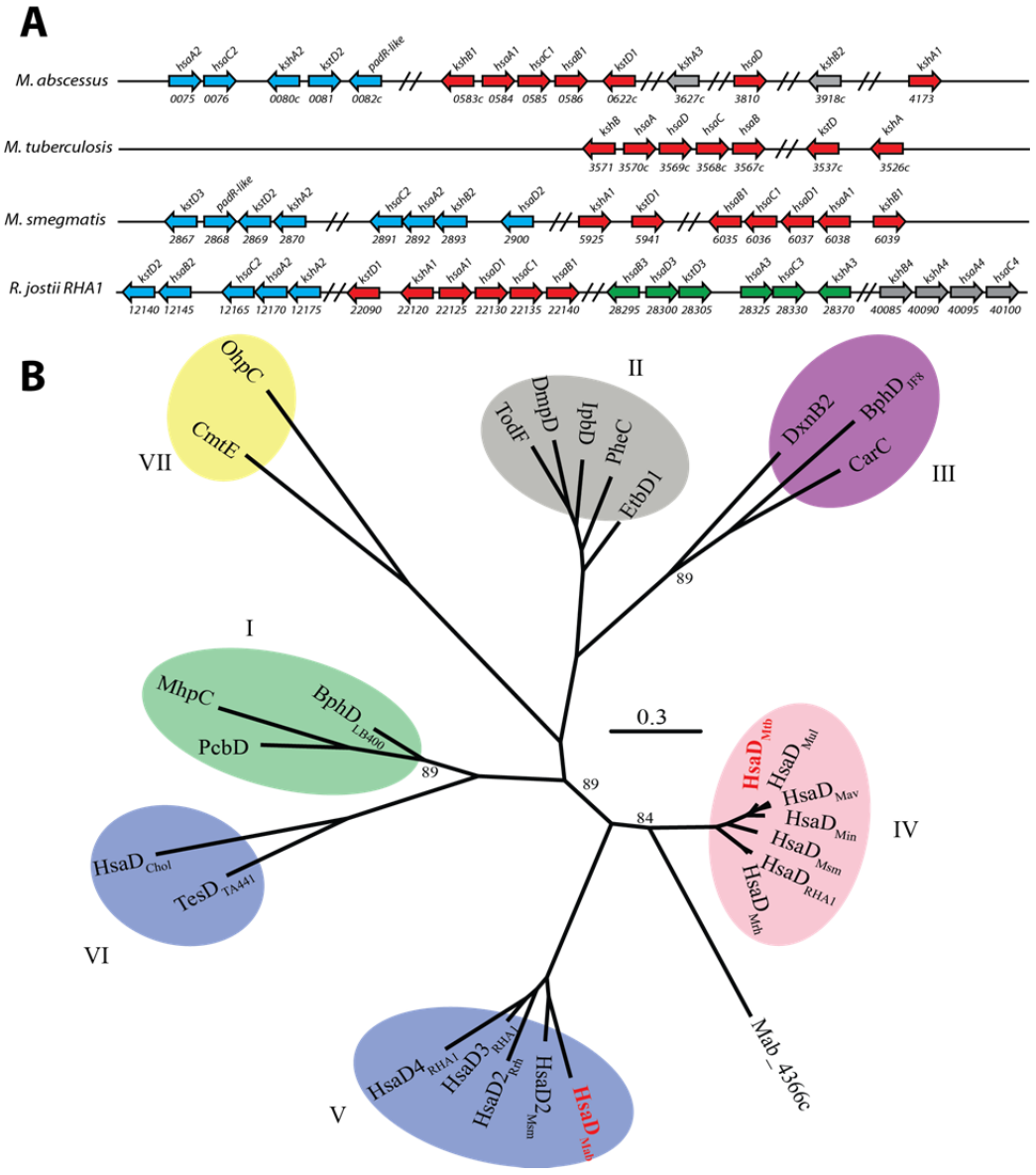


Figure 2.1 The steroid catabolic gene clusters in *M. abscessus*
(A) Organization of actinobacterial Rings A and B degrading enzymes involved in cholesterol (red), ADD (blue), cholate (green), and undetermined (grey) steroid catabolism. Genes are identified above by their respective gene product and below by their loci. Omitted are *mab_, rv, msmeg_,* and *RHA1_RS* prior to the gene loci for *M. abscessus, M. tuberculosis, M. smegmatis,* and *R. jostii* RHA1. (**B**) Phylogenetic analysis of MCP hydrolases**.** The tree was generated using a structure-based alignment of sequences the following MCP hydrolases: HOPDA hydrolase from *Bacillus sp.* Strain JF8 ($BphD_{JF8}$, BAC79225); HOPDA hydrolase from *Burkholderia xenovorans* LB400 ($BphD_{LB400}$, WP011494293.1); HOPDA hydrolase from *E. coli* (MhpC, BAA13054.1); 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoate hydrolase from s*phingomonas sp. strain KA1* (CarC, BAC56762); 6-isopropyl-HOHD hydrolase from *Pseudomonas putida F1* (CmtE, BAB17778); MCP hydrolase from *Pseudomonas sp. strain*

CF600 (DmpD, P19076); chloro-HOPDA hydrolase from *Sphingomonas wittichii RW1* (DxnB2, WP012049251.1); HOHD hydrolase from *R. jostii RHA1* (IpbD, BAA31163); HOHD hydrolase from *Pseudomonas fluorenscens* (EtbD1, BAA12150); DSHA hydrolase from *M. avium* ($HsaD_{Mav}$, WP033725372.1); DSHA hydrolase from *M. intracellulare* ($HsaD_{Min}$, WP014381744.1); DSHA hydrolase from *M. smegmatis* ($HsaD_{Msm}$, AFP42309.1); DSHA hydrolase from *M. tuberculosis* ($HsaD_{Mtb}$, NP_218086.1); DSHA hydrolase from *M. ulcerans* ($HsaD_{Mul}$, WP011741784.1); DSHA hydrolase from *Rhodococcus rhodochrous* ($HsaD_{Rrh}$, WP019746906.1); DSHA hydrolase from *R. jostii RHA1* ($HsaD_{RHA1}$, BAA98136); DSHA hydrolase from *M. smegmatis* ($HsaD2_{Msm}$, AFP39291.1); MCP hydrolase from *R. rhodochrous* ($HsaD2_{Rrh}$, WP085469424.1); TDSHA hydrolase from *R. jostii* RHA1 ($HsaD3_{RHA1}$, ABG97574); MCP hydrolase from *R. jostii* RHA1 ($HsaD4_{RHA1}$, ABH00058); DSHA from *M. abscessus* ($HsaD_{Mab}$, Mab_3810); 2-hydroxymuconic-semialdehyde hydrolase from *R. jostii* RHA1 (OhpC, ABG92355); HOPDA hydrolase from *Pseudomonas sp. strain DJ12* (PcbD, BAA07955); 2-hydroxymuconic-semialdehyde hydrolase from *R. jostii* RHA1 (PheC, ABG99128); TDSHA hydrolase from *Comamonas testosteroni TA441* ($TesD_{TA441}$, BAC67693); TDSHA hydrolase from *Pseudomonas putida* Chol1 ($HsaD_{Chol1}$; WP008568663.1); 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase from *Pseudomonas putida F1* (TodF, P23133); and putative HOHD hydrolase from *M. abscessus* (Mab_4366c, CAM64437).

2.3.3 *M. abscessus* grows on cholesterol and 4-AD but not cholate

The identification of multiple clusters of steroid catabolic genes in *Mab* prompted investigations into the ability of the bacterium to grow on different steroids. The strain grew to stationary phase within 72 h on cholesterol with a doubling time of approximately 12 hours, and a concurrent depletion of cholesterol from the media was observed by GC-MS (Figure 2.2A). *Mab* also grew on 4-AD, although to a lower yield, commensurate with the steroid's lack of a side chain (Figure 2.2B). Further, 4-AD-grown cultures turned pink, consistent with the accumulation and non-enzymatic oxidation of DHSA, a catechol [39] (Figure 2.2D). GC-MS analysis confirmed the presence of DHSA based on its molecular ion (*m/z* = 460) and fragmentation pattern (Figure 2.2E). The accumulation of DHSA could reflect the inefficient expression of *hsaD* in the presence of 4-AD due to repression by KstR leading to upstream metabolite accumulation. Finally, *Mab* did not grow on cholate or detectably deplete cholate from the media (Figure 2.2C), consistent with bacterium's predicted lack of cholate catabolic genes.

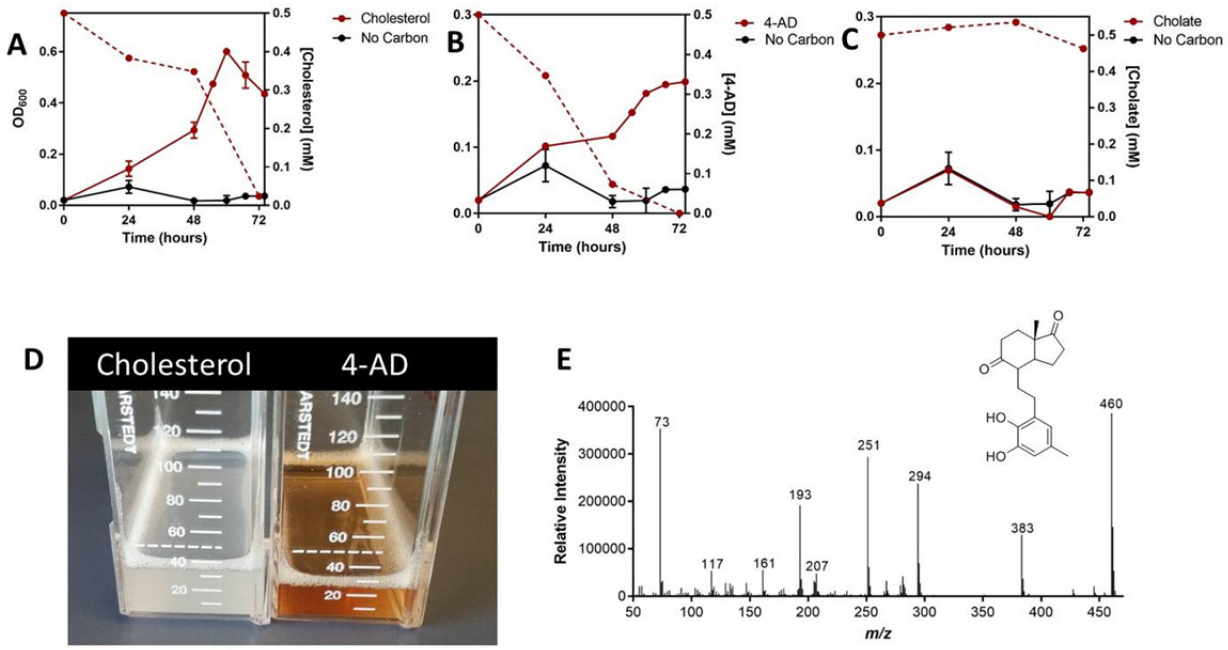


Figure 2.2 Growth of *M. abscessus* on different steroids.
M. abscessus was grown in M9 minimal medium supplemented with 0.5 mM cholesterol (**A**), 0.5 mM 4-AD (**B**), or 0.5 mM cholate (**C**). Growth was measured by optical density at 600 nm (solid lines), and substrate depletion was measured by GC-MS (dotted lines). Error bars indicate standard deviation of biological triplicates. (**D**) Colour of culture supernatants of *M. absceussus* grown on cholesterol and 4-AD. (**E**) GC-MS acquired mass spectra of DHSA identified in culture supernatant of ADD grown *M. absceussus.*

2.3.4 *ΔhsaC* and *ΔhsaD* RHA1 accumulate metabolites with incompletely degraded sidechains.

The data imply that $HsaD_{Mab}$ is involved in growth on cholesterol. However, the phylogenetic data indicate that the enzyme's substrate specificity (k_{cat}/K_M) differs from that of other HsaDs involved in cholesterol catabolism, such as $HsaD_{Mtb}$ and $HsaD_{RHA1}$. To probe these differences, metabolite accumulation in the previously described *hsaC* and *hsaD* mutants of RHA1 [25] was first investigated. RHA1 is used to understand cholesterol catabolism and its mutants are used to generate metabolites to study *Mtb* enzymes [58, 76]. When incubated with cholesterol, an *hsaC* mutant accumulated two metabolites that were detected using GC-MS and that were not detected in cultures of WT cells (Figure 2.3, red). The first was DHSA, as

previously observed for this mutant [39]. The second had a parent ion with *m*/*z* = 590 and contained the 294 *m*/*z* fragment characteristic of the catechol of DHSA (Figure 2.3, red). NMR analysis identified this metabolite as 3,4-dihydroxy-17-isopropionoyl-9,10-seconandrost-1,3,5(10)-triene-9-one (DHSBNC), a DHSA analog with a C-17 isopropionyl moiety (data not shown). Cultures of Δ*hsaD* cells incubated with cholesterol were characterized by yellow-colored supernatants (data not shown). This coloration is consistent with excretion of an MCP such as DSHA [77, 159]. GC-MS analysis identified a metabolite in these supernatants whose parent ion had *m*/*z* = 549 (Figure 2.3, blue). This value is consistent with the 2-amino derivative of DSHBNC (DSHA with a C-17 isopropionyl group). MCPs have been reported to react with ammonia forming pyridine analogs [194]. These results are consistent with the mutant excreting DSHBNC which then reacts with ammonium chloride, present at ~20 mM in M9 media, to yield 2-amino DSHBNC. No metabolites with masses consistent with DSHA or 2-amino-DSHA were detected, implying that DSHBNC is the primary metabolite that is excreted in Δ*hsaD* RHA1.

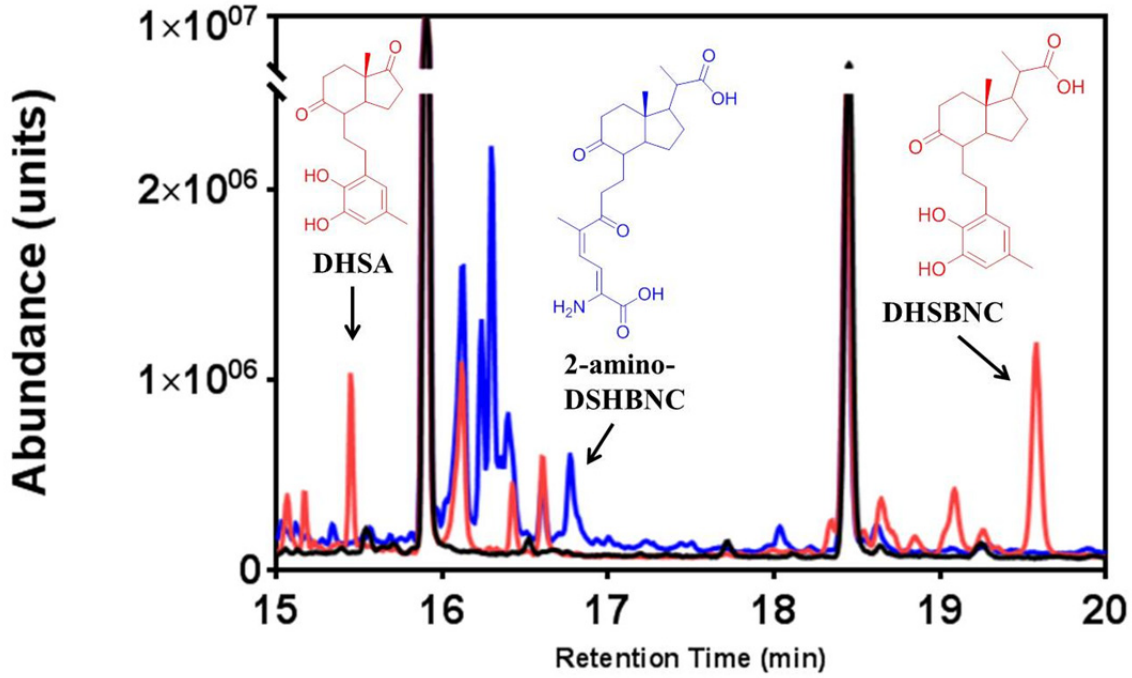


Figure 2.3 Δ*hsaC* and Δ*hsaD* RHA1 accumulate cholesterol derived metabolites with partially degraded side chains.
Shown are gas chromatograms of ethyl acetate extracted supernatants from Δ*hsaC* (red), Δ*hsaD* (blue), and WT RHA1 following 48 h of incubation in M9G containing 0.5 mM cholesterol. Structures displayed represent species confirmed using authentic standards.

2.3.5 The substrate specificities of $HsaC_{Mtb}$, $HsaD_{Mab}$ and $HsaD_{Mtb}$.

The accumulation of DHSBNC in the culture supernatant of Δ*hsaC* RHA1 indicates that HsaC may act on this substrate or a CoA thioester, thereof. DHSBNC-CoA was produced from DHSBNC using the CoA synthetase CasI. Steady state kinetic analysis of $HsaC_{Mtb}$ revealed a ~4 times higher substrate specificity for DHSBNC-CoA than either DHSA or DHSBNC (Figure 2.4, Table 2.2).

To further characterize $HsaD_{Mab}$ and $HsaD_{Mtb}$, their respective substrate specificities were compared. Both enzymes had higher specificity (k_{cat}/K_M) for a CoA thioesterified substrate, DSHBNC-CoA, than for non-thioesterified substrates, DSHA and DSHBNC (Figure 2.4, Table

2.2). However, the difference was two orders of magnitude in $HsaD_{Mtb}$ and only a single order of magnitude in $HsaD_{Mab}$. This effect can largely be attributed to the significantly lower K_M value for DSHBNC-CoA as compared DSHA or DSHBNC for either enzyme. Despite both enzyme's higher specificity for DSHBNC-CoA, $HsaD_{Mab}$ nevertheless had a reasonable k_{cat}/K_M value for DHSA, the 4-AD metabolite, ~40-fold higher than that of $HsaD_{Mtb}$. Finally, $HsaD_{Mtb}$ had remarkably similar-steady state kinetic parameters for DSHA and DSHBNC, suggesting that the C-17 isopropionyl group contributes little to the substrate specificity. In an attempt to determine the affinity of $HsaD_{Mtb}$ for the CoA moiety, competitive inhibition of the DSHBNC-CoA reaction was tested using CoASH but no significant inhibition was detected ($K_{i,c}$ > 2 mM, data not shown).

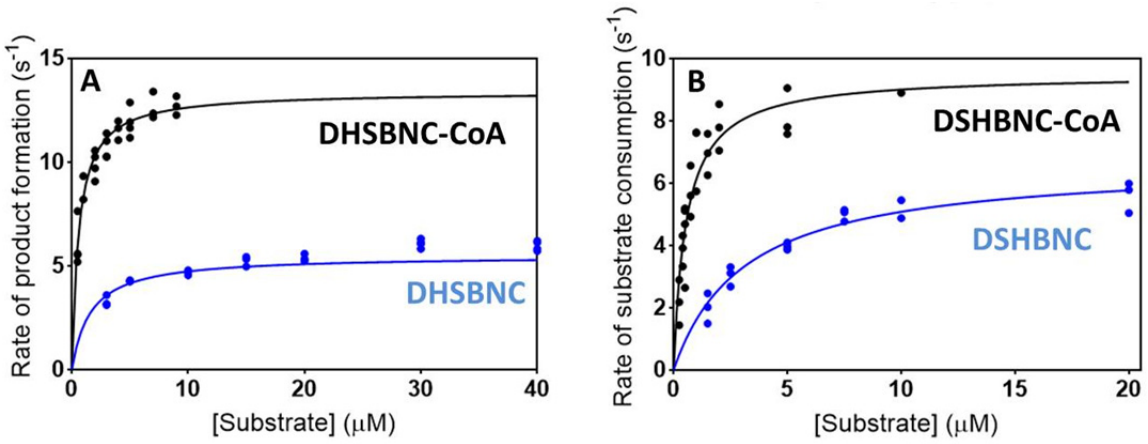


Figure 2.4 Steady-state kinetic parameters of $HsaC_{Mtb}$ and MCP hydrolases towards substrates with and without partially degraded side chains
Shown are steady state analyses of **(A)** $HsaC_{Mab}$ using DHSBNC (blue) or DHSBNC-CoA (black) and **(B)** $HsaD_{Mab}$ using DSHBNC (blue) or DSHBNC-CoA (black) as substrates. Rates reported are normalized to enzyme concentration. Reactions were performed in 200 µl potassium phosphate pH 7.5 (I = 0.1 M) at 25°C and followed spectrophotometrically at 392 nm. Curves represent best fit of the Michaelis-Menten equation.

Table 2.2 **Steady-state kinetic parameters of $HsaC_{Mtb}$, $HsaD_{Mtb}$ and $HsaD_{Mtb}$**

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat} / K_M ($x\ 10^5\ M^{-1}\ s^{-1}$)
$HsaC_{Mtb}$	DHSA	8.0 (0.2)	1.6 (0.1)	50 (4)
	DHSBNC	6.6 (0.1)	3.2 (0.3)	21 (2)
	DHSBNC-CoA	13.4 (0.3)	0.63 (0.06)	210 (30)
$HsaD_{Mtb}$	DSHA	0.93 (0.04)	34 (3)	0.27 (0.02)
	DSHBNC	0.91 (0.04)	35 (3)	0.26 (0.02)
	DSHBNC-CoA	3.4 (0.1)	0.6 (0.2)	57 (3)
$HsaD_{Mab}$	DSHA	9 (1)	8(2)	11 (1)
	DSHBNC	6.7 (0.3)	3.2 (0.4)	21 (2)
	DSHBNC-CoA	9.5 (0.5)	0.5 (0.1)	180 (20)

2.3.5.1 $HsaD_{Mtb}$ has high affinity for substrates with partially degraded side chains

Steady-state kinetic analysis established that $HsaD_{Mtb}$ has >200 times higher specificity for substrates with partially degraded side chains. This effect is largely attributed to the presence of the CoA moiety in DSHBNC-CoA (Table 2.2). We speculated that this effect is reflective of tight binding between DSHBNC-CoA and $HsaD_{Mtb}$. The affinity (K_D) of $HsaD_{Mtb}$ for substrates with a partially (DSHBNC-CoA, DSHBNC) and completely degraded (DSHA) side chain was determined by titrating solutions of substrate with the catalytically inactive variant $HsaD_{Mtb}$ S114A. The $HsaD_{Mtb}$ S114A·MCP complex has a characteristic orange color whose absorption spectrum is similar to that of the catalytic intermediate, ES^{red} [77]. Titration of $HsaD_{Mtb}$ S114A into DSHBNC-CoA resulted in the same bathochromic shift of the substrate from a λ_{max} of 392 nm to 459 nm (Figure 2.5, *inset*). $HsaD_{Mtb}$ S114A had over 500-times greater affinity for

DSHBNC-CoA than for either DSHA or DSHBNC (Figure 2.4;Table 2.3). The similar affinity of $HsaD_{Mtb}$ S114A for DSHA and DSHBNC is consistent with $HsaD_{Mtb}$'s similar specificity for these substrates. The quadratic binding equation calculated a lower A_{max} for DSHA and DSHBNC/DHSBNC-CoA, suggesting a smaller extinction coefficient of ES^{red} for the $HsaD_{Mtb}$ S114A·DSHA complex than with the other two substrates (Table 2.3). Overall, the presence of the CoA moiety could contribute and additional ~16 kJ mol^{-1} of binding energy of DSHBNC-CoA over DSHBNC which may be responsible for the higher substrate specificity of $HsaD_{Mtb}$ towards DSHBNC-CoA.

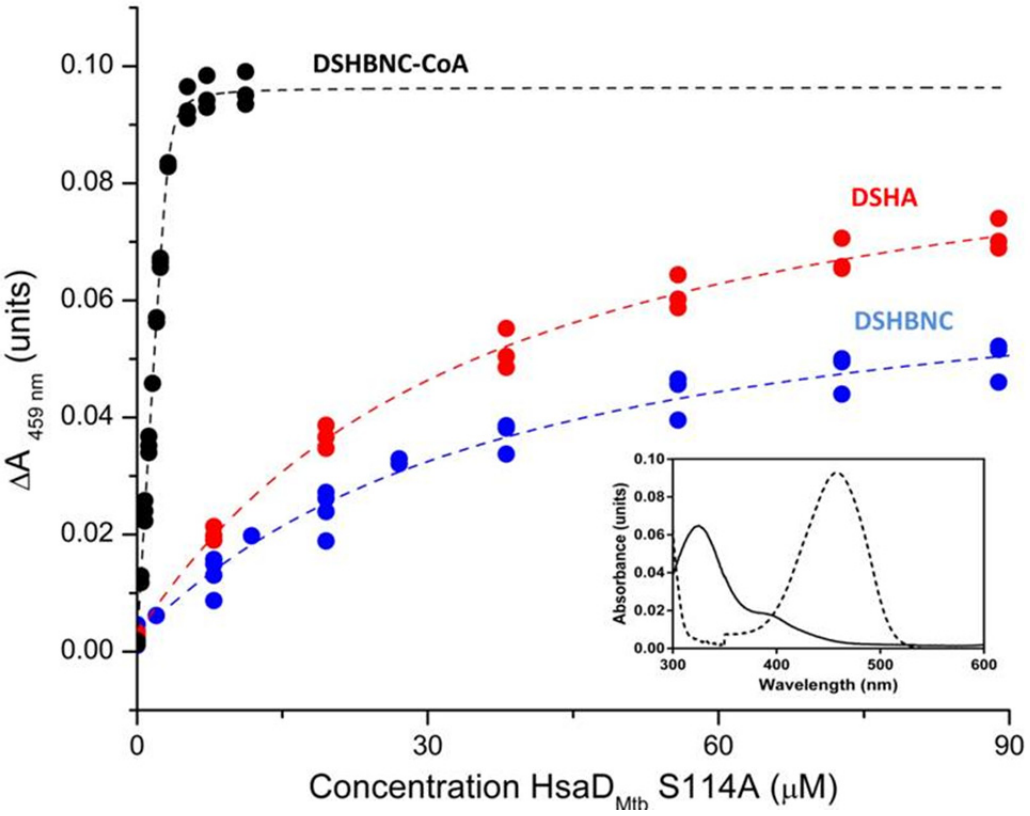


Figure 2.5 Dissociation constant (K_d) determination for $HsaD_{Mtb}$ towards DSHBNC-CoA.
Shown is the increase in absorbance at 459 nm following the titration of HsaD from *M. tuberculosis* into 5 μM DSHA (red), DSHBNC (blue) or DSHBNC-CoA (black). Spectra were recorded in 200 μl potassium phosphate pH 7.5 (I =0.1M) at 25°C. Curves represent the best fit of the quadratic binding equation to the data. Inset displays the change in spectra of DSHBNC-CoA with (dotted line) and without (solid line) HsaD from *M. tuberculosis*..

Table 2.3 Parameters for dissociation constant determination of $HsaD_{Mtb}$

Substrate	A_o (units)	A_{max} (units)	K_D (μM)
DSHA	0.002 (0.001)	0.066 (0.004)	34 (6)
DSHBNC	0.003 (0.001)	0.094 (0.003)	32 (4)
DSHBNC-CoA	0.0014 (0.0007)	0.095 (0.001)	0.06 (0.02)

Chapter 3: **Structural and functional characterization of the KstR2·HIP-CoA complex**

3.1 Introduction.

HIP, the metabolite resulting from the degradation of cholesterol's alkyl side chain and Rings A and B, is a major metabolic check point for the degradation of the final steroid fragment. Casabon *et al.* (2013) demonstrated that the degradation of HIP commences by its thioesterification to HIP-CoA. Therefore, subsequent steps in its catabolism were proposed to involve CoA thioesters. Interestingly, HIP-CoA regulates the expression of a ~14-gene regulon by acting as the effector of KstR2, the regulon's TetR family transcriptional repressor [91]. However the mechanism by which HIP-CoA acts on KstR2 is unknown.

Herein, a combination of isothermal titration calorimetry (ITC), electrophoretic mobility assays (EMSA), X-ray crystallography and directed mutagenesis was used to characterize the molecular function of KstR2 from *M. tuberculosis*, $KstR2_{Mtb}$. The data define interactions between KstR2 and HIP-CoA, and provide insights into the function of this regulator in the bacterial catabolism of steroids as well as into TFRs in general.

The contents of Chapter 3 were published in the *Journal of Biological Chemistry* in 2015 [177]. Experiments pertaining to the crystallization and structure refinement were all performed by Dr. Peter Stogios and Elena Evdokimova in the Savchenko Lab at the University of Toronto. Sections 3.2.5 and 3.3.2 were written primarily by PS. All other sections were written and edited by LE and myself.

3.2 Materials and methods

3.2.1 Chemicals and reagents

ATP, CoASH, and cholesterol (>99%) were purchased from Sigma-Aldrich. NdeI and HindIII Fast Digest restriction enzymes were purchased from Thermo Fisher Scientific Inc. T7 DNA Ligase and DpnI were purchased from New England Biolabs. Oligonucleotides were ordered from Integrated DNA Technologies. FadD3 and poly-His-tagged tobacco etch virus protease (TEV^{Pro}) were produced as previously described [79, 195]. Water for buffers was purified using a Barnstead Nanopure Diamond^{TM} system to a resistivity of at least 18 MΩ. Reagents were of HPLC or analytical grade.

3.2.1.1 Preparation of HIP and HIP-CoA

HIP was obtained using a Δ*fadD3* mutant of RHA1 as previously described [91]. HIP-CoA was produced by incubating 2 mM HIP with 2.25 mM ATP, 2.25 mM CoASH, 5 mM $MgCl_2$ and 5 μM FadD3 in 800 μl 25 mM HEPES, pH 7.5, 50 mM KCl for 30 min. HIP-CoA was purified at room temperature by high-performance liquid chromatography (HPLC) using a Luna 3 μm PFP(2) 50 × 4.6 mm column (Phenomenex) in 100 mM ammonium acetate, pH 4.5 at 1 ml min^{-1} over a 20 ml linear gradient of 0-90% methanol. HIP-CoA containing fractions were pooled and methanol was removed under nitrogen. HIP-CoA was purified to >95% purity and its identity was confirmed by ESI-MS as described [91]. HPLC purified HIP-CoA was desalted using a Strata-X 33u 30 mg solid phase extraction (SPE) column (Phenomenex). The SPE column was equilibrated with 1 ml methanol then 1 ml water. The HIP-CoA solution was passed through the column, washed with 1 ml water, and eluted in 100% methanol. HIP-CoA was dried under nitrogen and solubilized in 100 μl 25 mM HEPES, pH 7.5, 50 mM KCl. Its concentration was determined spectrophotometrically using $\varepsilon_{260\ nm} = 11.9\ mM^{-1}\ cm^{-1}$. Typical mole recoveries

ranged from 70-80%. HIP-CoA for co-crystallization was produced as described above but in a final volume of 6 ml (2.9 mg HIP). HPLC-eluted fractions containing high concentrations of HIP-CoA were desalted on the SPE column, dried under nitrogen, suspended in 250 µl water and lyophilized overnight. The residue was suspended in 50 µl water to a final concentration of 62 mM.

3.2.2 DNA manipulation

Plasmid DNA was manipulated and propagated using standard procedures [180]. Oligonucleotide-directed mutagenesis was performed using the Quickchange^{TM} PCR protocol with slight modifications. Briefly, a single 5' phosphorylated mutagenic DNA primer was annealed to pETKstR2 carrying a gene encoding poly-His tagged (Ht-)KstR2$_{Mtb}$ [91], then amplified using Phusion DNA Polymerase. T7 DNA ligase was added to the reaction mixture to form single stranded mutagenized plasmid DNA. Template DNA was removed using DpnI and the remaining ssDNA was electroporated into *Escherichia coli* NovaBlue. The R162M and W166L variants were producing using primers with the following respective nucleotide sequences: 5'-pGTCTACCGATTCATCATGGACACCACCTGGGTG-3' and 5'-pCATCCGTGACACCACCCTCGTGTCGGTGCGCTGG-3'. The nucleotide sequences of variant *kstR2* were confirmed.

3.2.3 Purification of KstR2 and variants

Wild-type and variant KstR2$_{Mtb}$ were produced using *E. coli* Rosetta 2(DE3)pLysS carrying the appropriate derivative of pETKstR2 as previously described [91]. The proteins were purified as previously described [91] with the following modification. The affinity-purified Ht-KstR2$_{Mtb}$ was dialyzed overnight against cleavage buffer (25 mM HEPES, pH 7.5, 50 mM KCl, 1 mM DTT, and 0.5 mM EDTA). The affinity tag was removed by incubating ~100 mg of Ht-KstR2$_{Mtb}$

with 0.5 mg TEV^{Pro} in 10 ml cleavage buffer overnight at 4°C. Complete digestion was confirmed by SDS-PAGE analysis. TEV^{Pro}-digested $KstR2_{Mtb}$ was loaded onto Mono-Q 10/100 HR (GE Healthcare) and eluted as previously described [91]. Proteins were exchanged into 25 mM HEPES, pH 7.5, 50 mM KCl, concentrated to ~20 mg ml^{-1} and flash frozen in liquid nitrogen as beads. Typically, 50 mg of protein were purified per 1 l culture. Protein concentrations were measured using the bicinchoninic acid (BCA) protein assay with bovine serum albumin as a standard.

3.2.4 Functional characterization of KstR2

3.2.4.1 Isothermal titration calorimetry

ITC experiments were performed using an ITC200 instrument (GE Healthcare) operated at 25°C and a stirring speed of 1000 rpm. Titrations were performed using 25 mM HEPES, pH 7.5, 50 mM KCl. $KstR2_{Mtb}$ (20 μM or 40 μM variant) was titrated with 40 × 1 μl injections of HIP-CoA (200 or 400 μM). For HIP and CoASH, 40 μM $KstR2_{Mtb}$ was titrated with 20 × 4 μl injections of 400 μM titrant. Injections of buffer into $KstR2_{Mtb}$ and variants showed no significant background heats. The data were processed by subtracting the background heats and removing outlier data points. One- and two-site models were fit using Origin 7.0. Experiments were independently repeated at least three times.

3.2.4.2 Electrophoretic mobility shift assays

A dsDNA probe of the $KstR2_{Mtb}$ operator sequence located in the intergenic region of *rv3557c* and *rv3558* was prepared by heating complementary ssDNA oligomers to 95°C and annealing at room temperature in 20 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, and 75 mM NaCl as previously described [91]. DNA probes were labelled with DIG-11-ddUTP using the second generation DIG gel shift kit from Roche according to the manufacturer's protocol. Binding

assays contained 0-2 pmol $KstR2_{Mtb}$ (WT or variant), 0.04 pmol DIG-labelled DNA probe and 0-10 nmol HIP-CoA in 20 µl 20 mM HEPES, pH 7.6, 10 mM $(NH_4)_2SO_4$, 1 mM DTT, 0.2% (w/v) Tween 20, 30 mM KCl, and 1 mM EDTA. Assays were incubated for 30 min at 37^{o}C then loaded onto 9% polyacrylamide gels containing 0.5× TBE. Gels were run for 45 min at 105 V then blotted onto positively charged Hybond-N^{+} nylon membranes (GE Healthcare). DNA was viewed using anti-DIG-alkaline phosphatase and chemi-luminescent substrate, CSPD, as described by the manufacturer (Roche). Sequences of DNA probes were 5'-GCGTACCAAGCAAGTGCTTGCTTAGGTAGC-3' and 5'-GCTACCTAAG-CAAGCACTTGCTTGGTACGC-3'.

3.2.4.3 Size exclusion chromatography

The oligomeric state of KstR2 was analyzed using size exclusion chromatography multi-angular light scattering (SEC-MALS). Twenty five µl of 80 µM $KstR2_{Mtb}$ was injected onto a HPLC 1260 Infinity LC (Agilent Technologies) coupled to a Superdex 200 5/150 column (GE Healthcare). A second sample was incubated at 37^{o}C for 30 min with 20 µM of a 24-bp DNA fragment representing the KstR2 operator from the intergenic region of *rv3549c*/*rv3550* [83]. SEC-MALS was operated at 0.2 ml min^{-1} in 25 mM HEPES, 50 mM KCl, pH 7.5. Data were collected using a miniDAWN TREOS multi-angle static light scattering device and an Optilab T-rEX refractive index detector (Wyatt Technologies). The molecular weight of complexes was determined using the ASTRA6 program (Wyatt Technologies). The nucleotide sequences of the oligonucleotides used to generate the DNA fragment were 5'-ACCTAAGCAAGC-ACTTGCTTGGTA-3' and its complement. Oligo-nucleotides were HPLC-purified by the manufacturer (Integrated DNA Technologies) and annealed as described above in 25 mM HEPES, 50 mM KCl, pH 7.5.

3.2.5 Crystallization of $KstR2_{Mtb}$:HIP-CoA

Crystals of the $KstR2_{Mtb}$·HIP-CoA complex were obtained by mixing 2 µl of 48 mg ml^{-1} protein with HIP-CoA at final concentration of 1 mM and 2 µl of reservoir solution (0.2 M ammonium sulfate, 0.1 M bis-Tris, pH 5.5 and 25% (w/v) PEG3350) using the hanging drop vapor diffusion method. Crystals appeared at room temperature and were flash frozen in liquid nitrogen after being cryoprotected with paratone oil. X-ray diffraction data was collected at 100 K using a Rigaku HomeLab system featuring Micromax-007 HF rotating copper anode fitted with a Rigaku R-AXIS IV++ image plate detector. Diffraction data was processed and reduced using the HKL-3000 software package [196]. The crystal structure was solved by molecular replacement using MrBump from the CCP4 software package [197] with the structure of $KstR2_{RHA1}$ (Ro04598) from *R. jostii* RHA1 (PDB 2IBD) as a search query. Structure refinement was carried out using Phenix.refine [198] and Coot [199]. Geometry was verified using Phenix.refine, Coot and the RCSB PDB Validation server. The final asymmetric unit (AU) contains one copy of the $KstR2_{Mtb}$ protein chain encompassing residues 3-198. The presence of one copy of HIP-CoA in the AU was verified using simulated annealing (Cartesian) omit maps using Phenix.refine with default parameters, followed by model building into residual positive F_o-F_c density and occupancy refinement of HIP-CoA

3.2.5.1 Structural analysis

The PDBePISA server was used to analyze inter-protomeric and protein-ligand interactions [200]. The DaliLite and PDBeFold servers were utilized for structure comparisons [201, 202]. Electrostatic surfaces were analyzed using Chimera [203]. Binding cavity properties were analyzed using the CASTp server [204].

3.3 Results

3.3.1 KstR2$_{Mtb}$ binds HIP-CoA with high affinity

HIP-CoA is the chemical effector of KstR2$_{Mtb}$, relieving the binding of the repressor to its operator DNA upon interaction with this molecule [91]. ITC was performed to better analyze the interaction of KstR2$_{Mtb}$ with its effector. The binding was exothermic and driven by enthalpy with an unfavorable entropic contribution (Figure 3.1,Table 3.1). The one-site equation best fit the binding isotherms. No cooperativity was detected and attempts to model the data using a two-site equation yielded poor fits. Replicate titration curves were centered at a mole ratio of 0.89 consistent with a one-to-one stoichiometry between the KstR2$_{Mtb}$ protomer and HIP-CoA. Under the experimental conditions, the K_d was 80 ± 10 nM (25 mM HEPES, pH 7.5, 50 mM KCl).

ITC was further employed to test whether KstR2$_{Mtb}$ binds either HIP or CoASH. Titrations of 400 μM HIP or CoASH into 40 μM KstR2$_{Mtb}$ (Figure 3.1 B and C, respectively) gave heats that were slightly above background. However, neither compound yielded a titration curve. Increasing the concentrations to 1 mM titrant and 100 μM KstR2$_{Mtb}$ generated a proportional increase in measured heats for both compounds and no titration of CoASH. Titrations at 1 mM HIP were unreliable due to precipitation of KstR2$_{Mtb}$ during titration.

Table 3.1 Thermodynamic parameters of KstR2$_{Mtb}$ binding HIP-CoA.

KstR2 variant	Ligand	N	K_d (μM)	ΔH (kJ mol^{-1})	ΔS (J K^{-1}mol^{-1})
WT	HIP	NB[a]	-	-	-
WT	CoASH	NB	-	-	-
WT	HIP-CoA	0.89 (0.01)	0.08 (0.01)	-69.4 (0.5)	-56
R162M	HIP-CoA	1.77 (0.02)	16 (2)	-43.0 (0.9)	-66
W166L	HIP-CoA	NB	-	-	-

[a]No binding detected.

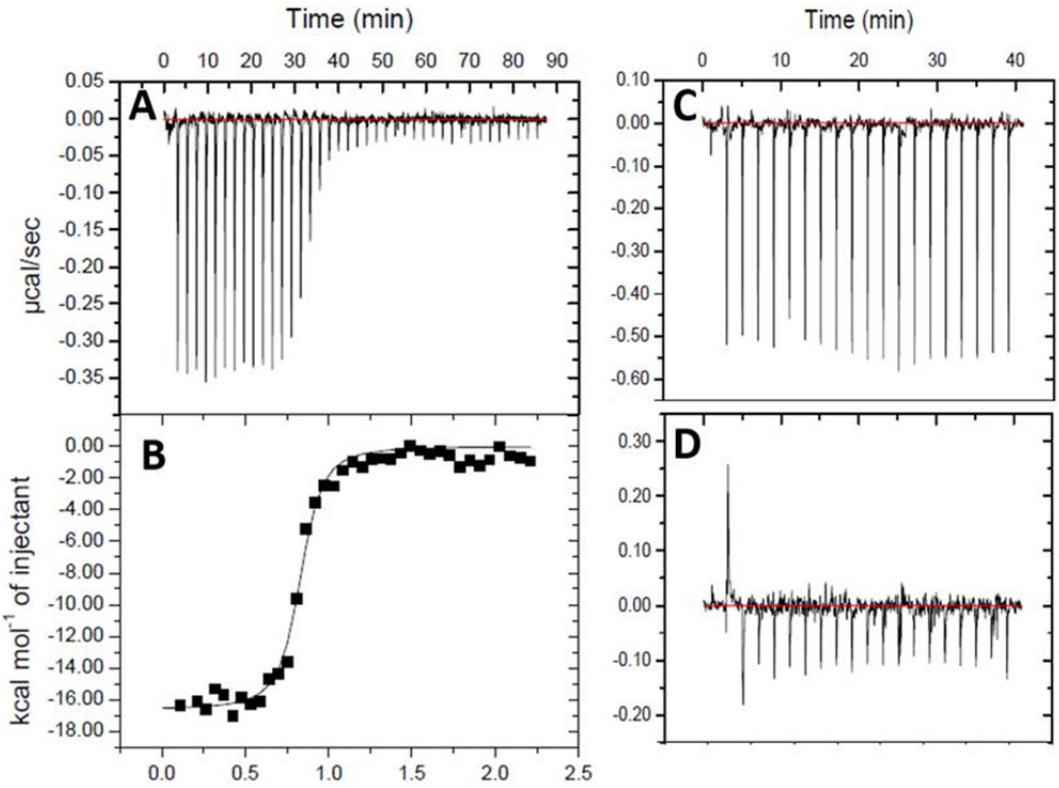


Figure 3.1 Representative isotherms of potential KstR2$_{Mtb}$ ligands.
(**A**). Titration of 20 µM KstR2$_{Mtb}$ with aliquots of 200 µM HIP-CoA. (**B**). Corresponding isotherm with the best fit single binding site model. (**C**). Titration of 40 µM KstR2$_{Mtb}$ with aliquots of 400 µM CoASH. (**D**). Titration of 40 µM KstR2$_{Mtb}$ with aliquots of 400 µM HIP. Titrations were performed using 25 mM HEPES 50 mM KCl, pH 7.5 at 25^{o}C.

3.3.2 Structure of KstR2$_{Mtb}$:HIP-CoA reveals an effector binding cleft spanning the two protomers of the dimer

To further understand the molecular function of KstR2$_{Mtb}$, KstR2$_{Mtb}$ was crystallized in the presence of HIP-CoA and the structure of the complex was solved to 1.6 Å. The asymmetric unit of the KstR2$_{Mtb}$·HIP-CoA structure contained a single protomer of KstR2$_{Mtb}$ associated with one molecule of HIP-CoA (Figure 3.2) in agreement with our ITC analysis. Structure determination statistics are presented in Appendix B.

Similarly to previously characterized TFRs, the $KstR2_{Mtb}$ protomer adopted an all α-helical L-shaped fold covering approximately 25 × 38 × 57 Å (Figure 3.2A). The short axis of the protomer comprises the N-terminal DBD (residues 6 to 54) with most α-helices in this domain arranged perpendicular to the long axis of the protein representing the C-terminal EBD (residues 55-198) (Figure 3.2A). The two domains are connected by a kinked α-helix (α6) with one face of this helix interacting with the DBD and the other with the EBD.

The $KstR2_{Mtb}$ protomer formed an extended interface of 1687 $Å^2$ with an adjacent $KstR2_{Mtb}$ protomer related to the first by a crystallographic two-fold symmetry axis. This arrangement likely represents the biological dimer, the typical minimal oligomeric state of TFRs. Dimerization of $KstR2_{Mtb}$ is mediated by contacts between 24 residues, 15 of which are hydrophobic, belonging to α-helices α8 and α9 of the C-terminal domain of each protomer.

The electron density corresponding to the HIP-CoA molecule (Figure 3.2B) occupies a large extended cavity (2637 $Å^2$ in surface area) that spans the two $KstR2_{Mtb}$ protomers and that sequesters over half of the ligand molecule from the solvent. This cavity, whose shape and chemical nature closely complement that of the ligand, is composed of two elements: a positively charged pocket lined by helices α8, α9, and their connecting loop in one protomer that binds the adenosine moiety; and a deep hydrophobic pocket defined by helices α4', α5', α6', α7' and α8' in the second protomer that binds the HIP moiety (' identifies elements of the second protomer). Thus, each HIP-CoA molecule binds across the $KstR2_{Mtb}$ dimer interface (Figure 3.2), and the two binding clefts are independent of each other. Indeed, the two ligands approach no closer than 7.6 Å.

A total of 17 amino acids from each $KstR2_{Mtb}$ protomer are located within 4.0 Å and HIP-CoA and interact with the HIP-CoA molecule. The adenine moiety of the ligand is anchored

primarily through interactions with residues belonging to the loop connecting helices α8 and α9. The diphosphates of the CoA moiety are stabilized by four hydrogen bonds, including three with Arg-162'. The cycloalkanone rings of the HIP moiety form stacking interactions with the aromatic side chains of Trp-166' (from α8') and Tyr-108' (from α6') that make up the deep hydrophobic pocket. In addition, the HIP moiety forms many hydrophobic interactions with side chains of residues that line the pocket: Phe-65', Leu-66', Leu-69', Phe-70', Tyr-73' and Val-105'. Finally, the 5-carbonyl oxygen forms a hydrogen bond with the side chain of Gln-109'. The high number of protein-ligand contacts and the close complementarity between the chemical environment of $KstR2_{Mtb}$'s binding cleft and the specific chemical groups of the ligand suggests that $KstR2_{Mtb}$ is highly specific toward HIP-CoA. This is in line with the biophysical characterization of $KstR2_{Mtb}$ interactions with this ligand presented above.

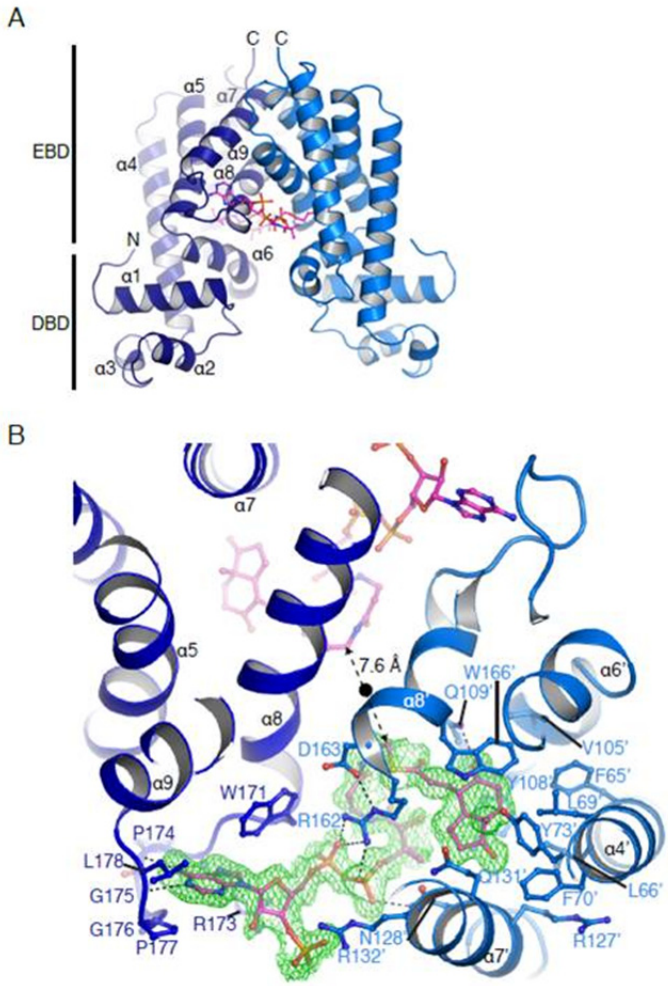


Figure 3.2 Crystal structure of $KstR2_{Mtb}$·HIP-CoA complex.
(**A**) Overall structure. Chain A from the crystal's AU is colored dark blue and chain B, produced by crystallographic symmetry, is colored light blue. HIP-CoA is shown in pink sticks. (**B**). Details of interactions between $KstR2_{Mtb}$ and HIP-CoA. The shown electron density is the F_o-F_c simulated annealing map calculated in the absence of HIP-CoA. Dashes indicate hydrogen bonds; black circle, the non-crystallographic symmetry axis; and dashed arrow, the closest approach between the dimer-bound ligands.

3.3.3 Binding of HIP-CoA alters the conformation of KstR2

Effector-binding typically triggers conformational changes in TFRs. To evaluate whether HIP-CoA binding triggers similar changes in $KstR2_{Mtb}$, structural characterization of the ligand-free form of the regulator is required. Attempts to obtain crystals of $KstR2_{Mtb}$ in the absence of HIP-CoA were unsuccessful. However, a structure of the ligand-free form of $KstR2_{RHA1}$ from *R. jostii* RHA1 is available (PDB 2IBD). $KstR2_{RHA1}$ shares 59% amino acid sequence identity with

$KstR2_{Mtb}$ including 19 of the 23 residues that interact with HIP-CoA. A superposition of the $KstR2_{Mtb}$·HIP-CoA complex and $KstR2_{RHA1}$ structures yielded an rmsd of 1.5 Å over 186 matching Cα atoms of their protomers (Figure 3.3A). Nevertheless, the ligand-binding clefts of these proteins were structurally different (Figure 3.3B), particularly surrounding the HIP moiety, suggesting that ligand binding induces the conformational differences observed in the $KstR2_{Mtb}$·HIP-CoA complex. The most striking alterations involve the positions of Trp-166 and Tyr-108, which shift up to 4.7 Å to stack on either face of bicycloalkanone rings of HIP in $KstR2_{Mtb}$·HIP-CoA (Trp-170 and Tyr-112, respectively, in $KstR2_{RHA1}$) (Figure 3.3B). The conformational differences include shifts in the positions of helices α4, α6 and α7 that surround the HIP-binding pocket and translate to regions beyond. Helix α4 connects the N- and C-terminal domains in both KstR2 structures and helix α6 forms multiple contacts with α1 of the N-terminal DBD (Figure 3.2 and 3.3B). Accordingly, the shift in position of these helices leads to a 15° outward rotation of the DBD in the ligand-bound $KstR2_{Mtb}$ in comparison to this domain's position in the ligand-free $KstR2_{RHA1}$ structure. In turn, this results in a significant difference in the relative position of the DBDs in the context of the KstR2 dimer (Figure 3.3A). These differences are consistent with HIP-CoA binding inducing a conformational change in KstR2.

To evaluate whether the conformational differences between the ligand-bound and ligand-free KstR2 structures would affect the interaction of the regulator with its operator DNA, the $KstR2_{Mtb}$·HIP-CoA structure was compared with that of a TFR bound to its operator DNA. The closest suitable match retrieved by our PDB search was SlmA from *Vibrio cholera* (PDB: 4GCT [205]). SlmA and $KstR2_{Mtb}$ protomers super-imposed with rmsd 2.8 Å over 179 matching Cα atoms (Figure 3.4A) facilitating the identification of potential DNA-binding secondary structure

elements in $KstR2_{Mtb}$. As is characteristic of TFRs, the SlmA dimer formed symmetric contacts in adjacent major grooves of its operator DNA through helices α2 (the recognition helix) and α3 in each protomer (Figure 3.4). The equivalent region in the $KstR2_{Mtb}$ structure (between residues 33 and 50) featured prominently exposed residues that could form contacts with DNA (*i.e.*, Val-33, Ser-44 and Tyr-48). Importantly, the relative position of the recognition helices differed dramatically in the operator-bound SlmA and ligand-bound $KstR2_{Mtb}$ dimers (Figure 3.4B): in $KstR2_{Mtb}$·HIP-CoA, this helix is positioned further away from the DNA major groove. Overall, these analyses suggest that the conformation of the DBD domain in $KstR2_{Mtb}$·HIP-CoA is not compatible with binding to its operator. This further suggests that HIP-CoA regulates the DNA-binding activity of KstR2 in the same manner as that established for other TFRs where effector-binding induces conformational changes that result in relieving the binding of the TFR to its operator DNA.

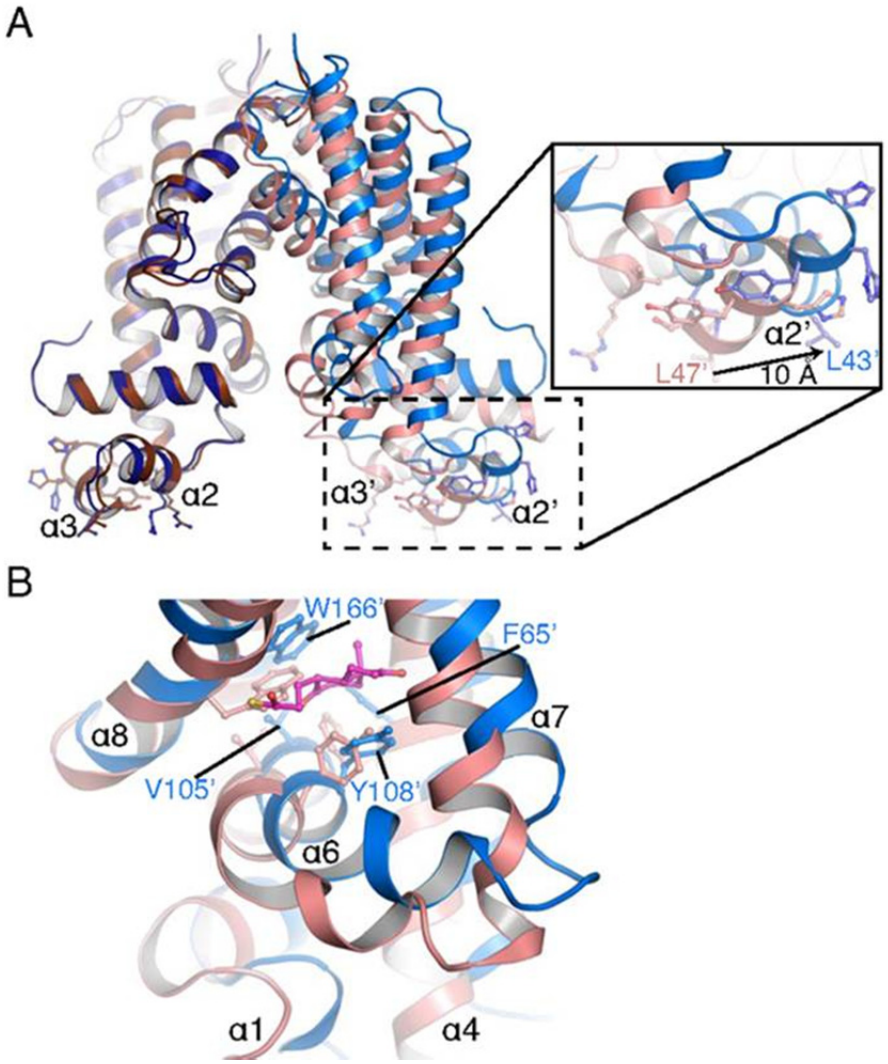


Figure 3.3 Conformational differences between $KstR2_{Mtb}$·HIP-CoA and ligand-free $KstR2_{RHA1}$
$KstR2_{Mtb}$·HIP-CoA and ligand-free $KstR2_{RHA1}$ are blue and brown, respectively. Chains A and B are dark and light shades, respectively. HIP-CoA is shown as pink sticks. (**A**). Overall comparison of the two conformations. Helices α2 and α3 are the putative DNA binding elements (helix-turn-helix motif). Inset shows a detail of the helix-turn-helix motif and the positional difference of a representative residue ($KstR2_{Mtb}$ Leu-43' / $KstR2_{RHA1}$ Leu-47'). (**B**). Localized conformational changes upon HIP-CoA binding. Single chains of the KstR2 dimer are shown for clarity. The positions of helices α1, α4, α6 and α8 differ between ligand-bound and ligand-free KstR2 structures

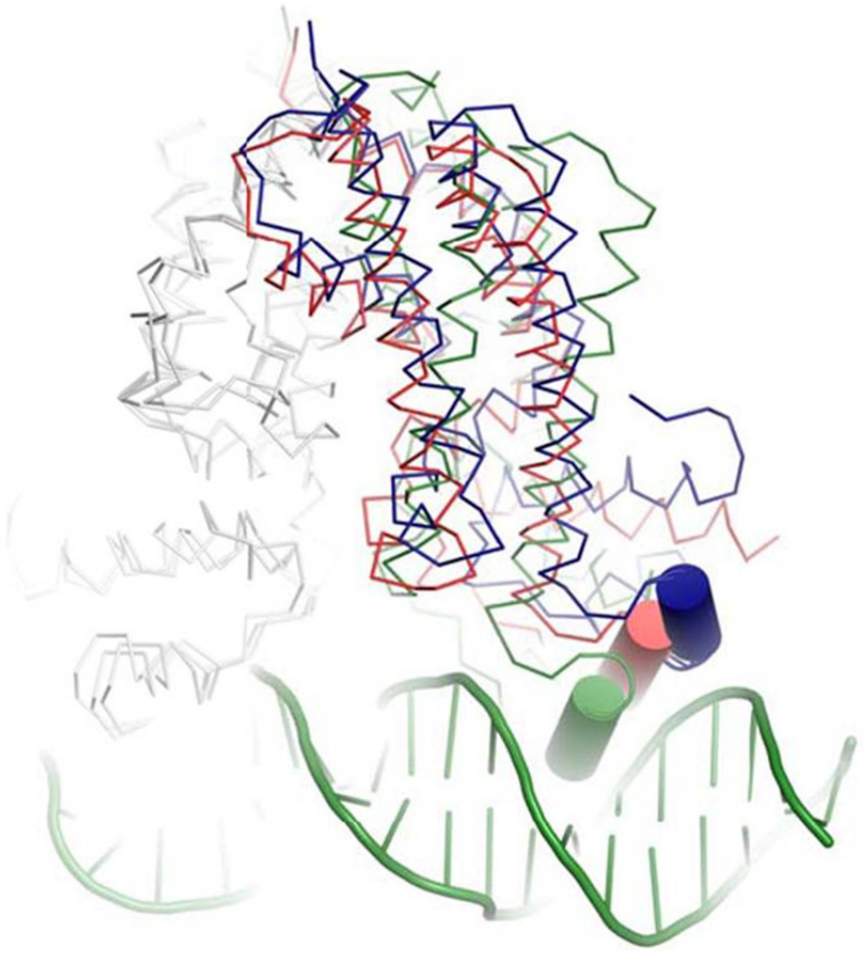


Figure 3.4 . Comparison of KstR2 structures with a TFR·DNA complex
Superposition of $KstR2_{Mtb}$·HIP-CoA (blue), $KstR2_{RHA1}$ (red) and SlmA (green, PDB 4GCT) bound to its operator DNA. Superposition performed on subunit chains colored white, which adopt the same conformation. The recognition helices (α2), which adopt different positions, are shown as cylinders.

3.3.4 Functional validation of KstR2:HIP-CoA interactions

To functionally validate the $KstR2_{Mtb}$·HIP-CoA structural model, two key HIP-CoA binding residues were individually substituted using directed mutagenesis and the resulting $KstR2_{Mtb}$ variants were characterized using ITC and EMSA. More specifically, the structural data indicate that Arg-162 and Trp-166 form important interactions with HIP-CoA (Figure 3.2B) and that their substitution with methionine and leucine, respectively, should disrupt the binding of $KstR2_{Mtb}$ to its effector but not to its operator DNA.

Isotherms showed that both KstR2 variants were significantly impaired with respect to HIP-CoA binding. The R162M variant bound HIP-CoA with an affinity ~200 times lower than WT (Table 3.1). Like WT, HIP-CoA binding to the R162M variant was enthalpically driven with an unfavorable entropic contribution (Table 3.1). Unlike WT, the isotherm of the R162M variant showed a shallow titration curve consistent with the variant not being saturated at a 3-fold molar excess of HIP-CoA (Figure 3.5A). The one-site equation fit poorly to the isotherm, yielding a stoichiometry of N=1.77. The W166L variant showed no titration with HIP-CoA: the generated heats were equal to background (Figure 3.5B).

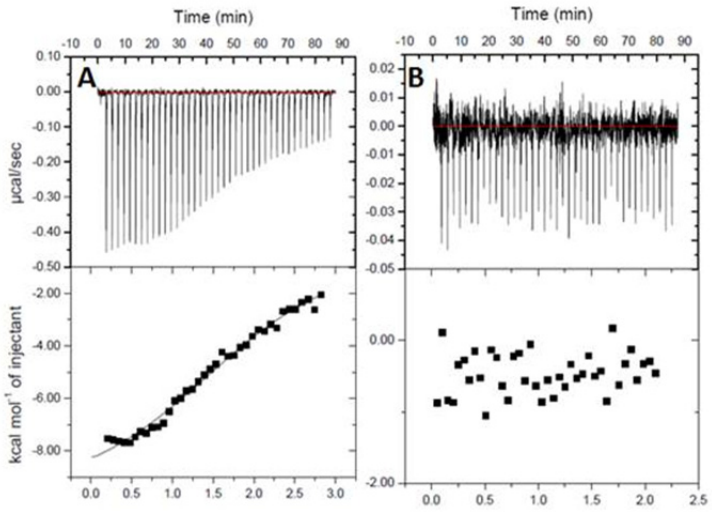


Figure 3.5 Isotherms of KstR2$_{Mtb}$ variants.
(**A**) and (**B**). show representative titrations of 400 µM R162M and W166L, respectively, with aliquots of 40 µM HIP-CoA (25 mM HEPES, 50 mM KCl, pH 7.5 at 25^{o}C). The bottom panel shows the corresponding isotherms with the best fit of the single binding site model.

Using EMSA, both the R162M and W166L variants bound to the KstR2 operator sequence with comparable affinity as WT KstR2$_{Mtb}$ (Figure 3.6). More specifically, WT and variants formed DNA: protein complexes at a mole ratio of 1:1 and no protein-free DNA probe was detected at a ratio of 1:50 DNA:protein. Consistent with previous results, the binding of DNA by WT KstR2$_{Mtb}$ was relieved in the presence of 50 µM HIP-CoA. Consistent with the ITC results,

500 μM HIP-CoA was required to detectably relieve binding to DNA by R162M. Moreover, the W166L·DNA complex was not detectably disrupted even at high concentrations of HIP-CoA.

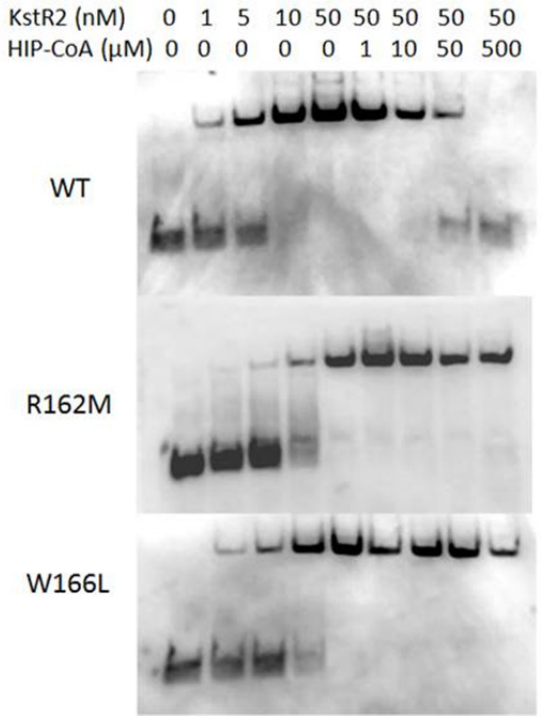


Figure 3.6 EMSA of $KstR_{Mtb}$ and variants.
Each lane contains 2 nM DIG-labelled DNA probe and the indicated amount $KstR2_{Mtb}$ and HIP-CoA. Additional experimental details are in Materials and Methods.

3.3.5 A KstR2 operator sequence binds two KstR2 dimers

The oligomeric state of $KstR2_{Mtb}$ was investigated using SEC-MALS. $KstR2_{Mtb}$ eluted as a single peak (t_R = 10.2 min; 25 mM HEPES, 50 mM KCl, pH 7.5) with a molecular weight of 42.9 ± 0.2 kDa determined using the Rayleigh ratio (Figure 3.7). This is within 10% of the predicted mass of the $KstR2_{Mtb}$ dimer (46 kDa). The small discrepancy between the two may be partly due to the elongated structure of $KstR2_{Mtb}$: the ASTRA6 software calculates molecular weight from molecular radius using spherical structures. To investigate the oligomeric state of $KstR2_{Mtb}$ bound to its operator, a DNA fragment containing a 14-bp KstR2 box flanked by 5 bp

on either side was synthesized. A sample of $KstR2_{Mtb}$ incubated with this 24-bp fragment yielded a single protein-containing peak (t_R = 8.5 min) with a molecular weight of 97 ± 1 kDa. This is within 10% of the molecular weight predicted for a complex of two $KstR2_{Mtb}$ dimers and one DNA fragment (106 kDa). The DNA fragment alone (14.7 kDa) eluted at 10 min and caused negligible light scattering (data not shown).

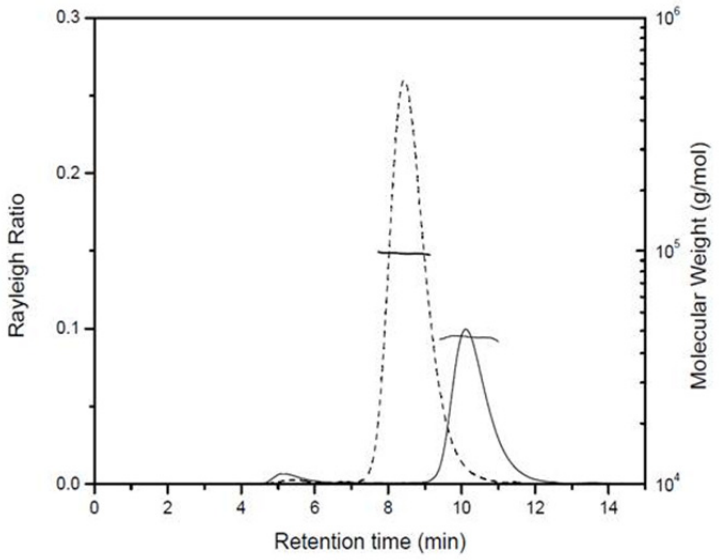


Figure 3.7 SEC-MALS of $KstR2_{Mtb}$
Rayleigh ratios of 80 µM $KstR2_{Mtb}$ with (dotted line) or without (solid line) 20 µM DNA (a 24-bp duplex containing a KstR2 box) resolved using a Superdex 200 5/150 column. Calculated molecular weights from the corresponding peaks are shown on the right axis. Results are representative of two independent experiments

Chapter 4: **Elucidation of HIP catabolism in *M. tuberculosis* and actinobacteria**

4.1 Introduction

The pathway(s) by which steroid Rings C and D are degraded in bacteria is largely unknown. Work by Casabon *et al.* (2013) and that presented in Chapter 3, strongly implicate the ~14 genes of the KstR2 regulon, in the degradation of steroid Rings C and D in Actinobacteria [83, 91, 177]. Rings C and D degradation starts with the CoA thioesterification of HIP to HIP-CoA by FadD3 [79]. Given the absence of a CoA hydrolase in the cluster and the large number of β-oxidative enzymes predicted to be encoded by the KstR2 regulon, subsequent catabolic steps are predicted to involve CoA thioesterified catabolites.

Herein, the catabolism of HIP was elucidated. Using a number of mutants in *Mtb*, *M. smegmatis*, and RHA1 unique HIP catabolites were identified by developing a LC-MS based approach to analyze intracellular CoA thioesters. Novel catabolites were confirmed using a combination of NMR and chemical synthesis. Many of the HIP catabolic steps were reconstituted *in vitro* using recombinantly expressed enzymes encoded within the KstR2 regulon. The presented data provide the first description of the bacterial catabolism of steroid Rings C and D.

The work outlined in Chapter 4 was published in *mBio* in 2017 [178]. This study was highly collaborative. A complete list of author contributions is available in the preface. In brief, much of the LC-MS data was generated by Dr. Israel Casabon with help from Jason Rogalski from the Foster Lab at UBC. Experiments relating to the growth of *Mtb* were performed by Kirstin Brown or Jie Liu. Chemical synthesis was performed by Timothy Hurst from the Snieckus Lab at Queen's University, Ontario.

4.2 Materials and methods

4.2.1 Chemicals and reagents

ATP, *p*-coumaric acid, propionic anhydride, succinic anhydride, CoASH, phorbol 12-myristate 13-acetate (PMA), sodium acetate, sodium propionate, and cholesterol (≥99%) were purchased from Sigma-Aldrich. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. ACCUZYME was purchased from Bioline. All other reagents were of HPLC or analytical grade. Buffers and solvents were prepared as previously described [79].

4.2.1.1 Preparation of steroid metabolites and CoA thioesters

HIP and HIP-CoA were prepared as described previously [79, 91]. *p*-Coumaroyl-CoA was synthesized as described using CouL [62] and was quantified spectrophotometrically in 50 mM sodium phosphate, pH 7.1 (ε_{333} = 21,000 M^{-1} cm^{-1}) [206]. Propionyl-CoA and succinyl-CoA were synthesized from their corresponding anhydride as previously described [207]. 5α-OH HIC-CoA was synthesized from 5α-OH HIC using the mixed anhydride method [48]. Total synthesis of the 5-OH HICs will be published elsewhere. Briefly, 5 mg 5αOH-HIC dissolved in 0.5 ml dry THF was reacted with 10 µl ethylchloroformate and 26 µl triethylamine (TEA) for 20 min at room temperature. The resulting mixed anhydride was filtered through glass wool into 15 mg of CoASH dissolved in 1 ml THF:water (2:3, v:v), adjusted with NaOH to ~pH 8, and reacted at 37^{o}C for 1 h. The reaction was stopped with the addition of 20 µl acetic acid and THF was removed under nitrogen. Typical mole yields were 80%.

CoA thioesters were HPLC-purified using a Luna 3 µm PFP(2) 50 × 4.6 mm column (Phenomenex) equilibrated with 0.1 M ammonium acetate, pH 4.5. CoA thioesters were eluted using a 20 min linear gradient of 0 to 90% methanol in 0.1 M ammonium acetate, pH 4.5. The

eluate was monitored at 260 nm. Methanol was removed under N_2 and compounds stored at -80°C. CoASH, acetyl-CoA, propionyl-CoA succinyl-CoA, 5α-OH HIC-CoA, and HIP-CoA were quantified at 260 nm using an ε_{260} of 11,900 M^{-1} cm^{-1} in 50 mM sodium phosphate, pH 7.1. The identities of the CoA thioesters were verified by LC/MS/MS using a Zorbax SB300-C18 150 × 0.075 mm column (Agilent Technologies) and an Agilent 6550 ToF mass spectrometer operated as described in Section 4.2.6.2.

4.2.1.1.1 Purification of COCHEA-CoA

COCHEA-CoA was obtained from Δ*ipdAB* RHA1 using a protocol similar to that described above in the *preparation of steroid metabolites* to prepare CoA thioester metabolites for MS analysis with the following modifications. Phospholipid-free CoA metabolomes prepared from 10 × 4 l of *ΔipdAB* RHA1 were pooled, dried using a SpeedVac™, suspended in 1 ml water and filtered using 0.2 μm PTFE membrane. COCHEA-CoA was purified from the CoA metabolome using a HP1100 series HPLC (Agilent Technologies) equipped with a Luna 3u PFP(2) 50 x 4.6 mm column (Phenomenex) operated at 1 ml min^{-1} and separated over a gradient of 0-60% methanol (90%) in 100 mM ammonium acetate, pH 4.5 over 12 min. COCHEA-CoA eluted at 8.4 min as a single species with a λ_{max} = 250 nm. HPLC-purified COCHEA-CoA fractions were dried using a SpeedVac™, suspended in 2 ml water and dialyzed against 2 l of water using a 100-500 Da cellulose ester dialysis membrane (Spectrum Laboratories Inc.). Desalted COCHEA-CoA was dried using a SpeedVac™, washed and dried twice in deuterated methanol, then dissolved in 450 μl deuterated water (D_2O). The final concentration of COCHEA-CoA was estimated to be 600 μM as determined using an HPLC standard curve of CoASH (ε_{260} = 11.9 mM^{-1} cm^{-1}) and using an extinction coefficient of 16.4 mM^{-1} cm^{-1} due to the additional

absorbance from a double bond [86]. Each biotransformation, using cells from 4 l of culture, yielded ~48 nmol of COCHEA-CoA. COCHEA-CoA was confirmed via LC/MS/MS prior to NMR.

4.2.1.1.2 Purification of MOODA

MOODA was purified from the supernatant of cholesterol-incubated Δ*fadE32 M. smegmatis* as follows. Four × 1 l cultures were grown to mid log (OD_{600} = 0.6) in 7H9 media + 0.5% Tween 20 + 0.2% glycerol and harvested by centrifugation (4000 × g, 20 min at 16°C). Cells were washed once using M9 salts, suspended in 200 ml M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.5 mM cholesterol, split into 2 × 100 ml, then incubated at 37°C for 24 h in 250 ml baffled flasks. Cells were harvested by centrifugation and discarded. The supernatant was collected, acidified to ~pH 2 using HCl and extracted 3× with 1:1(v:v) ethyl acetate. The organic phases were pooled, dried over anhydrous $MgSO_4$ and filtered through Whatman paper. Ethyl acetate was removed using a rotavap. The oily residue was dissolved in water and brought to pH 7 with NaOH. MOODA was purified using a Strata-X-A strong anionic exchange solid phase extraction column (Phenomenex) according to the manufacturer's protocol. GC/MS analysis indicated that MOODA was >95% homogeneous. The yield of MOODA was ~125 μg l^{-1}. For NMR characterization, ~0.5 mg MOODA was dried using a SpeedVac™, washed twice with deuterated methanol (MeOD), and dissolved in 500 μl D_2O.

4.2.2 DNA manipulation, plasmid construction, and gene deletions

DNA was propagated, amplified, digested, ligated, and transformed using standard protocols [180]. Genes were amplified using the primers and template genomic DNA listed in Appendix A. Amplicons were digested with the enzymes indicated in the descriptions of the

oligonucleotides. The nucleotide sequence of all constructs was verified prior to their use. Plasmids based on pTipQc2 and pMV361.apr were electroporated into RHA1 and *M. smegmatis*, respectively, as previously described [79, 208].

Mycobacterial genes were deleted using homologous recombination [209]. Allelic exchange substrate (AES) constructs were generated using the oligonucleotides listed in Appendix A to amplify the up- and downstream regions of the genes to be deleted and cloning them on either side of the hyg^R cassette in pYUB854. The linearized AES was electroporated into *Mtb* and *M. smegmatis* harboring pJV53. In RHA1, mutants were obtained using a SacB-based selection as previously described [210]. Briefly, pK18-derived plasmids were electroporated into *E. coli* S17.1 and then conjugated into RHA1. After the second recombination, kanamycin-sensitive/sucrose-resistant colonies were screened and confirmed using PCR.

4.2.3 Growth of bacteria

Strains and plasmids used in this study are provided in Appendix A. RHA1 strains were cultivated aerobically at 30°C on M9 mineral medium as described previously [79] containing either 1 mM cholesterol, 1.5 mM HIP, or 10 mM pyruvate. *M. smegmatis* strains were cultivated aerobically at 37°C in M9 supplemented with 2 mM $MgSO_4$ and 0.1 mM $CaCl_2$ as described previously [211]. *Mtb* strains were grown on 7H9 media supplemented with 0.5% tyloxapol and either 0.2% glycerol or 0.5 mM cholesterol as described previously [40]. *Escherichia coli* strains were grown at 37°C in Luria-Bertani (LB) broth. Hygromycin (150 μg ml^{-1} for *E. coli*, 50 μg ml^{-1} for mycobacteria), kanamycin (25 μg ml^{-1} for *E. coli*, 20 μg ml^{-1} for mycobacteria), ampicillin (100 μg ml^{-1}), chloramphenicol (34 μg ml^{-1}) and apramycin (30 μg ml^{-1}) were used for selection

where appropriate. Growth of RHA1 and *M. smegmatis* was followed using $OD_{600\ nm}$. Growth of *Mtb* was followed using $OD_{600\ nm}$ and CFU ml^{-1} by serially diluting in saline containing 0.05% tween-80 and plating on 7H10 + OADC plates.

4.2.4 Macrophage infections

THP-1 cells (American Type Culture Collection, TIB-202), were cultured in GIBCO® RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and 1 mM sodium pyruvate and were maintained between 2 and 5 x 10^5 cells ml^{-1}. *Mtb* strains were grown to late log phase in Middlebrook 7H9 supplemented with OADC, aliquots were frozen at -80°C, and CFU ml^{-1} was enumerated. THP-1 cells were seeded in 24-well flat-bottom tissue culture plates and allowed to adhere in the presence of 50 ng ml^{-1} PMA for 48 hours at 37°C in a humidified, 5% CO_2 atmosphere. Cells were washed to remove PMA and incubated for a further 48 hours prior to infection. Bacteria were added to THP-1 cells at an MOI of 1:1 for 6 hours. Cells were washed three times to remove extracellular bacteria and were then incubated for 7 days. At each time point, THP-1 cells were lysed by adding 0.06% SDS. Bacteria were then serially diluted in saline containing 0.05% tween-80, and plated on Middlebrook 7H10 with OADC for enumeration.

4.2.5 Preparation of CoA metabolomes

Cells were grown in 900 ml pyruvate or glycerol minimal medium as described above. Cells were harvested at mid-log phase, washed with fresh medium lacking growth substrate, then suspended in 100 ml growth medium supplemented with either 0.5 mM cholesterol, 20 mM pyruvate (RHA1) or 0.2% glycerol (*M. smegmatis* and *Mtb*). Biotransformations were incubated for 48 hours at 30°C/200 rpm for RHA1 and 37°C/200 rpm for *M. smegmatis* and 37°C in roller

bottles for *Mtb*. Cells were cooled on ice, harvested by centrifugation, washed once with ice-cold minimal medium, and then stored at -80°C until use.

CoA thioesters were extracted from cell pellets using a modified protocol described for eukaryotic cells [212]. Preparations were kept on ice or 4°C unless otherwise noted. Frozen cell pellets were suspended in 4 ml of acetonitrile:isopropanol (3:1 v:v) containing 15-50 nmol *p*-coumaroyl-CoA. Cells were disrupted using a FastPrep-24 bead beater (6 × 40 s at 6.5; 5 min pauses on ice between rounds). After the first three rounds, 0.1 M KH_2PO_4, pH 6.7 was added to a final concentration of ~25 mM KH_2PO_4. The supernatant was recovered by centrifugation (15,000 × *g* for 30 min), filtered through a 0.2 μm regenerated cellulose membrane (Phenomenex), and acidified with 0.25 ml glacial acetic acid per ml of extract. The acidified extract was applied to a 100 mg 2-(2-pyridyl)ethyl-functionalized silica column (Supelco, 54127-U), equilibrated using 1 ml "equilibration" solution (acetonitrile:isopropanol:water:acetic acid (9:3:4:4 v:v:v:v)) at -20°C. The resin was washed with 2 ml equilibration solution at -20°C before CoA metabolites were eluted with 2 ml methanol:0.25 M ammonium acetate, pH 7 (4:1 v:v) at -20°C. Methanol was evaporated under N_2, then the sample was flash frozen in liquid N_2 and lyophilized overnight. The lyophilized sample was suspended in 0.6 ml methanol, deposited on a Phree column (Phenomenex) to remove phospholipids and recovered by centrifugation (500 × *g* for 10 min). This sample eluate was dried under N_2, suspended in 0.2 ml methanol, and stored at -80°C. Immediately prior to LC/MS analysis (see below), the cellular extracts was diluted 10-fold in 0.1 M ammonium acetate, pH 4.5 and filtered 0.2 μm (PTFE).

4.2.6 CoA metabolite profiling by LC-MS

4.2.6.1 Targeted LC-MS

CoA thioesters were detected in cellular extracts using an Agilent 6460 Triple Quadrupole (QQQ) operated in positive ion mode and connected to a 80 x 0.25 mm Luna 3 μm PFP(2) (Phenomenex) analytical column through a 15 x 0.25 mm PFP(2) trapping column. CoA

thioesters were separated using a gradient of 100 mM ammonium acetate, pH 4.5 into 20 mM ammonium acetate, pH 4.5 in 98% methanol over 30 min, operated at 3 µl min^{-1}. A mixture of CoA thioester standards was run prior to each CoA metabolome to verify column performance and MRM sensitivity. Collision Energy Dissociation (CID) and fragmentor voltages were selected based on signal optimization using CoA thioester standards.

High resolution MS, MS^2 and MS^3 analyses of CoA thioesters were performed in positive ion mode on a Bruker Impact-II Q-ToF equipped with a 150 x 0.25 mm Luna 3 µm PFP(2) (Phenomenex) column. CoA thioesters were eluted using a gradient of 100 mM ammonium acetate in 2% methanol and 20 mM ammonium acetate in 98% methanol. The mass spectrometer was calibrated daily.

4.2.6.2 Untargeted LC-MS

To ensure that no CoA thioesters were missed using our targeted analysis method, representative CoA metabolomes were analyzed using LC/MS/MS as previously described [150]. Briefly, cellular extracts were diluted 1:49 with acetonitrile:water (3:97 v:v) supplemented with 0.1% formic acid then injected onto a Zorbax SB300-C18 150 x 0.075 mm column (Agilent Technologies) operated at 0.3 µl min^{-1} and eluted using a 10 min gradient from 3 - 97% acetonitrile. Mass spectra were recorded in positive ion mode on an Agilent 6550 Time of Flight (ToF) mass spectrophotometer using a scanned mass range of 50-1100 Da. Species were determined to be CoA thioesters based on the characteristic $[M+H]^+$-507 and 428 m/z fragments.

4.2.6.3 NMR characterization of metabolites

^{1}H-NMR, ^{1}H-^{13}C HMBC, ^{1}H-^{13}C HSQC, ^{1}H-^{1}H COSY, and ^{1}H-^{1}H TOCSY spectra were recorded at 25^{o}C using a Bruker 850 MHz NMR spectrometer. ^{1}H-NMR spectra were recorded

before and after each experiment to ensure no degradation had occurred during data collection. NMR data were analyzed using the ACD/NMR Processor v12.0 Academic Edition software.

4.2.6.4 Analysis of CoA metabolomics data

Peak integration, retention time, and signal-to-noise ratio (S/N) were calculated using MassHunter Qualitative Analysis B.06.00 (Agilent Technologies). Peaks were defined as having S/N > 3. Analysis of CoA metabolomic data was completed using the [M+H] $\Rightarrow$ -507 *m*/*z* transitions due to the higher signal intensity as compared to the [M+H] $\Rightarrow$ 428 *m*/*z* transitions, although the latter transition was confirmed for each CoA thioester characterized. CoA thioester levels were normalized to the internal standard (*p*-coumaroyl-CoA) prior to calculating their relative concentrations and proportion of the total cellular CoA pool.

4.2.7 Protein production and purification

Proteins were produced in and purified from *E. coli* Rosetta 2 pLysS ($IpdF_{Mtb}$), *E.coli* BL21 (DE3) (MBP-$IpdC_{DOC21}$) or RHA1 ($EchA20_{RHA1}$, $FadA6_{Mtb}$ and $IpdAB_{RHA1}$). Cells were grown on LB supplemented with carbenicillin, ampicillin or chloramphenicol at 50, 100 and 34 μg ml^{-1}, respectively, as appropriate. Single colonies of freshly transformed cells were used to inoculate 50 ml growth medium and incubated overnight at 37^{o}C, 200 rpm (*E. coli*) or 30^{o}C, 200 rpm (RHA1). Ten ml of overnight culture was used to inoculate 1 l fresh medium. At an OD_{600} of ~0.6, inducer was added (0.5 mM IPTG for *E. coli*; 10 μg ml^{-1} thiostrepton for RHA1) and cultures were incubated for an additional 16 h, then cells were harvested by centrifugation. For *E. coli* cultures, this 16 h incubation was done at 25^{o}C. In addition, *E. coli* cultures producing MBP-$IpdC_{DOC21}$ were supplemented with 0.1% glucose to repress intracellular expression of amylases. Cell pellets were stored at -80^{o}C until use.

To purify the various proteins, cell lysis buffers contained 2 U ml^{-1} DNaseI and one tablet protease inhibitor cocktail (Roche). *E. coli* cells were lysed using five passages through an Avestin Emulsiflex-05 homogenizer operated at 10,000 p.s.i. and RHA1 cells were lysed using an MP Biomedicals FastPrep-24 bead beater (five rounds of 40 s). Cell lysates were clarified by ultracentrifugation (40000 × g, 45 min at 4°C) then filtered through a 0.45 μm membrane. Proteins were buffer-exchanged and concentrated using a Centricon 30 K (Millipore) or an Amicon Stirred Cell (Millipore) equipped with a 30 K regenerated cellulose Ultrafiltration Membrane (Millipore), then flash frozen in liquid nitrogen. Protein preparations were evaluated using SDS PAGE.

To purify $IpdF_{Mtb}$, pellets of *E.coli* Rosetta 2:pETRv3559c were suspended in 25 ml 50 mM sodium phosphate, pH 8.0, 10% glycerol. The clarified lysate was loaded onto 3 ml Ni-Sepharose 6 Fast Flow resin (GE Healthcare) and $IpdF_{Mtb}$ was eluted using a gradient of 10-500 mM imidazole in 50 mM sodium phosphate, pH 8.0 according to the manufacturer's protocol. Fractions containing $IpdF_{Mtb}$ were pooled, dialyzed overnight against 25 mM HEPES, pH 7.5, 50 mM KCl and 10% glycerol, and concentrated to ~15 mg ml^{-1}.

To purify MBP-$IpdC_{DOC21}$, pellets of *E.coli* Rosetta2*:* pMALDOC21 were suspended in 30 ml 50 mM Tris, pH 8.0, 100 mM NaCl. The clarified lysate was loaded onto a 15 ml column of amylose resin (New England Biolabs). The resin was washed with ~75 ml 50 mM Tris, pH 8.0, 100 mM NaCl (Buffer A), then MBP-$IpdC_{DOC21}$ was eluted using ~75 ml Buffer A containing 20 mM maltose. Fractions containing >70% MBP-$IpdC_{DOC21}$ were pooled, concentrated to ~3 ml and exchanged into Buffer A. MBP-$IpdC_{DOC21}$ was precipitated using a final concentration of 1.75 M ammonium sulfate in Buffer A. The white precipitate was collected

via centrifugation (4000 × g, 5 min, 4°C), washed with fresh Buffer A containing 1.75 M ammonium sulfate and solubilized in 1 ml Buffer A.

To purify $EchA20_{RHA1}$, $FadA6_{Mtb}$, and $IpdAB_{RHA1}$, pellets RHA1: pTipR1EchA20, RHA1:pTipFadA6, RHA1:pTipR1IpdAB, respectively, were suspended in 25 ml 50 mM sodium phosphate, pH 8.0 containing 300 mM NaCl and 10 mM imidazole. Clarified lysates were loaded onto 2 ml Ni-Sepharose G Fast Flow resin (GE Healthcare) and eluted using a gradient of 10-500 mM imidazole according to the manufacturer`s protocol. Fractions containing the desired protein were pooled and dialyzed overnight against either 25 mM HEPES, pH 7.5 and 50 mM NaCl ($FadA6_{Mtb}$ and $IpdAB_{RHA1}$) or 25 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM $MgCl_2$, 1 mM $NaHCO_3$ and 10% glycerol ($EchA20_{RHA1}$). Proteins were concentrated to ~10 mg ml^{-1}.

4.2.8 Enzymatic transformations

Assays using purified enzymes were performed in a final volume of 100 µl containing 10 mM sodium phosphate, pH 8.0, 100 µM 5α-OH HIC-CoA, 100 µM NAD^+, 5 µM FMN and 2 µM of each relevant enzyme. Assays containing $FadA6_{Mtb}$ also contained 50 µM CoASH. Reactions were incubated at 37°C for 1 h. Proteins were removed before HPLC and MS analysis by the addition of 200 µl acetonitrile + 5% acetic acid. Volatile solvents were removed using a SpeedVac™ concentrator and precipitated proteins were pelleted by centrifugation (16,000 × g, 5 min, 4°C). Samples were filtered through a 0.2 µm membrane, diluted 2:1 in water and run on an HP1100 series HPLC equipped with a Luna 3u PFP(2) column. The eluate was monitored at 260 nm. The identity of new peaks was confirmed by LC/MS/MS as described above.

4.2.9 Bioinformatic analyses

The amino acid sequences of Rings C/D catabolic enzymes were obtained from the NCBI using their *Mtb* H37Rv gene loci. These sequences were used to search for homologs in the genomes of RHA1, *M. smegmatis* MC^2155, *C. testosteroni* CNB-2, and *S. denitrificans* DSM 18526 using BLAST-P. Best hits were used to search the *Mtb* H37Rv genome to evaluate whether they were reciprocal best hits. Closest characterized homologs were determined by using the amino acid sequences of *Mtb* Rings C/D catabolic enzymes to search against the Protein Data Bank database using BLAST-P and manually identifying the best characterized result for each enzyme.

4.3 Results

4.3.1 The *ipdABC* genes are required for growth on cholesterol and HIP

The KstR2 regulon has been strongly implicated in the catabolism of HIP in mycolic acid-containing Actinobacteria [91]. Because FadD3, encoded by the KstR2 regulon, initiates HIP catabolism, we hypothesized that the other regulon-encoded enzymes act downstream of FadD3. Deletion mutants of *ipdAB* and *ipdC* in each of *Mtb* (*Rv3551-Rv3553*) and RHA1 (RHA1_*RS22695-22685*) were initially focused on to gain insights into HIP catabolism.

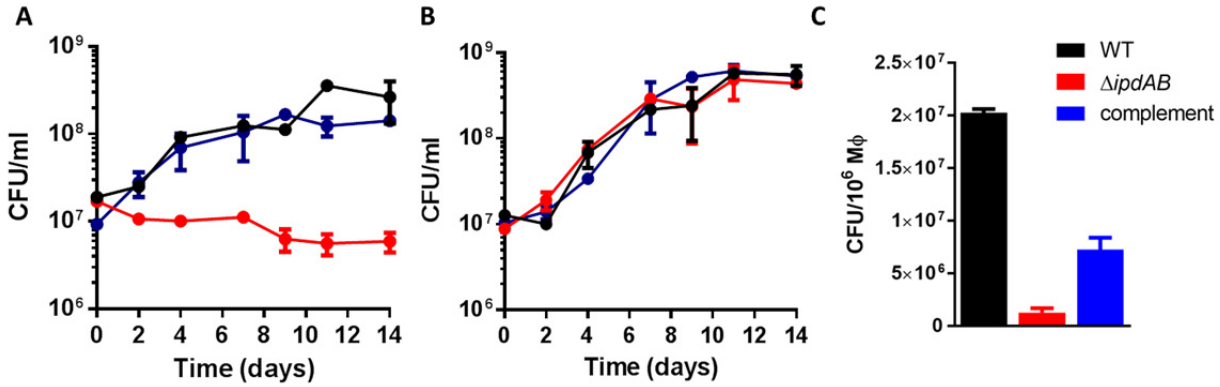


Figure 4.1 Growth of Δ*ipdAB Mtb*.

WT *Mtb Erdman* (black), Δ*ipdAB Mtb* (red), or Δ*ipdAB Mtb::ipdAB* (blue) were grown on (A) 0.5 mM cholesterol, (B) 0.2% glycerol, or (C) in PMA-differentiated THP-1 cells. Data represent the mean of biological triplicates

An Δ*ipdAB* mutant constructed in *Mtb* Erdman did not grow on cholesterol (Figure 4.1A), but grew as the wild-type on glycerol (Figure 4.1B). The growth defect on cholesterol was restored through complementation. Because mutations in *Mtb* can exhibit unexplained strain differences [213] an equivalent Δ*ipdAB* mutant constructed in *Mtb* CDC1551 was verified to have the same phenotype as Δ*ipdAB Mtb* Erdman (data not shown). Finally, an Δ*ipdAB* mutant of RHA1 exhibited a similar phenotype: it did not grow on either HIP or cholesterol, and grew normally on pyruvate (Figure 4.2). The growth defects on cholesterol and HIP were restored through complementation with *ipdAB* of *Mtb*.

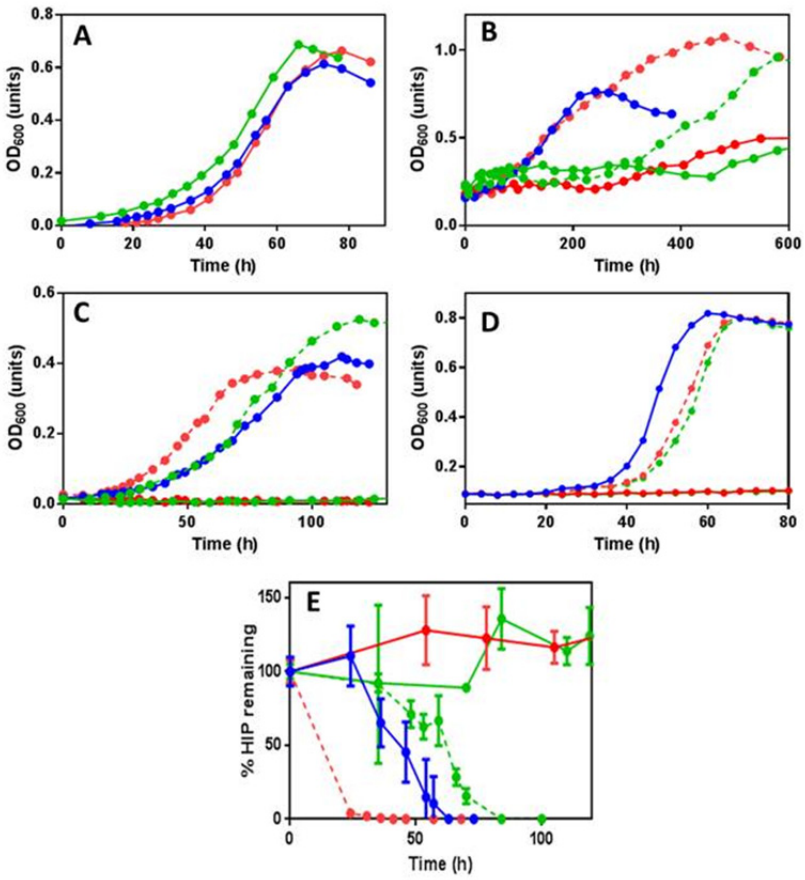


Figure 4.2 Growth and CoA metabolites of RHA1 strains
Growth of WT::pTip-Qc2 (blue), Δ*ipdAB* ::pTip-Qc2 (red), Δ*ipdAB*:: pTipCoL51 (red, dashed), Δ*ipdC*::pTipQc2 (green) and Δ*ipdC*:: pTipRv3553 (green, dashed) on: (A) 10 mM pyruvate; (B) 1 mM cholesterol; (C) 1.5 mM HIP; and (D) 1 mM HIP plus 10 mM pyruvate. (E) Depletion of HIP by RHA1 strains, color-coded as in growth curves as measured by GC/MS and reported as % of initial levels. Data are the mean of triplicates. Error bars show standard deviation. Data for panel D were acquired using a BioScreen C (Growth Curves USA)

An Δ*ipdC* mutant of RHA1 also did not grow on either cholesterol or HIP but grew normally on pyruvate (Figure 4.2). The growth defect on cholesterol and HIP was restored through complementation with $ipdC_{Mtb}$. Gas chromatography-coupled mass spectrometry (GC/MS) established that while cholesterol and HIP were depleted in the wild-type and complemented strains, they were not detectably depleted by the *ipdC* mutant (Figure 4.2E). Similar results were obtained in *Mtb* Erdman: the Δ*ipdC* mutant did not grow on cholesterol but

grew normally on glycerol (Figure 4.3). Moreover, this phenotype was restored through complementation with an integrative plasmid harboring *ipdC*.

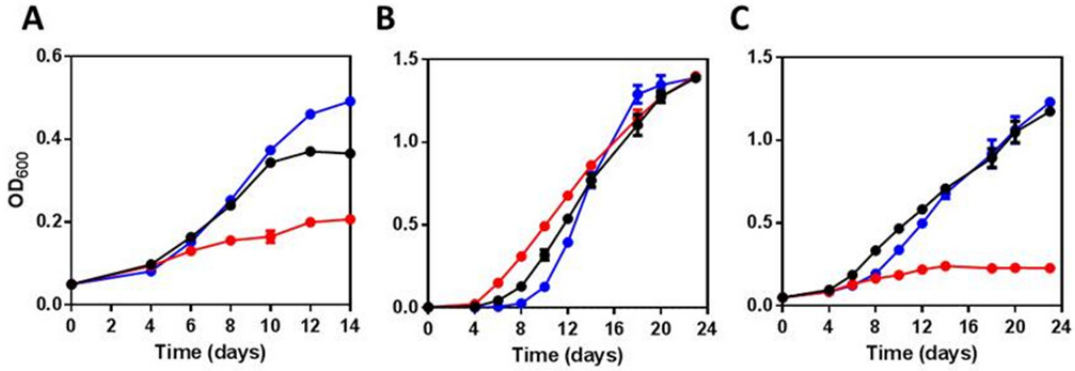


Figure 4.3 Growth and CoA metabolites of *ΔipdC Mtb.*
WT (black), *ΔipdC* (red), *ΔipdC::ipdC* (blue) *Mtb* Erdman were grown on (A) 1 mM cholesterol; (B) 0.2% glycerol; or (C) 0.5 mM cholesterol and 0.2% glycerol. (D) CoA metabolome of *ΔipdC* (red) and WT *Mtb* (blue) incubated with 0.5 mM cholesterol. Arrows indicate the peaks corresponding to the 5-OH HIC-CoA in the *ΔipdC* RHA1 CoA metabolome.

4.3.2 Growth in macrophages

Transposon mutagenesis studies suggest that *ipdA* is essential for *Mtb* survival in macrophages [52]. Moreover, the gene is essential for survival of *R. equi* in foals [23]. Therefore the growth of Δ*ipdAB Mtb* in PMA-differentiated THP-1 cells was tested (Figure 4.1C). WT Mtb increased >350-fold over 7 days, corresponding to a doubling time of 19.6 hours. The mutant increased ~10-fold over this time, corresponding to a doubling time of 46.5 hours, while complementation restored intracellular replication to 131-fold. These results are consistent with *Mtb* catabolizing cholesterol during intracellular growth [134].

4.3.3 The accumulation of cholesterol catabolites in the *ipd* mutants

In an attempt to identify the respective substrates of IpdAB and IpdC, the occurrence of metabolites in the RHA1 mutants was investigated. GC/MS analyses revealed that, when incubated with cholesterol, Δ*ipdC* RHA1 accumulated low amounts of a metabolite with an m/z

of 356 (Figure 4.4A). No metabolites were detected in the culture supernatant when cells of Δ*ipdAB* RHA1 were incubated with cholesterol (Figure 4.4A).

Failure to detect significant amounts of extracellular metabolites in the supernatants of cholesterol-incubated Δ*ipdC* and Δ*ipdAB* mutants was hypothesized to be due to the accumulation of intracellular, CoA thioesterified metabolites that are not readily excreted. To test this hypothesis, CoA thioesters were extracted from cells and analyzed them using LC/MS. LC was performed using a pentafluorophenyl (PFP) resin to maximize resolution of the CoA thioesters. *Mtb*, RHA1 and *M. smegmatis* cells incubated with various substrates contained CoASH, acetyl-CoA, propionyl-CoA, and/or succinyl-CoA irrespective of the growth substrate (Table 4.1). The identity of these metabolites was based on their *m/z*-values and their retention time (R_t) on the PFP column with respect to synthetic standards. Their concentrations were quantified relative to *p*-coumaroyl-CoA, the internal standard. Some strains also contained low amounts of dephospho-CoASH depending on the substrate, as has been reported in other cells [150]. When incubated in the presence of cholesterol, the *ipd* mutants accumulated CoA thioesters that were not detected either in the wild-type strains or in *ipd* mutants incubated with glycerol or pyruvate (data not shown). More specifically, cholesterol-incubated cells of Δ*ipdC Mtb* and RHA1 contained significant amounts of two CoA thioesters with *m*/*z* values of 962, one of which was more abundant than the other (Figure 4.4B, Table 4.1). The main CoA thioester that accumulated in cholesterol-incubated cells of Δ*ipdAB Mtb* and RHA1 eluted with a R_t of 22.7 min and had an *m*/*z* value of 976 (Figure 4.4C, Table 4.1**).**

Table 4.1 Characterization of CoA metabolites found in this study

Compound	Full Name	M+H (QQQ)[a]	MRM Transition	M+H (Bruker)[b]	Molecular Formula	Retention time (min)	NMR Confirmed	Reference	Major peak observed in CoA metabolome[c]	*in vitro* enzymatic production[d]
CoASH	Coenzyme-A	768.1	768=> 261,428	768.1210	$C_{21}H_{37}N_7O_{16}P_3S^+$	17.0-18.4	Yes	SIGMA	All	
Dephos-CoASH	Dephospho-CoASH	688.1	688=>348	688.1562	$C_{21}H_{36}N_7O_{13}P_2S^+$	19.3	No		All	
Acetyl-CoA		810.1	810=> 303,428	810.1315	$C_{23}H_{39}N_7O_{17}P_3S^+$	19.9-20.9	yes	SIGMA	All	COCHEA-CoA + $IpdAB_{RHA1}$ + $FadA6_{Mtb}$
Propionyl-CoA		824.1	824=> 317,428	824.1466	$C_{24}H_{41}N_7O_{17}P_3S^+$	21.6-22.9	Yes		All	
Succinyl-CoA		868.1	868=> 361,428	868.1400	$C_{25}H_{41}N_7O_{19}P_3S^+$	21.6-22.2	Yes		All	
Unknown	Unknown	992.2	992=> 485,428	992.1901	$C_{32}H_{49}N_7O_{21}P_3S^+$	20.8-22.0	No		RHA1 *ΔipdAB* *Mtb ΔipdAB Δ* *MC²155 ΔipdAB*	
5β-OH HIC-CoA	3aβ-*H*-4*α*(Carboxylic acid)-5*α*-hydroxy-7a*β*-methylhexahydro-1-indanone	962.1	962=> 455,428	962.2159	$C_{32}H_{51}N_7O_{19}P_3S^+$	22.0-22.9	Yes (5βOH-HIC)	This Study	RHA1 *ΔipdC,* *ΔipdABC Mtb* *ΔipdC*	
HIEC-CoA	(7aS)-7a-Methyl-1,5-dioxo-2,3,5,6,7,7a-hexahydro-1H-indene-4-carboxyl-CoA	958.1	958=> 451,428	958.1899	$C_{32}H_{47}N_7O_{19}P_3S^+$	23.2-23.4	No	This Study	*MC²155 ΔechA20*	5α-OH HIC-CoA + $IpdC_{DOC21}$+$IpdF_{Mtb}$

Compound	Full Name	M+H (QQQ)[a]	MRM Transition	M+H (Bruker)[b]	Molecular Formula	Retention time (min)	NMR Confirmed	Reference	Major peak observed in CoA metabolome[c]	*in vitro* enzymatic production[d]
COCHEA-CoA	(3R)-2-(2-Carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA	976.2	976=> 469,428	976.1920	$C_{32}H_{49}N_7O_{20}P_3S^+$	22.7-23.4	Yes	This Study	RHA1 Δ*ipdAB* *MC²155* Δ*ipdAB* *Mtb* Δ*ipdAB*	HIEC-CoA + $EchA20_{RHA1}$
MOODA-CoA	4-Methyl-5-oxooctanedioyl-CoA	952.2	952=> 445,428	952.1960	$C_{30}H_{49}N_7O_{20}P_3S^+$	23.0-23.2	Yes (MOODA)	This Study	*MC²155 ΔfadE32* (MOODA in Sup.)	COCHEA-CoA + $IpdAB_{RHA1}$ + $FadA6_{Mtb}$
5α-OH HIC-CoA	3a*α-H-*4*α*(Carboxylic acid)-5*α*-hydroxy-7a*β*-methylhexahydro-1-indanone	962.1	962=> 455,428	962.2159	$C_{32}H_{51}N_7O_{19}P_3S^+$	23.0-23.8	Yes (5αOH-HIC)	This Study	RHA1 *ΔipdC, ΔipdABC Mtb ΔipdC* *MC²155 ΔechA20, ΔipdF*	
3'oxo-5OH-HIP-CoA		1004.2	1004=> 453,428	1004.2265	$C_{34}H_{53}N_7O_{20}P_3S^+$	23.3-24.1	No	This Study		
3',5-diOH-HIP-CoA		1006.2	1006=> 499,428	1006.2420	$C_{34}H_{55}N_7O_{20}P_3S^+$	24.0-24.2	No	This Study		

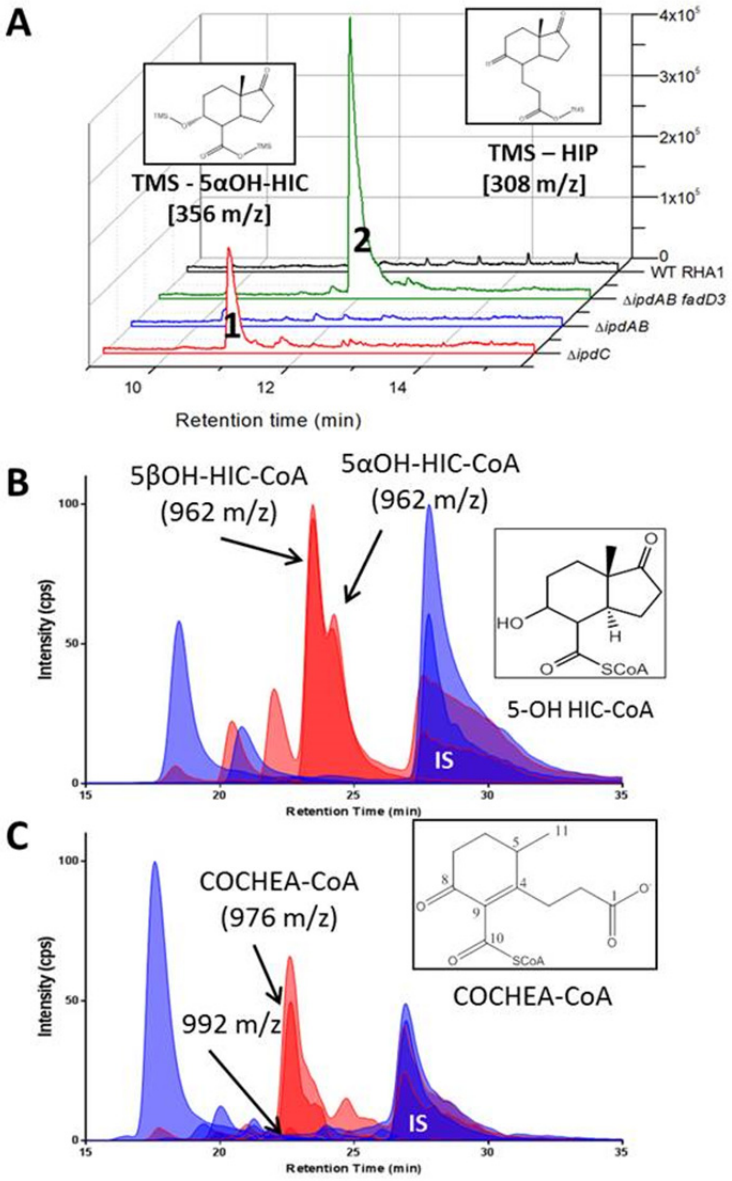


Figure 4.4 Accumulation of cholesterol derived metabolites from Δ*ipdAB* and Δ*ipdC* strains.
(A) GC/MS traces of culture supernatants of Δ*ipdC,* Δ*ipdAB,* Δ*fadD3ipdAB* and WT RHA1 incubated with cholesterol. Peaks 1 and 2 correspond to TMS-5α-OH HIC and TMS-HIP, respectively. CoA metabolome of cholesterol-incubated cells of (B) WT (blue) and Δ*ipdC* (red) RHA1 or (C) WT (blue) and Δ*ipdAB* (red) *Mtb*. The major unique peaks in the Δ*ipdC* and Δ*ipdAB* metabolomes correspond to 5αOH-HIC-CoA and COCHEA-CoA, respectively (inset). Lighter shaded curves in panels B and C are based on the 768=>261 transition observed in free CoASH as well as CoA thioesters subjected to in-source fragmentation.

4.3.4 Identification of CoA metabolites in the *ipd* mutants

The 962 *m/z* metabolite that accumulated in *ΔipdC* RHA1 and *Mtb* was predicted to be produced by the β-oxidative cleavage of acetyl-CoA from HIP-CoA and reduction of the 5-oxo group to yield 3aα-*H*-4α(3'-carboxyl-CoA)-5-hydroxy-7aβ-methylhexahydro-1-indanone (5-OH HIC-CoA; Figure 4.4B). To identify the metabolites that accumulated in the Δ*ipdC* mutants, the 5α and 5β isomers of 5-OH HIC were synthesized and confirmed by NMR (Appendix C). Each isomer was thioesterified to yield 5α- and 5β-OH HIC-CoA, respectively, which were then purified by HPLC. The high resolution $[M+H]^+$ *m/z* value and MS^3 fragmentation pattern (of the $[M+H]^+$-507 fragment), of synthetic 5-OH HIC-CoA corresponded to those of the most abundant CoA thioesters in cholesterol-incubated Δ*ipdC* mutants. The presence of both α and β diastereomers was confirmed by the R_t values of their corresponding standards. Interestingly, the *m/z* value of the metabolite that accumulated in the supernatant of cholesterol-incubated Δ*ipdC* RHA1 corresponds to that of 5-OH HIC (Figure 4.4A).

The most abundant CoA thioester in cholesterol-incubated Δ*ipdAB* mutants accumulated in sufficient quantities to allow its isolation from the RHA1 mutant for further characterization. The metabolite was identified as 2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA (COCHEA-CoA) based on high resolution mass spectrometry (976.1960 m/z), ^{1}H, COSY, TOCSY, HMBC, and HSQC NMR (Appendix C). Most diagnostically, the carbon-11 (C-11, see Figure 4.4C for numbering of carbons) methyl protons appear as a doublet, establishing that C-5 bears a hydrogen and that Ring D is open. C-10 is thioesterified based on C-1 having the same ^{13}C NMR chemical shift as in 2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1 [214].

4.3.5 Enzymatic transformation of 5-OH-HIC-CoA

To further elucidate the catabolism of HIP, various KstR2-encoded enzymes were purified, including IpdAB, IpdC, IpdF, EchA20 and FadA6. These preparations were used to evaluate their abilities to transform 5-OH HIC-CoA *in vitro*. I initially sought to work exclusively with the *Mtb* homologs. However, I was unable to obtain all of the *Mtb* homologs in stable, soluble forms despite testing various host strains and expression conditions. In some cases, the RHA1 homolog was more stable. For IpdC, I obtained the best preparations using the homolog from *P. putida* DOC21, a bile acid-degrading bacterium. The gene encoding $IpdC_{DOC21}$, DC0014-19, is the reciprocal best hit of $ipdC_{Mtb}$ and occurs in a predicted operon similar in structure to that of *S. denitrificans* which contains *ipdAB*. SDS-PAGE analyses of the various protein preparations are provided in Figure 4.5. The subscripts Mtb, RHA1 and DOC21 identify the parent strain of each enzyme. Transformation experiments were performed by incubating synthetic 5-OH HIC-CoA with various enzymes and characterizing the reaction products using LC/MS.

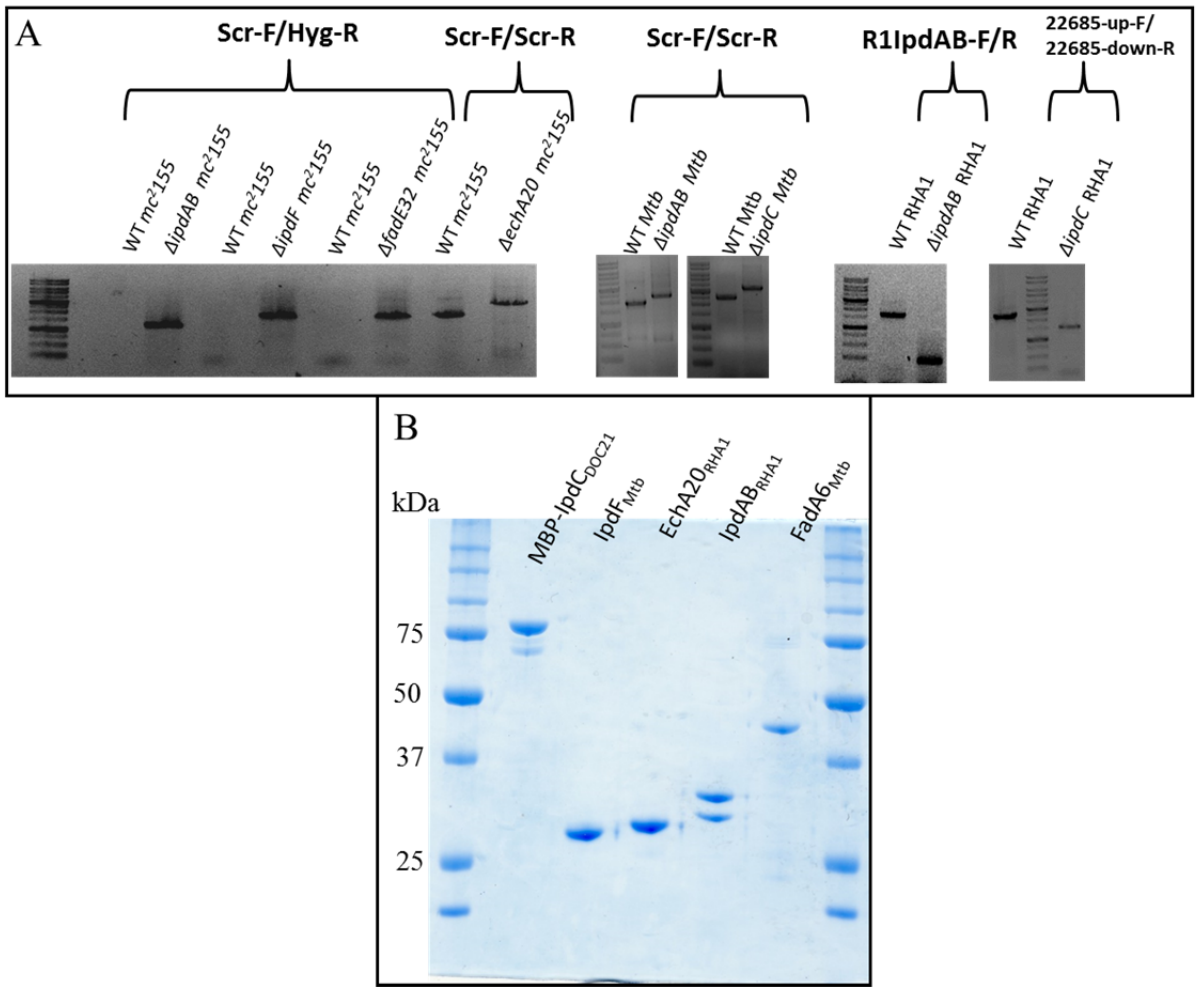


Figure 4.5 Electrophoretic analyses of gene deletion mutants and purified proteins.
(A) PCR confirmation of Δ*ipdAB* in RHA1 and Mtb; Δ*ipdC* in RHA1 and Mtb; Δ*ipdF*, Δ*echA20*, and Δ*fadE32* in *M. smegmatis* using the listed primer sets (Appendix A). Thick bands in DNA ladder depict 3 kb (top) and 1 kb (bottom). Gene deletions in mycobacteria introduced a ~1.5 kb *hyg*^{R} marker whereas RHA1 gene deletions are markerless (Section 4.2.2). (B) SDS PAGE loaded, from left to right, with 0.5 µg each of, MBP-IpdC$_{DOC21}$, IpdF$_{Mtb}$, EchA20$_{RHA1}$, IpdAB$_{RHA1}$, and FadA6$_{Mtb}$. Purified proteins are flanked by molecular weight standards.

Incubation of 5α-OH HIC-CoA with IpdF$_{Mtb}$ and IpdC$_{DOC21}$ yielded a compound with *m*/*z* value of 958 (Figure 4.6A and B; red trace), consistent with the oxidation of the 5-OH group of HIC and the introduction of a double bond. Interestingly, the two enzymes did not detectably transform 5β-OH HIC-CoA (data not shown). Based on its mass and the structure of the downstream metabolite, COCHEA-CoA, the IpdC/IpdF transformation product was

provisionally identified as (7aS)-7a-methyl-1,5-dioxo-2,3,5,6,7,7a-hexahydro-1*H*-indene-4-carboxyl-CoA (HIEC-CoA). This assignment is consistent with the function of the next step of the pathway, as discussed below. However, NMR data are required for a definitive identification. Neither $IpdF_{Mtb}$ nor $IpdC_{DOC21}$ transformed 5-OH HIC-CoA in the absence of the other enzyme. Moreover, as described below, Δ*ipdF M. smegmatis* accumulated the same major metabolite as the Δ*ipdC* mutants, 5α-OH HIC-CoA. Therefore, I was unable to determine the order of reaction of IpdF and IpdC. However, incubation of 5-OH HIP-CoA with $IpdF_{Mtb}$ yielded HIP-CoA, demonstrating that this enzyme catalyzes oxidation of the 5-OH.

Incubation of 5-OH HIC-CoA with $IpdF_{Mtb}$, $IpdC_{DOC21}$ and $EchA20_{RHA1}$ yielded a compound whose R_t and *m*/*z* values were identical to those of COCHEA-CoA (Figure 4.6C; green trace), the major metabolite that accumulated in the Δ*ipdAB* mutants.

Finally, incubation of 5-OH HIC-CoA with $IpdF_{Mtb}$, $IpdC_{DOC21}$, $EchA20_{RHA1}$, $IpdAB_{RHA1}$ and $FadA6_{Mtb}$ yielded a compound with *m*/*z* value of 952 (Figure 4.6D; blue trace). Hydrolysis of the CoA thioester yielded a compound that GC/MS revealed to be 4-methyl-5-oxo-octanedioc acid (MOODA), which accumulated in a Δ*fadE32* mutant of *M. smegmatis* when incubated with cholesterol (Figure 4.7). MOODA-CoA has a predicted *m*/*z* value of 952. Consistent with such a role, the enzymatic transformation of COCHEA-CoA to MOODA-CoA required CoASH and yielded stoichiometric amounts of acetyl-CoA (Figure 4.6).

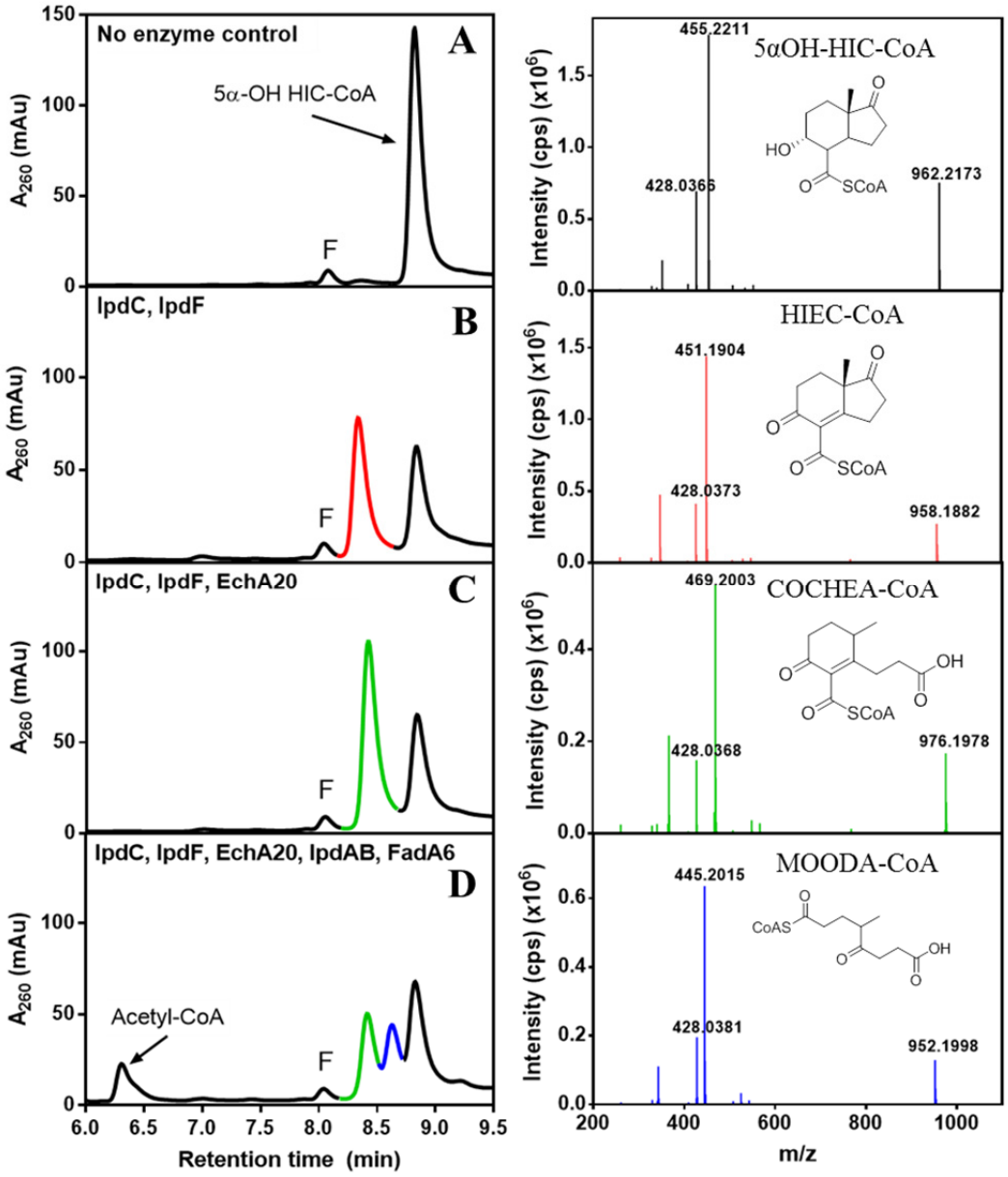


Figure 4.6 LC/MS analyses of the transformation of 5-OH HIC-CoA by purified enzymes.
Left panels show HPLC traces of reaction mixtures containing 100 µM 5-OH HIC-CoA, 125 µM NAD^+, 50 µM CoASH, 5 µM FMN (10 mM phosphate, pH 7.5) and: (A) no enzyme (control); (B) $IpdF_{Mtb}$ and $IpdC_{DOC21}$; (C) $IpdF_{Mtb}$, $IpdC_{DOC21}$, and $EchA20_{RHA1}$; or (D) $IpdF_{Mtb}$, $IpdC_{DOC21}$, $EchA20_{RHA1}$, $IpdAB_{RHA1}$ and $FadA6_{Mtb}$. LC/MS analyses of the reaction products identified the major HPLC Major peaks are color-coded with fragmentation patterns in the right-hand panels and correspond to: 5α-OH HIC-CoA (962 *m/z*), HIEC-CoA (958 *m/z*), COCHEA-CoA (976 *m/z*), and MOODA-CoA (952 *m/z*). Other LC peaks correspond to acetyl-CoA (810 m/z) and (F) FMN.

4.3.6 Bioinformatic analysis of HIP catabolic enzymes

Table 4.2 Annotation of KstR2 regulon

Gene[a]	H37Rv[b]	RHA1[c]	*M. smeg*[d]	CNB-2[e]	*S. denit*[f]	Annotation of gene product	Best hit[g]	Identity[h]
	Rv3548c	RS22710	5999	1286	00355	Short chain type dehydrogenase/ reductase	P22414	42
	Rv3549c	RS22705	6000	1330	12450[i]	Short chain type dehydrogenase/ reductase	A6CQL2	34
echA20	Rv3550	RS27700	6001	1280	00335	HIEC-CoA hydrolase	1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB), P76082	28
ipdA	Rv3551	RS22695	6002	1276	00310	COCHEA-CoA hydrolase, α subunit	glutaconate CoA transferase, α subunit, Q59111	26
ipdB	Rv3552	RS22690	6003	1277	00315	COCHEA-CoA hydrolase, β subunit	glutaconate CoA transferase, β subunit, Q59111	25
ipdC	Rv3553	RS22685	6004	1279	00330	5-OH HIC-CoA reductase	Enoyl-[ACP] reductase II (FabK), Q9FBC5	30
fadA6	Rv3556c	RS22430	6008	1283	00350	β-keto CoA thiolase	3-oxo-acyl CoA thiolase, I6XHI4	38
kstR2	Rv3557c	RS22425	6009	-	-	HIP-CoA repressor[j]	-	-
ipdF	Rv3559c	RS22420	6011	1289	00370	5-oxo HIC-CoA oxidase	levodione reductase, Q9LBG2	39
fadE30	Rv3560c	RS22415	6012	1288	00365	acyl-CoA dehydrogenase	I6YCA3	31
fadD3	Rv3561	RS22410	6013	1360	09100[i]	HIP-CoA synthetase[j]	-	-
fadE31	Rv3562	RS22400	6014	1281	00340	acyl-CoA dehydrogenase	I6YCA3	29
fadE32	Rv3563	RS22395	6015	1282	00345	MOODA-CoA dehydrogenase	I6YCA3	23
fadE33	Rv3564	RS22390	6016	1287	00360	acyl-CoA dehydrogenase	I6YCA3	23

[a]Name assigned based on current study. Identification numbers for the corresponding genes in [b]*Mtb* H37Rv, [c]RHA1 and [d]*M. smegmatis*, Msmeg_ omitted from IDs for simplicity. [e]*C.testosteroni* CNB-2, CtCNB2_ omitted from IDs for simplicity, [f]*Steroidobacter denitrificans* DSM 18526, ACG33_ omitted from IDs for simplicity. [g]Accession number of functionally characterized best hit in NCBI database. [h]Amino acid sequence identity of the *Mtb* enzyme and its experimentally characterized best hit based on full sequence alignment. [i]Not clustered with other Rings C/D catabolic genes. [j]*Mtb* enzyme characterized.

To better understand the activities of IpdC, IpdF, EchA20, IpdAB and FadA6, bioinformatic analyses were performed (Table 4.2). Among characterized homologs, IpdF shares 39% amino acid sequence identity with levodione reductase from *Corynebacterium aquaticum* M-13 [215], which catalyzes the NADH-dependent reduction of a ring ketone. For its part, IpdC shares 30% amino acid sequence identity with FabK from *Streptococcus pneumoniae*, an enoyl-acyl ACP reductase that catalyzes double bond reduction [216], the reverse of the predicted IpdC reaction. Like FabK, purified $IpdC_{DOC21}$ contained a flavin (data not shown). Overall, these analyses are consistent with the ability of IpdF and IpdC to catalyze the transformation of 5α-OH HIC-CoA to HIEC-CoA.

EchA20 is one of twenty-one EchAs in *Mtb*. EchAs are members of the crotonase superfamily that are predicted to catalyze the hydration of enoyl-CoAs [217], of which HIEC-CoA, the substrate of EchA20, is an example. A phylogenetic analysis revealed that among *Mtb* EchAs, only EchA20 clustered with MenB although it shares ~28% amino acid sequence identity, with each of EchA8, EchA18 and MenB. EchA8 and EchA18 are uncharacterized. However, MenB is a 1,4-dihydroxy-2-naphthoyl-CoA synthase that catalyzes an intramolecular Claisen condensation, or Dieckmann cyclization, in menaquinone biosynthesis [218]. This reaction is essentially the reverse of the reaction catalyzed by EchA20. EchA20 also shares 24% amino acid sequence identity with BadI of *Rhodopseudomonas palustris* (Table 4.2). BadI, a β-ketocyclohexanecarboxyl-CoA hydrolase involved in the anaerobic catabolism of benzoate [219], catalyzes a hydrolytic ring-opening reaction similar to that of EchA20. Overall, the bioinformatics analyses indicate that EchA20 catalyzes the hydrolytic ring-opening of HIEC-CoA to COCHEA-CoA via a reverse Dieckmann cyclization. Nevertheless, it is unclear whether other *Mtb* EchAs catalyze similar reactions.

IpdAB shares 24% amino acid sequence identity with glutaconate CoA transferase (GCT) of *Acidaminococcus fermentans* [169], a Class I CoA transferase. Most Class I CoA transferases characterized to date catalyze the transfer of CoA between short acyl chains [167, 169, 220, 221]. However, how a CoA transferase is involved in transforming COCHEA-CoA to MOODA-CoA is not immediately apparent. Finally, FadA6 shares 38% amino acid sequence identity with FadA5, a β-ketoacyl-CoA thiolase involved in cholesterol side chain degradation [65] (Table 4.2), consistent with it catalyzing the thiolysis of a COCHEA-CoA ring-opened product containing a β-keto thioester moiety, to MOODA-CoA and acetyl-CoA.

4.3.7 Validation of HIP catabolism using additional mutants

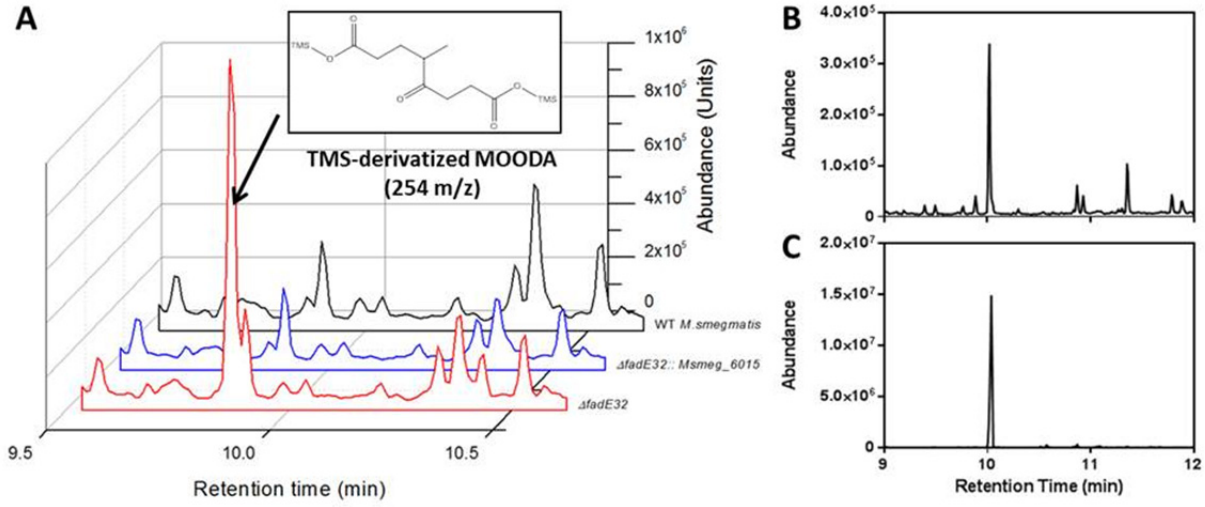


Figure 4.7 Cholesterol-derived metabolite of Δ*fadE32 M.smegmatis*.
(A) GC/MS traces of culture supernatants of cholesterol-grown Δ*fadE32* (red), Δ*fadE32* ::*msmeg_6015* (blue) and WT (black) *M. smegmatis*. The major metabolite observed in the mutant was MOODA (inset). (B) GC/MS trace of the product following hydrolysis of metabolite 952 m/z in 1 M NaOH. (C) MOODA purified from Δ*fadE32 M. smegmatis* incubated with cholesterol

To obtain further evidence for the HIP catabolic pathway suggested by analyses of the *ipd* mutants and the enzymatic transformations of 5-OH HIC-CoA, the following KstR2-regulated genes were deleted in *M. smegmatis* MC^2155 and the corresponding mutants analyzed : Δ*ipdF*, Δ*ipdAB*, Δ*echA20*, and Δ*fadE32* (*msmeg_6011*, *msmeg_6002-6003, msmeg_6001*, and

msmeg_6015, respectively). All four mutants were defective for growth on HIP: the Δ*fadE32* strain grew more slowly on HIP while the other three did not grow at all (Figure 4.8). The growth defect of each mutant on HIP was complemented by the *Mtb* or *M. smegmatis* gene supplied in *trans*. GC/MS analysis of the culture supernatants revealed that the Δ*ipdF* and Δ*echA20* mutants accumulated low amounts of 5-OH HIC (Figure 4.9). Moreover, Δ*fadE32* accumulated a metabolite whose TMS-derivative had an *m*/*z* value of 346 (Figure 4.7). The metabolite that accumulated in the supernatant of the cholesterol-grown Δ*fadE32* mutant was purified and was identified as MOODA based on ^{1}H, COSY, HMBC, and HSQC-NMR analysis (Appendix C).

The Δ*ipdAB* mutant of *M. smegmatis* differed from those of RHA1 and *Mtb* in that it accumulated two major metabolites in the supernatant when incubated with cholesterol. These had *m*/*z* values of 254 and 326 when derivatized with TMS (Figure 4.9). The former was purified, and, based on ^{1}H NMR, was identified as an analog of COCHEA lacking the C-10 carboxyl (Appendix C). A similar result was reported in an *ipdAB* deletion mutant of *Comamonas testosteroni* [214].

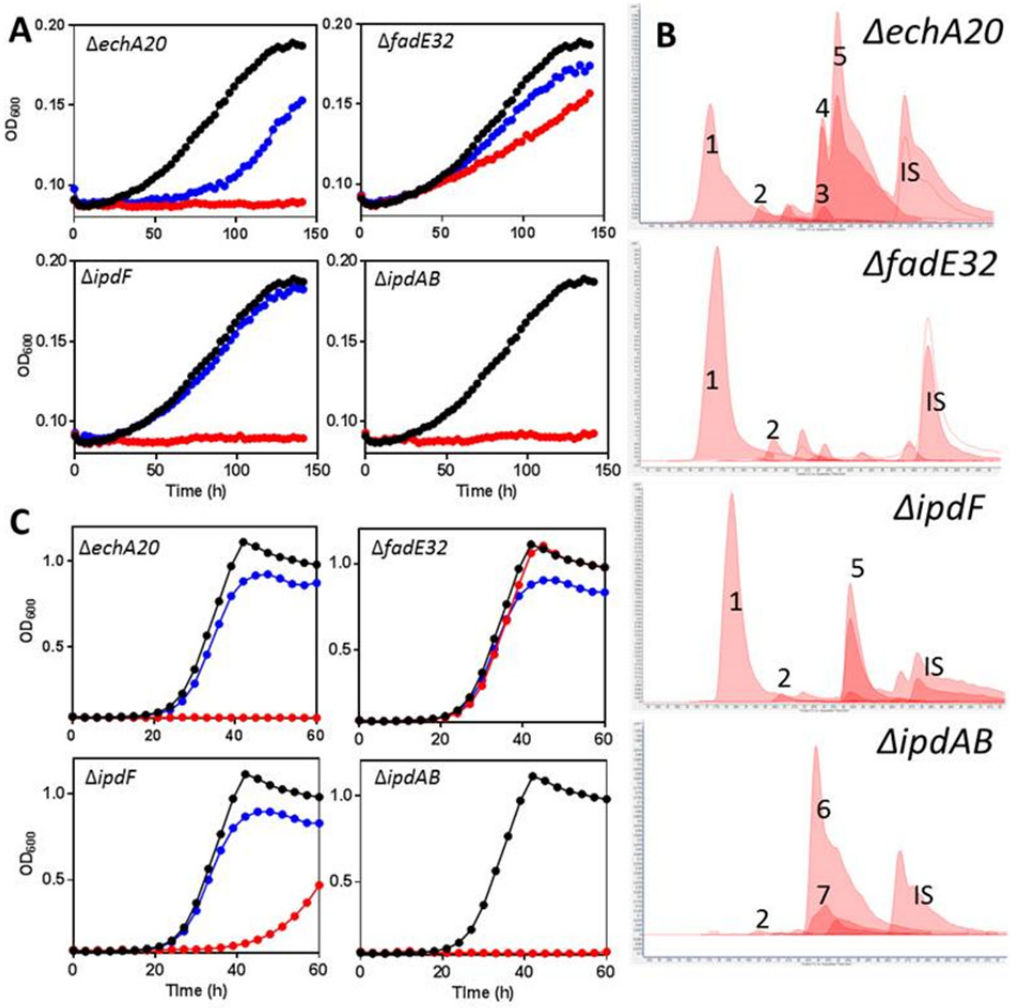


Figure 4.8 Characterization of KstR2 regulon mutants of *M. smegmatis.*
Growth of Δ*echA20,* Δ*fadE32,* Δ*ipdF,* and Δ*ipdAB M. smegmatis* mutants on (A) 1.5 mM HIP or (C) 1 mM HIP + 0.2% glycerol. Curves show WT (black), KstR2 regulon mutants (red), and corresponding complements (blue) and are the mean of three biological replicates. (B) CoA metabolomes of mutants. Numbers correspond to CoASH (1), acetyl-CoA (2), HIEC-CoA (3), 5β-OH HIC-CoA (4), 5α-OH HIC-CoA (5), COCHEA-CoA (6), and unknown CoA thioester 992 m/z (7); IS = *p*-coumaroyl-CoA internal standard. Lighter shaded curves indicate the 768=>261 transition (see Figure 4.4).

The profile of Δ*ipdF* was similar to that of Δ*ipdC*, containing a significant amount of 5α-OH HIC-CoA and a lesser amount of 5β-OH HIC-CoA (Figure 4.8B). This is consistent with the enzymatic studies inasmuch as neither IpdC nor IpdF alone significantly transformed 5-OH HIC-CoA. The CoA metabolome of Δ*echA20 M. smegmatis* also contained significant quantities of

the 5-OH HIC-CoAs (Figure 4.8B). However, it also contained a small amount of a metabolite whose retention time and *m/z* value (958) corresponded to that of the transformation product of 5-OH HIC-CoA by IpdC and IpdF (Figure 4.8B). The CoA metabolome of Δ*ipdAB M. smegmatis* was very similar to those of the corresponding RHA1 and *Mtb* mutants (Figure 4.8B). Finally, the CoA metabolome of Δ*fadE32* was indistinguishable from that of WT *M. smegmatis*.

A summary of the data from Sections 4.3.3 – 4.3.7 is provided in Section 6.3.1.

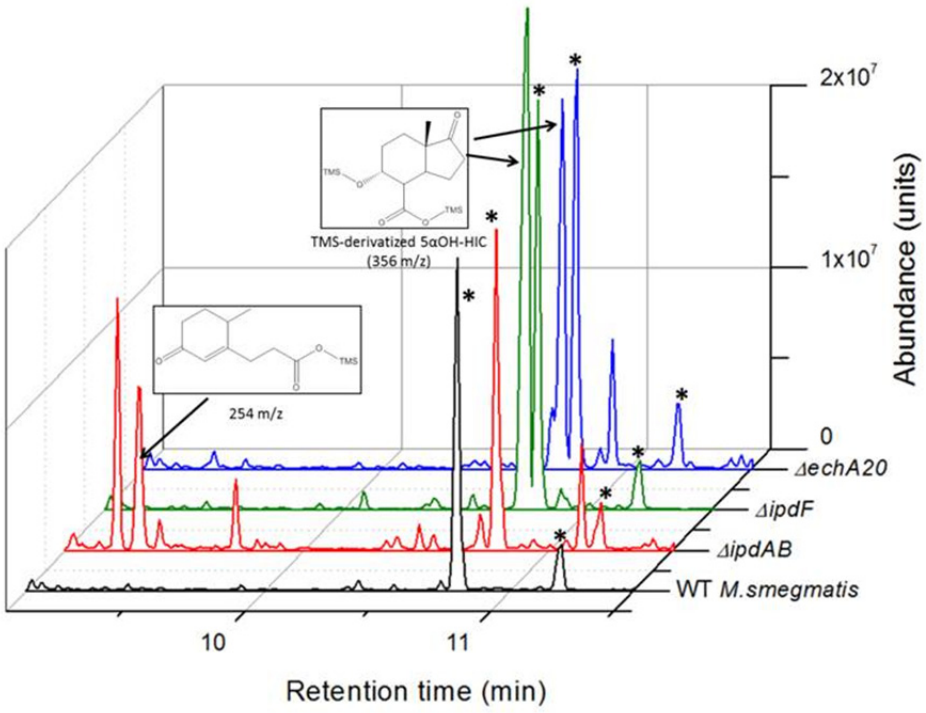


Figure 4.9 Metabolites produced by *M. smegmatis* strains.
GC/MS of culture supernatants of cholesterol-incubated *M. smegmatis* strains. Cells of the indicated strains were incubated with 0.5 mM cholesterol. Insets show structures of TMS-derivatized 5α-OH HIC and 2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1. * indicates unidentified compounds present in all *M. smegmatis* extracts that is unrelated to cholesterol catabolism.

4.3.8 HIP-dependent toxicity

The failure of the Δ*ipdAB* and Δ*ipdC* mutants to grow on cholesterol (Figure 4.1 and 4.2) is in marked contrast to the phenotype of Δ*fadD3* RHA1, which grows on cholesterol to ~50%

the yield of wild-type [79]. More specifically, the failure of the *ipd* mutants to grow on cholesterol despite the fact that the encoded enzymes act downstream of FadD3, suggests that the *ipd* deletions induce some form of toxicity. To explore this further, KstR2 regulon mutants were grown on a second carbon source in the presence of HIP. Interestingly, the Δ*fadE32* and Δ*ipdF* mutants grew on other carbon sources in the presence of HIP (Figure 4.8C), while the Δ*ipdAB*, Δ*ipdC* and Δ*echA20* mutants did not (Figure 4.10A, 4.8C, and 4.2D). The inability to catabolize a secondary carbon source in the presence of HIP indicates that there is a HIP-dependent toxicity in some of the mutants, similar to the cholesterol dependent toxicity observed for *ipdAB* and *ipdC* mutants described above. One possible form of cholesterol (or HIP)-dependent toxicity is the accumulation of propionyl-CoA, which can be relieved by supplementation with vitamin B_{12} [49]. However, supplementation of the Δ*ipdAB*, Δ*ipdC* and Δ*echA20* mutants with vitamin B_{12} did not relieve cholesterol-dependent toxicity, indicating that the basis of toxicity is independent of propionyl-CoA in these mutants. However, in analyzing the CoA metabolites of these mutants, the Δ*ipdAB*, Δ*ipdC* and Δ*echA20* mutants were noted to contain significantly lower levels (<20%) of CoASH as compared to WT when cells were incubated with cholesterol (Figure 4.10B). In contrast, the Δ*fadE32* and Δ*ipdF* mutants contained statistically similar CoASH levels to WT under these conditions. Indeed, the respective levels of CoASH and cholesterol-derived CoA thioesters appeared to be inversely related. For example, when incubated with cholesterol, 5-OH HIC-CoA and COCHEA-CoA accounted for 84 ± 2 and 94 ± 1 % of the total CoA detected in cells of Δ*ipdC* and Δ*ipdAB* RHA1, respectively, indicating that sequestration of CoASH by cholesterol-derived CoA-thioesters may be the basis of toxicity.

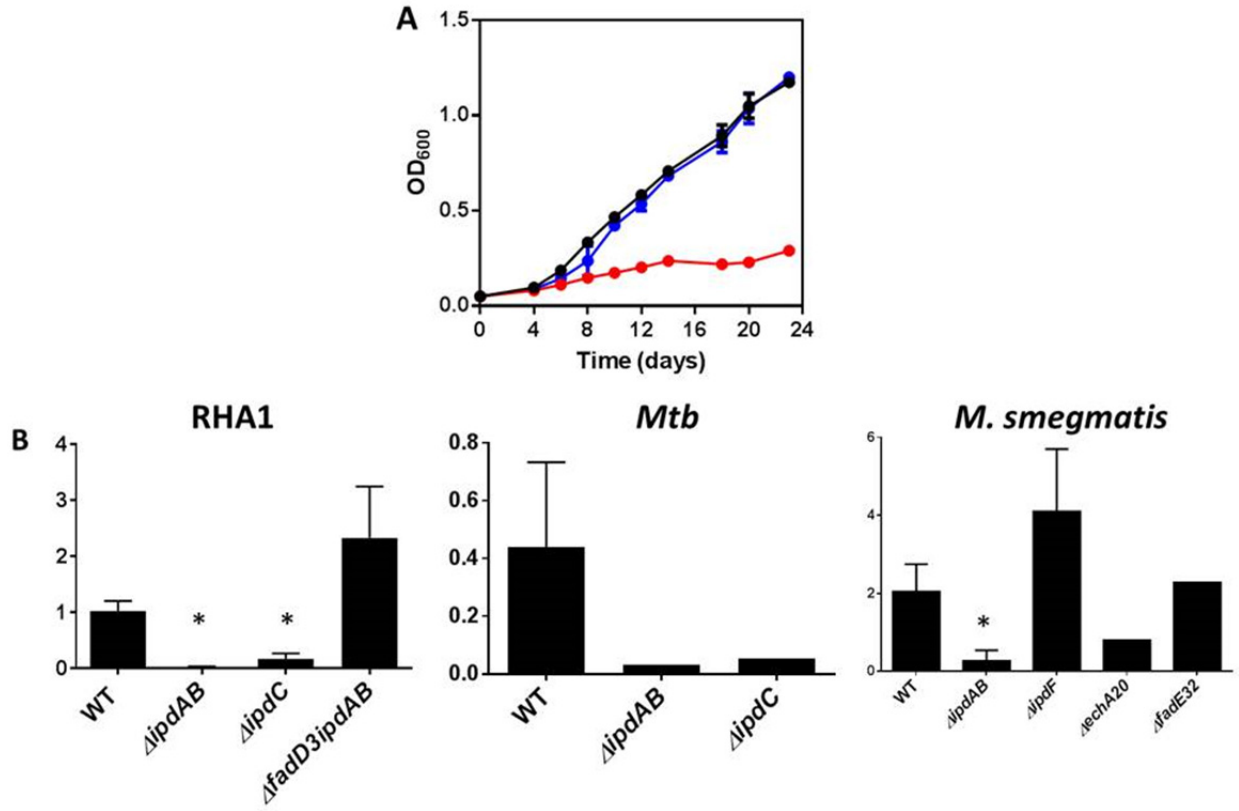


Figure 4.10 Cholesterol-dependent toxicity.
(A) Growth of WT (black), *ΔipdAB* (red), and *ΔipdAB::ipdAB* (blue) *Mtb* grown on 7H9 media containing 0.5 mM cholesterol and 0.2% glycerol. Data represent the average of biological triplicates. (B) The relative abundance of CoASH (768⇒261) was normalized to the internal standard (*p*-coumaroyl-CoA (914⇒407)) in KstR2 regulon mutants. * indicates $p < 0.05$ as compared to WT strain. Error bars represent standard deviation. The number of replicates, N, was: 5, 5, 5, and 3 for WT, Δ*ipdAB*, Δ*ipdC*, and Δ*fadD3* Δ*ipdAB* RHA1, respectively; 2, 1, and 1 for WT, Δ*ipdAB*, and Δ*ipdC* Mtb, respectively; and 4, 5, 1, 4, and 1 for WT, Δ*ipdAB*, Δ*echA20*, Δ*ipdF*, and Δ*fadE32 M. smegmatis*, respectively.

Chapter 5: **Structural and mechanistic characterization of IpdAB**

5.1 Introduction

Elucidation of the catabolism of cholesterol Rings C and D (Chapter 4) implicated IpdAB in the opening of Ring C. This result was unexpected as apparent hydrolytic cleavage of a carbon-carbon bond is unprecedented within the Class I CoA transferase family to which IpdAB belongs. IpdAB is of specific interest due to its role in pathogenesis. *Mtb* deficient in *ipdAB* were unable to grow in macrophages (Section 4.3.2) and *Rhodococcus equi* deficient in *ipdAB* and *ipdA2B2* is patented as a live vaccine for use in horses against pyogranulomatous pneumonia [23].

Herein, a combination of x-ray crystallography, steady-state kinetics, directed-mutagenesis, and isotopic labelling was employed to characterize the function of IpdAB from *Mtb* and RHA1. The data distinguishes IpdAB from Class I CoA transferases and indicate that IpdAB opens the last steroid ring of cholesterol by acting as a hydrolase. Overall, this study identifies a novel catalytic function within the CoT superfamily and provides important insights into a virulence determinant and potential target for novel anti-tuberculosis therapeutics.

Experiments pertaining to data collection,structure refinement, and structural analyses of the IpdAB X-ray crystal structures were performed by Sean Workman, Dr. Liam Worrall, and Dr. Nobuhiko Watanabe from the Strynadka Lab at UBC. Some LC-MS data was acquired by Dr. Israël Casabon. The contents of this chapter were prepared as a manuscript for submission in 2018 as follows:

Crowe, A.M, Workman, S.,Worrall, L., Watanabe, N., Strynadka, N., and Eltis, L.D. (**2018**). IpdAB, a virulence factor in *Mycobacterium tuberculosis*, is a cholesterol ring-cleaving hydrolase.

5.2 Materials and methods

5.2.1 Chemicals and reagents

CoASH, NAD^+, FMN, sodium acetate, sodium propionate, sodium succinate, acetyl anhydride, propionic anhydride, succinyl anhydride, acetoacetyl-CoA, and acetyl-CoA were purchased from Sigma- Aldrich. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Phusion DNA polymerase and Taq DNA ligase was purchased from Thermo Scientific. Oligonucleotides were purchased from Integrated DNA Technologies (IDT). 5αOH-HIC was provided by Victor Snieckus (Queen's University, Kingston, Canada). All other reagents were of HPLC or analytical grade. Buffers and solvents were prepared as previously described in Section 4.2.

5.2.2 Bioinformatic anaylsis of CoA transferases

Amino acid sequences of CoTs and IpdAB homologs were aligned using TCOFFEE-ESPRESSO [183]. For homomeric enzymes containing domains complementary to α and β subunits, the domains were treated separately. Phylogenetic trees were generated using Approximate Likelihood-Ratio Test (aLRT) via PhyML on the phylogeny.fr server [182].

5.2.3 DNA manipulation and plasmid construction:

DNA was propagated, amplified, digested, ligated, and transformed using standard protocols [180]. Oligonucleotide- directed mutagenesis was performed using the QuikChange™ PCR protocol with slight modifications. Briefly, a single 5' phosphorylated mutagenic DNA oligomer was annealed to pTipR1IpdAB then amplified using Phusion DNA polymerase. Taq DNA ligase was added to generate a single stranded mutagenized plasmid. DpnI was used to remove template DNA and the remaining ssDNA was transformed into *E. coli* NovaBlue. pTipR1IpdAB variants $E58^BA$, $E105^AA$, $E105^AD$, $R92^BM$, and $R126^BM$ were generated using

the following respective oligonucleotides (substituted nucleotides are underlined): 5'-pCTGATCACCGACGGTGCGGCCCTGATCTTCGCG-3', 5' pGTCCGCGAAATGGACGCG-GGCATGGTCAAGTGC-3', 5'-pGTCCGCGAAATGGACGACGGCATGGTCAAGTGC-3', 5'-pTCGCCTCCGGCCGGATGCACGTGGTGATGG-3', and 5'-pCAGATGTTCGGCGTC-ATGGGCGCACCCGGCAAC-3'. The nucleotide sequence of variants was confirmed.

5.2.4 Production of COCHEA-CoA:

COCHEA-CoA was produced enzymatically from 5αOH-HIC-CoA as described in Section 4.2.1.1 with the following variations. Following synthesis of 5αOH-HIC-CoA, the reaction mixture was loaded onto 100 mg prepacked 2-(2-pyridyl)ethyl-functionalized silica (Supelco, 54127-U), washed with acetonitrile: isopropanol: water: acetic acid (9:3:4:4 v:v:v:v) to remove unreacted precursors, then eluted in 20 mM ammonium acetate (pH 7.0) in 80% methanol. Methanol was removed under nitrogen. Four μmoles of 5αOH-HIC-CoA were added to a mixture containing $IpdC_{DOC21}$ (5 μM), IpdF (5 μM), $EchA20_{RHA1}$ (5 μM), $IpdAB_{RHA1}$ (1 μM), 1 mM NAD^+, and 10 μM FMN in 50 mM sodium phosphate, pH 8.0. The reaction mixture was incubated at 30°C until completion as determined using HPLC (~3 h). Proteins were precipitated using methanol and the reaction mixture was lyophilized overnight then suspended in 500 μl water. COCHEA-CoA (t_R = 6.1 min) was purified using an HP1100 series HPLC (Agilent Technologies) equipped with a Luna 3u PFP(2) 50 x 4.6 mm column (Phenomenex) operated at 1 ml min^{-1} and separated over a gradient of 20.0 – 30.2% methanol (90%) in 100 mM ammonium acetate, pH 4.5. HPLC- purified fractions of COCHEA-CoA were pooled, lyophilized, and then suspended in 400 μl water. To desalt, COCHEA-CoA was loaded onto an HPLC equipped with the PFP(2) column, washed for 2 min with water + 0.1% acetic acid then

eluted using 100% methanol. Typical mole yields of 25% were observed from racemic 5αOH-HIC-CoA

5.2.5 Protein production and purification:

$IpdAB_{RHA1}$ and variants were produced using RHA1 as a host strain as described in Section 4.2.3 with the following modifications. Following nickel affinity chromatography, $IpdAB_{RHA1}$ and its variants were dialyzed overnight against 25 mM HEPES, pH 7.5, 50 mM NaCl then concentrated to ~5 mg ml^{-1} to which 1:1000 (mol:mol) α-thrombin was added and allowed to digest at room temperature for up to 20 h until complete as determined using SDS PAGE. Digested $IpdAB_{RHA1}$ was loaded onto a 10/100 MonoQ anionic exchange column (GE Healthcare) and purified using an AKTA Purifier (GE Healthcare) operated at 2 ml min^{-1} in 25 mM HEPES, pH 7.0 using a gradient of 150 – 400 mM NaCl. $IpdAB_{RHA1}$ eluting at a conductance of ~36 mS/cm then exchanged into 25 mM HEPES, pH 7.5, 50 mM NaCl, and concentrated to ~15 mg ml^{-1} using a 10K Amicon centrifugation unit (Millipore). Enzyme was flash frozen in liquid nitrogen as beads and stored at -80 C until further use.

5.2.6 Characterization of IpdAB and variants:

Circular dichroic (CD) spectra were recorded at room temperature using a Jasco model J-810 spectropolarimeter. Far-UV CD spectra (190-250 nm) were recorded using a 1 mm quartz cuvette containing 3 μM protein in 10 mM sodium phosphate, pH 8.0. Cuvettes were continuously purged with nitrogen during spectra collection. Spectra were recorded in triplicate and averaged.

Size exclusion chromatography multiangle light scattering (SEC-MALS) data were obtained using a HPLC 1260 Infinity LC (Agilent Technologies) coupled to a Superdex 200 5/150 column (GE Healthcare). Data were collected using a miniDAWN TREOS multiangle

static light scattering device and an Optilab T-rEX refractive index detector (Wyatt Technologies). Sample of 25 µl containing 0.5 mg ml^{-1} protein were analyzed. The HPLC was operated at 0.25 ml min^{-1} in 25 mM HEPES, pH 7.5, 50 mM NaCl. Molecular weights of complexes were calculated using the ASTRA6 program (Wyatt Technologies).

Protein concentrations were measured using the bicinchoninic acid (BCA) protein assay with bovine serum albumin as a standard.

5.2.7 Crystallization of IpdAB and variant

Crystals of $IpdAB_{RHA1}$were obtained by sitting-drop vapour diffusion using a reservoir solution of 1.9 M ammonium sulphate and 0.2 M sodium potassium tartrate. Drops contained 1 µl of protein solution (10 mg ml^{-1}) mixed with an equal volume of reservoir solution and were incubated at room temperature. Single clear bipyramidal crystals were observed within 7 d for WT and 60 d for the $E105^{A}A$ mutant. To generate substrate-bound structures, 5 µl of ~10 mM COCHEA-CoA in 2.4 M ammonium sulphate, 0.2 M sodium potassium tartrate was added directly to protein crystals in the drop and allowed to soak for 24-72 h at room temperature.

5.2.8 Crystallographic analysis and refinement:

Single IpdAB crystals were looped and flash-frozen in liquid N_2 without additional cryoprotectant. Diffraction data for substrate-free $IpdAB_{RHA1}$ and the E105A·COCHEA-CoA complex were collected at Beamline 08B1-1 at the Canadian Light Source in Saskatoon, Saskatchewan. Diffraction data for $IpdAB_{RHA1}$ WT•COCHEA-CoA complex were collected at Beamline 5.0.2 at the Advanced Light Source in Berkeley, California. All data were processed using Xia2 [222], XDS [223], and Aimless [224]. The native structure was solved by molecular replacement using Phaser [225] using a starting model based on the backbone atoms of coordinate set PDB#: 1POI. Enzyme•substrate complexes were grown isomorphously and

determined using difference Fourier analysis. Iterative cycles of model-building and refinement were performed with Coot [199] and Phenix [198]. For the $IpdAB_{RHA1}$ WT•COCHEA-CoA complex, a feature enhanced map [226] was used to guide the placement of ligand in the active site. Active site volume for $IpdAB_{RHA1}$ and GCT (PDB# 1POI) was measured using the CASTp server [204] and a 2.0 Å probe.

5.2.9 *In vitro* activity of IpdAB

The transformation of various acyl-CoAs was performed in 100 µl 10 mM sodium phosphate, pH 8.0 containing 1 µM $IpdAB_{RHA1}$, 100 µM COCHEA-CoA, 125 µM CoASH and, as appropriate 5 µM FadA6 from *Mtb*. Reactions were incubated at room temperature for 1 h and terminated with the addition of 200 µl MeOH (0.5 % acetic acid). To test CoA transferase activity, the reactions contained either 100 µM acetyl-CoA, propionyl-CoA, succinyl-CoA, COCHEA-CoA, or MOODA-CoA and 100 µM each of acetate, propionate, succinate and MOODA. Reaction products were analyzed using a HP1100 Series HPLC attached to a 4.6 x 50 mm Luna 3u PFP(2) column operated as described in Section 4.2.1.

5.2.9.1 Steady-state kinetic characterization of IpdAB:

Steady- state kinetic parameters were evaluated using spectrophotometric assays recorded on a Cary 5K UV-Vis-NIR spectrophotometer (Agilent Technologies). The FadA6-catalyzed thiolysis of acetoacetyl-CoA was determined using 0.1 µM FadA6, 100 µM CoASH, and 50-600 µM acetoacetyl-CoA in 200 µl HEPES, pH 7.5, 10 mM $MgCl_2$ (I= 0.05 M) at 25^{o}C by following the decrease in absorbance at 303 nm due to loss of the acetoacetyl-CoA-Mg^{2+} enolate (ε = 16.9 $mM^{-1}cm^{-1}$ [47]) as previously described [65]. The IpdAB-catalyzed hydrolysis of COCHEA-CoA was followed using a coupled reaction with FadA6 by following the consumption of

COCHEA-CoA at 252 nm (ε = 17.2 mM^{-1} cm^{-1}). Reactions were performed in 200 µl HEPES, pH 7.5, 1 mM $MgCl_2$ (I= 0.01 M) at 25°C containing 50 µM CoASH, 5 µM FadA6, 0.01 µM IpdAB, and 20-110 µM COCHEA-CoA. The specific activity of $IpdAB_{RHA1}$ variants was determined using 2 µM FadA6, 50 µM COCHEA-CoA, 50 µM CoASH, and 0.01 – 3.5 µM IpdAB variant. The extinction coefficient of COCHEA-CoA was empirically derived by measuring the decrease in absorbance at 252 nm upon complete conversion to MOODA-CoA (ε_{310nm} = 11.9 $mM^{-1}cm^{-1}$). Steady-state kinetic parameters were determined by least fit squares fitting of the Michaelis-Menten equation to the data using the GraphPad™ analysis software.

5.2.10 Structure assignment for MOODA-CoA:

The location of the CoA moiety in MOODA-CoA was determined using NMR. 1 µmol COCHEA-CoA was incubated with 2 µM IpdAB, 10 µM FadA6, and 2 µM CoASH in 10 mM sodium phosphate, pH 8.0 at room temperature. Upon completion, the proteins were removed using methanol and the reactions were lyophilized overnight. The resulting residue was suspended and dried twice in 500 µl deuterated methanol (D-MeOD) then suspended in 400 µl D_2O. ^{1}H-NMR, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, and ^{1}H-^{13}C HMBC spectra were collected on a Bruker 600 MHz spectrophotometer. The sample contained MOODA-CoA (1.0 mM), acetyl-CoA (1.3 mM), and CoASH (1.3 mM). NMR data were analyzed using the Academic ACD NMR Processor Software (ACD Labs). Spectra of MOODA-CoA were compared with those recorded for MOODA.

5.2.11 Attempts to trap acyl-enzyme intermediates

Fifty µM COCHEA-CoA and 50 µM $IpdAB_{RHA1}$ were incubated in 50 µl of 10 mM sodium phosphate, pH 8.0 for 5 min then sodium borohydride was added to a final concentration

of 20 mM. After an additional 30 min of incubation, samples were thoroughly desalted and exchanged into water using a 10 kDa Centricon centrifugation unit (EMD Millipore). The molecular weight of intact protein samples was determined using a Waters Xevo G2 qTOF operated by The University of British Columbia Proteomic Core Facility. The specific activity towards COCHEA-CoA was tested for the desalted $IpdAB_{RHA1}$ as described above for the IpdAB variants.

5.2.12 K_D determination for IpdAB E105^{A}A:

The dissociation constant (K_D) of IpdAB E105^{A}A for COCHEA-CoA was determined by following the increase in absorbance at 310 nm upon formation of the complex. IpdAB E105^{A}A (5 µM) was titrated with 0-5.2 µM COCHEA-CoA in 200 µl HEPES, pH 7.5, 1 mM $MgCl_2$ (I=0.01 M) at 25^{o}C. The reference cuvette contained 200 µl HEPES, pH 7.5, 1 mM $MgCl_2$ (I=0.01 M) at 25^{o}C. Difference spectra were recorded after each addition of substrate to the two cuvettes. Data used for calculation represent the difference between titrations of COCHEA-CoA into IpdAB E105^{A}A or assay buffer alone. The instrument was blanked using the assay buffer. Steady-state kinetic parameters and dissociation constants (K_D) were determined by least fit squares fitting of the Michaelis-Menten equation or quadratic binding equation (Eq. 1, Section 2.2.9) to their respective data sets using the GraphPad^{TM} analysis software.

5.2.13 Deuterium incorporation into COCHEA-CoA:

100 µM COCHEA-CoA was incubated for 30 min with and without 2.5 µM IpdAB in 50 µl of 10 mM sodium phosphate, pH 8.0 prepared in D_2O. Reactions were terminated by 5 µl of acetic acid. Samples were centrifuged to remove precipitated protein then filtered 0.45 µm, diluted 1:49 in water and analyzed by LC-MS/MS and/or LC-MS^3.

5.2.14 Proton-deuterium exchange NMR experiments:

Proton-deuterium exchange experiments were performed in 400 µl 10 mM sodium phosphate, pH 8.0 on a Bruker 500 MHz spectrophotometer. Before use, $IpdAB_{RHA1}$ was exchanged into 10 mM sodium phosphate, pH 8.0 in D_2O using a Nanosep 10K Centrifugal Device (PALL Life Sciences). ^{1}H-NMR spectra of 1.6 mM COCHEA-CoA were recorded before and after addition of 50 nM IpdAB every minute for 45 minutes. A delay of 5 minutes occurred between the addition of IpdAB and the first recorded spectra. Aliquots of the reaction at the start and end of the experiments were analyzed by LC-MS to confirm incorporation of deuterium into COCHEA-CoA. A time course plot of relative proton integration was generated by normalizing each proton integration with the C-37 methyl group in the Coenzyme A moiety. The initial rate of deuterium exchange at C-9 of COCHEA-CoA was calculated using a least fit squares of Equation 3 to the first 10 minutes of the progress curve as performed by GraphPad. [S] was calculated using Equation 4 where $[S]_o$, $H_a(t)$ and $H_{0.91}(t)$ represent the initial substrate concentration and the integration of proton (a) or C-37 at time t, respectively.

Equation 3: $\frac{d[S]}{dt} \sim -k[S][E]$

Equation 4: $[S] = [S]_o \frac{H_a(t)}{H_{0.91}(t)}$

5.2.15 ^{18}O labelling of COCHEA-CoA, MOODA-CoA, and IpdAB

100 µM COCHEA-CoA was incubated with and without 5 µM $IpdAB_{RHA1}$ (or 5 µM IpdAB, 20 µM FadA6, and 125 µM CoASH in MOODA-CoA experiments) in 50 µl of 10 mM sodium phosphate, pH 8.0 repaired in 97% ^{18}[O] H_2O (Sigma Aldrich) for 15 minutes at room temperature. Reactions were terminated with the addition of 2 µl acetic acid or 200 µl methanol. Samples were ultracentrifuged (16,000 g, 5 min) to remove protein, filtered 0.45 µm through

PTFE, then immediately frozen in liquid nitrogen and stored at -80^{o}C until analysis by LC-MS/MS. To look at ^{18}O labelling of IpdAB, the reactions were performed as described above; however no acetic acid or methanol was added. IpdAB was desalted into water using a 10K centrifugation unit (Millipore) then subjected to intact protein MS as described above.

5.2.16 Mass spectrometry:

Mass spectra were recorded using a Qstar mass spectrometer (Agilent Technologies) coupled to an HP 1100 series HPLC (Agilent Technologies) equipped with a 50 x 0.3 mm C18(2) column operated at in positive ion mode (Ion spray voltage= +5500V, Ion source temperature= 350^{o}C). CoA thioesters were eluted using a gradient of 100 mM ammonium formate, pH 3.5 into acetonitrile + 0.1% formic acid at a flow rate of 4 µl min^{-1}. High resolution MS^{3} of CoA thioesters were analyzed on a Bruker Impact-II Q-ToF equipped with a 150 x 0.25 mm Luna 3 µm PFP(2) (Phenomenex) column as describedin Section 4.2. CoA thioesters were eluted using a gradient of 100 mM ammonium acetate in 2% methanol and 20 mM ammonium acetate in 98% methanol. Mass spectrometers were calibrated daily.

5.3 Results

5.3.1 Phylogenetic analysis of IpdAB and Class I and II CoA transferases

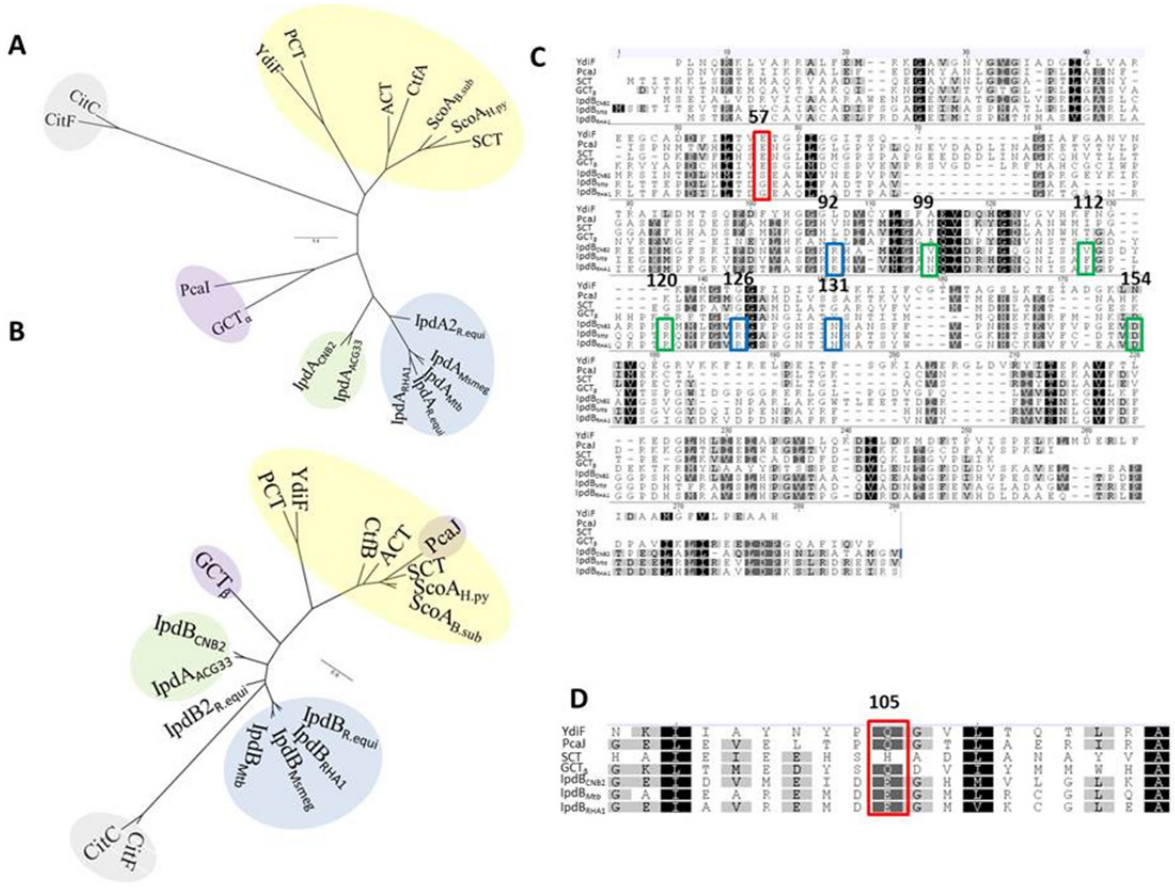


Figure 5.1 Bioinformatic analysis of IpdAB and homologs.
Phylogenetic trees displaying IpdA and α-subunits (**A**) or IpdB and β-subunits (**B**) from Class I and II CoA transferases. Shaded regions indicate Gram positive IpdA or IpdB (blue), Gram negative IpdA or IpdB (green), or Class I β-keto-CoA (purple), Class I (yellow), and Class II (grey) CoA Transferases. Proteins displayed are IpdA from *R.jostii* RHA1 (IpdA$_{RHA1}$), *R. equi* (IpdA$_{R.equi}$; IpdA2$_{R.equi}$), *M. smegmatis* (IpdA$_{M.smeg}$), *M. tuberculosis* (IpdA$_{Mtb}$), *Steroidobacter denitrificans* (IpdA$_{ACG33}$), and *Comamonas testosteroni CNB-2* (IpdA$_{CNB-2}$); β-ketoadipyl-CoA transferase from *Pseudomonas putida* (PcaI), Glutaconate CoA transferase from *Acidaminococcus fermentans* (GCT), Citrate Lyase from *Enterobacter aerogenes* (CitC), Citrate Lyase from *Clostridium argentinense* (CitF), Butyrate-acetoacetate CoA transferase from *Clostridium acetobutylicum* (CtfA), Acetate CoA transferase from *Escherichia coli* (ACT), Succinyl-CoA transferase from *Bacillus subtillus* (ScoA$_{B.sub}$), Succinyl-CoA transferase from *Helicobacter pylori* (ScoA$_{H.pylori}$), Succinyl-CoA transfease from pig heart, Propionyl-CoA transferase from *Clostridium propionicum* (PCT), and Acetate CoA transferase from *Escherichia coli* (YdiF). Additional information is available in the Supplemental Information (SI). (**C**) Amino acid alignment of IpdB from CNB-2, *Mtb,* and RHA1 with well characterized Type I CoA transferases. Red box indicates location of highly conserved catalytic glutamic acid in Class I CoA transferases. Numbering corresponds to residue number in IpdB$_{RHA1}$ (**D**) Conservation of active site glutamate in IpdAs.

To better understand the relationship of IpdAB to CoTs, the phylogeny of IpdA and IpdB with respect to the α and β subunits, respectively, of characterized Class I and II CoTs was analyzed. Sequence alignments were structure-guided and only sequences corresponding to the conserved core domains of the CoTs were used in the phylogenetic analysis to minimize biases resulting from insertions (Figure 5.1). This analysis revealed that IpdA homologs from steroid-degrading bacteria formed a clade distinct from each of the three formed by the α subunits of the Class I CoTs, Class I β-keto-CoTs and Class II CoTs, respectively (Figure 5.1). IpdAs clustered most closely with the heterotetrameric Class I β-keto- CoTs and, within the IpdA clade, orthologs from actinobacteria (blue) and proteobacteria (green) formed subclades. Analysis of IpdBs with the CoT β subunits revealed similar relationships except that PcaI, the β subunit of the β-ketoadipate:succinyl-CoA transferase PcaIJ, clustered differently than the α subunit, PcaJ (Figure 5.1).

5.3.2 Characterization of IpdAB

To biochemically and structurally characterize IpdAB, $IpdAB_{Mtb}$ and $IpdAB_{RHA1}$ were overproduced and purified to apparent homogeneity using metal affinity chromatography. Both orthologs were eluted as colourless preparations. RHA1 was a better expression host than *E. coli*, although $IpdAB_{RHA1}$ was produced to higher levels than $IpdAB_{Mtb}$ and was more stable once purified. SEC MALS analyses revealed that $IpdAB_{RHA1}$ had a molecular weight of 109.7 ± 0.1 kDa, corresponding to an $\alpha_2\beta_2$ heterotetramer as observed in Class I CoTs.

5.3.2.1 *In vitro* transformations and steady state kinetics

All CoTs characterized to date catalyze the inter- or intra-molecular transfer of CoA from a CoA thioester to a free carboxylate [167, 169]. To test whether $IpdAB_{RHA1}$ had similar activity, the enzyme was incubated with a variety of CoA donors and small organic acids. HPLC analysis failed to detect the transfer of CoA from any of acetyl-CoA, propionyl-CoA and succinyl-CoA to acetate, propionate, or succinate (data not shown). Additionally, IpdAB did not detectably catalyze the hydrolysis of CoA donors (data not shown).

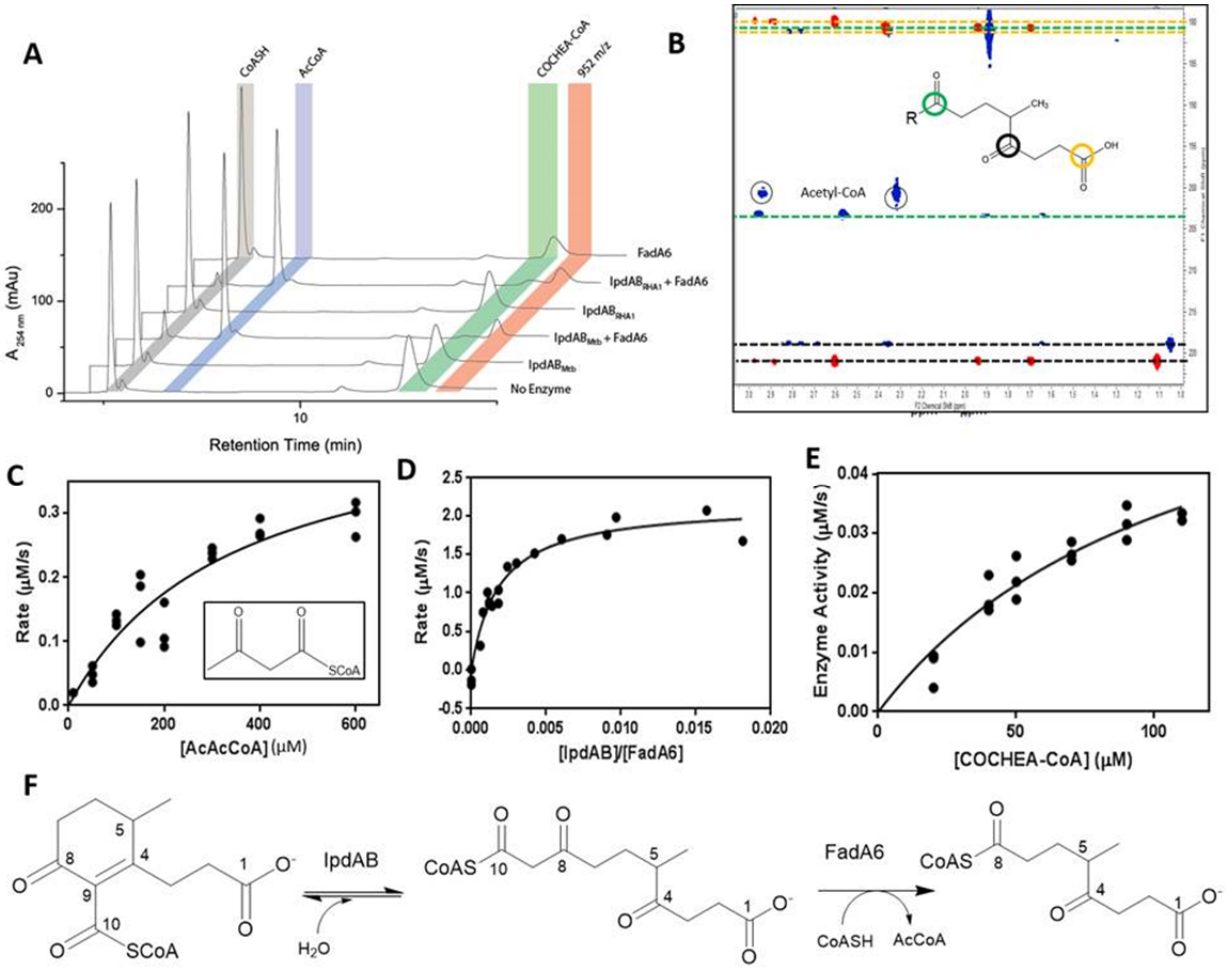


Figure 5.2 *In vitro* activity of IpdAB.
(**A**) HPLC trace of *in vitro* transformation of COCHEA-CoA with $IpdAB_{RHA1}$ and FadA6. 100 μM COCHEA-CoA was incubated with 1 μM $IpdAB_{RHA1}$/ $IpdAB_{Mtb}$, 5 μM FadA6, and/or 125 μM CoASH in 10 mM sodium phosphate pH 8.0. (**B**) ^{13}C – ^{1}H HMBC NMR spectral overlay (0-3 ppm ^{1}H; 180-230 ppm ^{13}C) of MOODA and MOODA-CoA. Spectra were recorded on a 600 MHz spectrophotometer in D_2O. (**C**) Steady state analysis of FadA6 and

acetoacetyl-CoA (AcAcCoA). (**D**) Effect on the ratio between IpdAB and FadA6 on *in vitro* turnover of COCHEA-CoA to MOODA-CoA. (**E**) Steady state analysis of IpdAB and FadA6 towards COCHEA-CoA. Curves for (**C**, **D,** and **E**) indicate best squares least fit of Michaelis-Menten equation to the data. Data was obtained in 200 µl HEPES pH 7.5 (I= 0.01 M) at 25°C. (**F**) Diagram outlining the proposed reaction catalyzed by IpdAB and FadA6. Carbon numbering used throughout this thesis are identified.

Table 5.1 Steady state parameters for FadA6 and IpdAB

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($\times 10^4$ M^{-1} s^{-1})
$FadA6_{Mtb}$	AcAcCoA	4.8 (0.8)	300 (100)	1.6 (0.8)
$IpdAB_{RHA1}$	COCHEA-CoA	5.8 (0.8)	120 (40)	5 (2)

5.3.3 IpdAB catalyzed the efficient transformation of COCHEA-CoA

Incubation of COCHEA-CoA with IpdAB and a thiolase yielded a metabolite presumed to be MOODA-CoA (Section 4.3.5). To extend this observation, the ability of IpdAB to transform COCHEA-CoA and other CoA thioesters was tested *in vitro*. In the presence of a thiolase, either FadA5 or FadA6 from *Mtb*, and CoASH, both $IpdAB_{Mtb}$ and $IpdAB_{RHA1}$ stoichiometrically transformed COCHEA-CoA into acetyl-CoA and a second CoA thioester of 952 *m*/*z*, consistent with MOODA-CoA (Figure 5.2A). No compounds other than COCHEA-CoA, MOODA-CoA, acetyl-CoA, and CoASH were detected using LC-MS. Neither IpdAB nor a thiolase alone detectably transformed COCHEA-CoA. Since $IpdAB_{Mtb}$ and $IpdAB_{RHA1}$ displayed similar activities, subsequent studies were performed using the more stable ortholog, $IpdAB_{RHA1}$.

To evaluate the efficiency of the IpdAB reaction, a coupled assay with $IpdAB_{RHA1}$ and FadA6 was established. The thiolytic activity of FadA6 towards β-keto CoA thioesters was first evaluated using acetoacetyl-CoA by following the decrease in absorbance at 310 nm from the

acetoacetyl-CoA-Mg^{2+} enolate. The steady-state kinetic parameters indicated that FadA6 readily cleaves acetoacetyl-CoA (Table 5.1). In a coupled assay with IpdAB, the rate of COCHEA-CoA consumption, as measured spectrophotometrically by the decrease in absorbance at 252 nm, was limited by FadA6 at mole ratios below 400:1 (FadA6:IpdAB; Figure 5.2C). The product of the IpdAB reaction was turned over by FadA6 at a rate of 0.014 ± 0.006 s^{-1}. In the presence of saturating amounts of FadA6, IpdAB had a relatively high specificity constant for COCHEA-CoA (k_{cat}/K_M = 5 × 10^4 $M^{-1}s^{-1}$; Figure 5.2D, Table 5.1). Similar results were obtained using FadA5 instead of FadA6 (data not shown).

NMR was used to definitively identify the second CoA thioester produced in the enzymatic transformation of COCHEA-CoA as MOODA-CoA in which the CoA moiety is attached at carbon-8 (C-8) of MOODA. More specifically, comparison of the ^{1}H-^{13}C HMBC spectra of MOODA-CoA and MOODA revealed a ~23 ppm increase in the chemical shift for C-8 in MOODA-CoA as compared to MOODA (Figure 5.2B). Given that FadA5 and FadA6, as well as all β-keto-CoA thiolases characterized to date, act on β-keto-CoA thioesters, the location of the CoA moiety on MOODA-CoA indicates that the substrate for FadA6 is 6-methyl-3,7-dioxodecanedioyl-CoA (MeDODA-CoA) (Figure 5.2F). Failure to observe this species in the absence of FadA6 may be because the equilibrium lies far to the left.

5.3.4 The structural fold of IpdAB is typical of Class I CoTs

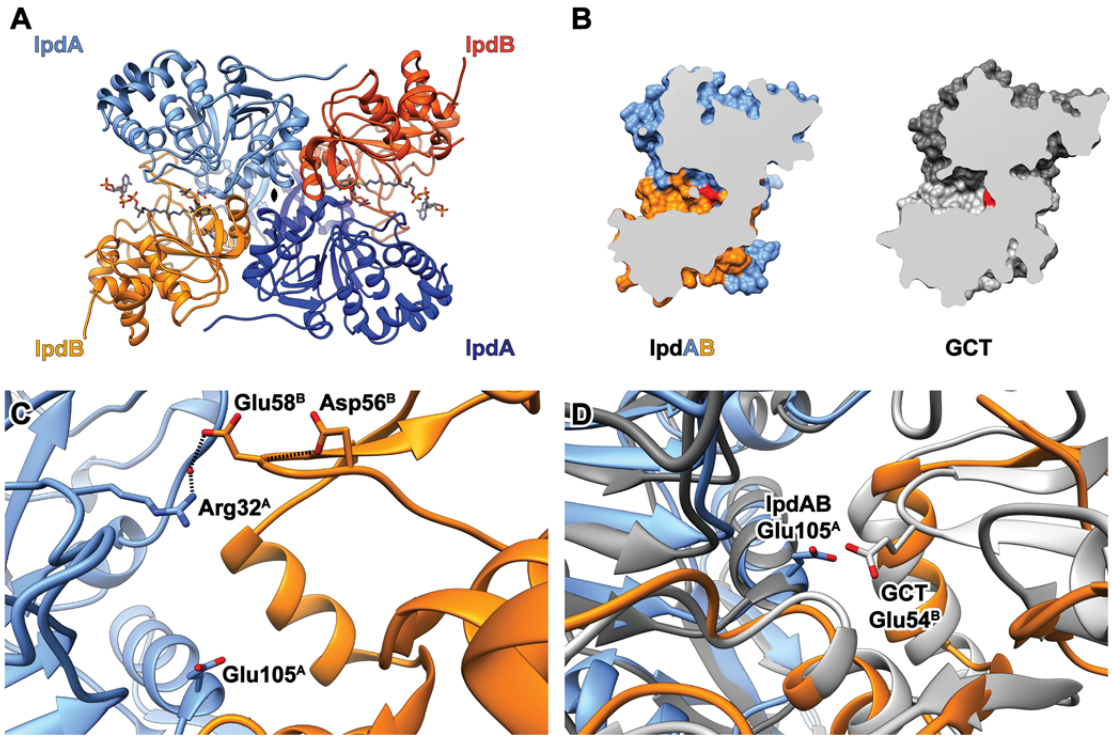


Figure 5.3 Structure of $IpdAB_{RHA1}$
(**A**) Biological assembly of $IpdAB_{RHA1}$. IpdA and IpdB subunits are depicted in blue and orange, respectively. COCHEA-CoA (lactonized) is displayed in the active sites. (**B**) Surface topology of the $IpdAB_{RHA1}$ (left) and GCT (right) active site pockets. The location of the catalytic Glutamate is displayed in red. (**C**) Location of candidates for the catalytic acidic residue in the β-subunit from structural alignments. Ordered water molecule is shown as a red dot. (**D**) Structural overlay of the $IpdAB_{RHA1}$ and GCT (grey) active sites including catalytically relevant residues.

To further characterize IpdAB, the X-ray crystallographic structure of $IpdAB_{RHA1}$ was solved to 1.7 Å resolution. The asymmetric unit of the P43212 $IpdAB_{RHA1}$ crystals contained a single αβ heterodimer. The $\alpha_2\beta_2$ assembly, observed in solution, was formed from two αβ protomers related via a two-fold axis of symmetry in the crystal lattice (Figure 5.3A). Ordered electron density for all the residues of IpdAB was observed except for the N terminal methionine of IpdA, which was removed in cleaving the affinity tag, and the first six residues of IpdB. Data collection and structure refinement statistics are presented in Appendix B.

IpdAB possesses the same core fold as Class I CoTs. More specifically, the core of IpdA consists of a seven-stranded parallel β-sheet sandwiched between helices α1-4 on one side and α5-7 on the other. The core of IpdB is similar except that the β-sheet is six-stranded (five parallel strands, one anti-parallel) and is sandwiched between fewer helices (α1-2 and α5-6, respectively). IpdA is further distinguished from IpdB by a 13-residue loop consisting of Thr144-Thr156 that overlaps the second IpdA subunit (denoted by $^{A'}$) and likely stabilizes the tetrameric assembly via two hydrogen bonds: one between Tyr147^{A} (O^{η}) and Arg141$^{A'}$ (peptide N), and a second between Arg141^{A} (N^{η}) and Pro146$^{A'}$ (peptide O). Consistent with the phylogenetic analyses, the structural fold of IpdAB closely resembles that of GCT (rmsd on 371 shared C_{α} with GCT, PDB# 1POI = 2.39 Å).

5.3.5 IpdAB has distinct active site residues

Each $IpdAB_{RHA1}$ protomer harbours a single, large active site pocket located at the interface between the two subunits. More specifically, the pocket is ~3300 Å^3 and lies between helix α5 of IpdA, helices α3,4 of IpdB, and the β-sheet of IpdB. A channel of ~28 Å projects out of the active site and follows the contours of the interface between the two subunits. The active site pocket of IpdAB is approximately two times larger than that of GCT (1780 Å^3) and is much less solvent-exposed (Figure 5.3B). These differences reflect IpdB's active site loop which, at 6 residues (Thr55^{B} to Leu60^{B}), is 18 residues shorter than that of GCT. The corresponding loop in GCT also carries Glu54$^{\beta}$, the key catalytic residue of Class I CoTs [169]. The IpdB loop carries two acidic residues, Asp56^{B} and Glu58^{B}. However, the C^{δ} of Glu58^{B} carboxylate is located ~13 Å away from the C^{δ} of Glu54^{B} in GCT when superimposed with IpdAB and appears to stabilize the protomer by forming a water mediated hydrogen bond with Arg32^{A} (Figure 5.3C). Asp56^{B} appears to contribute to the stability of the active site loop via hydrogen bonds to the peptide

nitrogen of Ala59^{B} (Figure 5.3C), and is not orientated towards the substrate-binding pocket, suggesting it does not have a role in catalysis. Nevertheless, IpdAB does contain an acidic residue at the center of its active site: Glu105^{A} (Figure 5.3D). Inspection of the sequence alignments used for the phylogenetic analyses revealed that Glu105^{A} is conserved in IpdAB orthologs but not in Class I CoTs (Figure 5.1D). By contrast, the typical catalytic base, Glu54$^{\beta}$, of Class I CoTs does not occur in any of the IpdAB orthologs (Figure 5.1C).

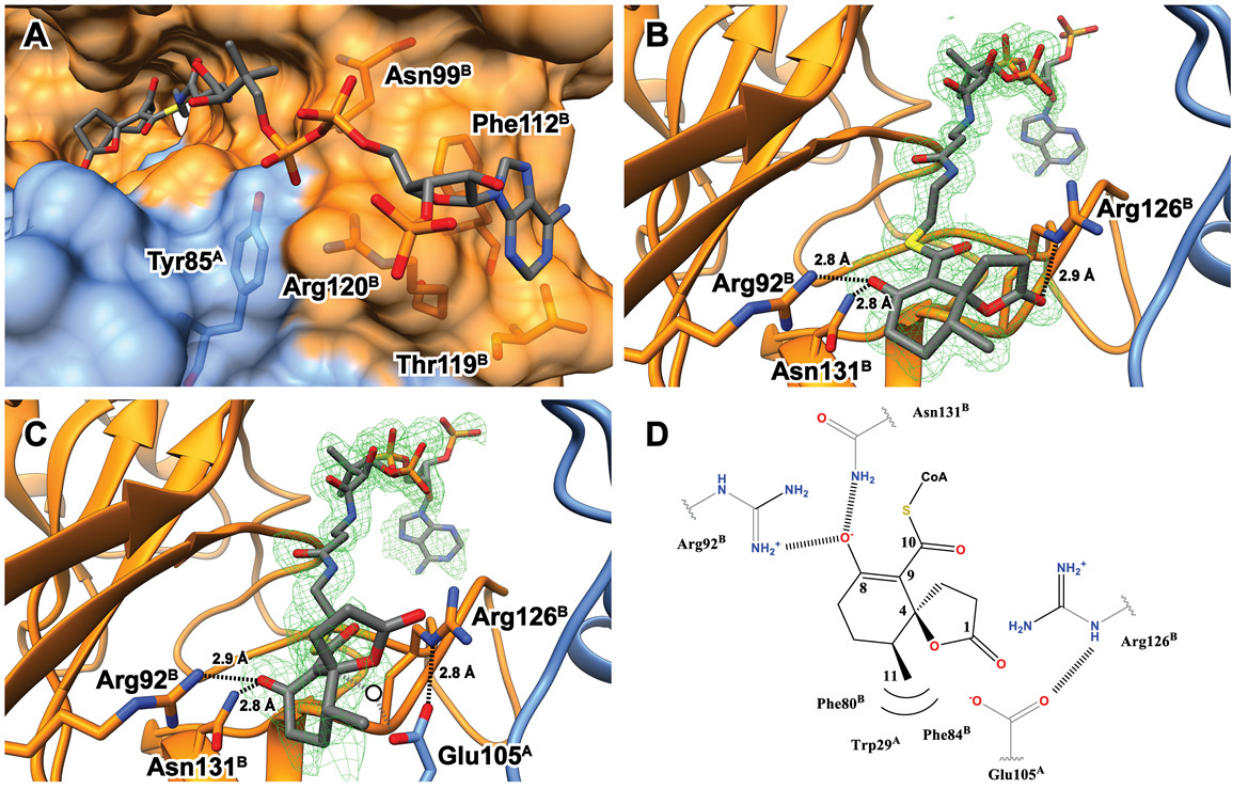


Figure 5.4 Structure of IpdAB·COCHEA-CoA.
(**A**) Lactonized COCHEA-CoA binds along a long channel between IpdA (blue) and IpdB (orange). Predicted CoA moiety binding residues are shown. (**B,C)** Active site of IpdAB$_{RHA1}$ E105^{A}A·COCHEA-CoA (**B**) and WT·COCHEA-CoA (**C**) displaying predicted catalytically relevant residues. Shown are the respective F_o-F_C maps (green mesh) around the substrate contoured at 2.0 σ. Location of proposed catalytic water in the absence of lactonized COCHEA-CoA is displayed as a white dot. (**D**) Diagram illustrating residues predicted to make contacts with the lactonized COCHEA-CoA in the IpdAB$_{RHA1}$ active site.

5.3.6 Structure of IpdAB·COCHEA-CoA complexes

To gain more insight into the function of IpdAB, structures of $IpdAB_{RHA1}$ WT and $E105^{A}A$ soaked with COCHEA-CoA were solved to 1.6 and 1.4 Å, respectively. The location of the thioester sulphur atom in the WT·COCHEA-CoA complex was determined using anomalous scattering. Refinement data are summarized in Appendix B. Structures of substrate-bound $IpdAB_{RHA1}$ were remarkably similar to the substrate-free $IpdAB_{RHA1}$ structure (rmsd on 462 shared C_{α}= 0.2 Å). The few notable differences include a displacement of the sidechain atoms of $Arg126^{B}$ and $Arg120^{B}$ by 3.3 and 1.0 Å, respectively, presumably to accommodate the substrate, and the movement of the side chain atoms of $Arg147^{B}$ and $Tyr85^{A}$ towards the substrate by 4.0 and 1.0 Å, respectively. With the exception of the acyl moiety in the active site and the substituted residue, the WT and $E105^{A}A$ substrate-bound structures were essentially indistinguishable (rmsd of 0.07 Å on 489 common C_{α} atoms). In both IpdAB·COCHEA-CoA complexes, the substrate is bound in a channel between the subunits with the CoA adenine binding ~21 Å away from the CoA thioester sulfur atom in the active site (Figure 5.4). This is similar to what has been reported in Class I CoTs [147]. For example, superposition of the IpdAB·COCHEA-CoA complexes with those of YdiF from *E.coli* (PDB# 2AHV) and pig heart SCOT (PDB# 3OXO) revealed that their respective CoA sulphur atoms are within 1.2 Å of each other. Nevertheless, the adenine-binding pocket of IpdAB is located ~7 Å from the position where it is found in CoTs. Notable interactions with the CoA moiety include coordination of the diphosphate group by the side chains of $Arg120^{B}$, $Asn99^{B}$, and $Tyr85^{A}$; π-π stacking between the adenine group and $Phe112^{B}$; and coordination of the adenine amino group by $Thr119^{B}$ (Figure 5.4A).

Electron density in the active site of the E105^{A}A·COCHEA-CoA complex clearly revealed a 5-membered lactone ring generated from the cyclization of the C-1 carboxylate to C-4 (Figure 5.4B). The cyclohexenone ring of the bound COCHEA-CoA is largely planar with the C-8 oxo and C-11 methyl group orientated 180^{o} from each other. The lactone ring is orientated perpendicularly to the cyclohexanone ring such that C-4 is an (*S*) stereocenter. No electron density was observed corresponding to an (*R*) C-4 stereocenter (Figure 5.4B). Torsional angles about C-4 deviate slightly from an optimal sp^3 hybridized center due to a slight bend of the lactone ring away from the C-11 methyl resulting in angles of 104 and 110^{o} for O-1/C-4/C-1 and O-1/C-4/C-5 bonds, respectively. Similarly, the C-8 oxo is bent slightly (Φ=116^{o}) towards Arg92^{B} and forms hydrogen bonds with that residue (2.9 Å) and Asn131^{B} (2.8 Å) (Figure 5.4B). The C-11 methyl group sits in a hydrophobic pocket formed by Phe80^{B}, Phe84^{B}, Val83^{B} and Trp29^{A}. The C-1 ester carbonyl hydrogen bonds (3.1 Å) to an ordered water molecule coordinated by Arg126^{B}, Arg32^{A}, and Asp80^{A}. The C-10 thioester carbonyl orients towards the N$^{\eta}$ of Arg126^{B} at a distance of 4.0 Å.

Electron density in the active site of the WT IpdAB$_{RHA1}$·COCHEA-CoA complex was less resolved than in the E105^{A}A·COCHEA-CoA complex. However, additional residual density outside of the active site fit the CoA moiety well, suggesting that the substrate may bind in more than one conformation (Figure 5.4C). Using a feature enhanced map [226], density for the 5-membered lactone observed in the E105^{A}A structure could be clearly identified. No density for the C-1 carboxylate was observed indicating that the un-lactonized form of COCHEA-CoA was not present. Comparison of the lactonized COCHEA-CoA from the two structures illustrates distinct differences (Figure 5.4B,C). First, the cyclohexanone ring is not planar in the WT complex: the C-8 oxo is bent upwards towards Arg92^{B}. Second, the C-8 oxo and C-10 thioester

are also not planar, suggesting C-9 is sp^3 hybridized and protonated. Third, the lactone exhibits bond torsion between C-1/C-2/C-3 (Φ=96^o) deviating from the predicted 105^o. Lastly, the lactone in the WT structure is rotated ~45^o away from its location in the E105^{A}A complex, presumably to avoid steric clash with Glu105^A. Interestingly, the O$^\varepsilon$ of Glu105^A forms a 2.8 Å hydrogen bond with the N$^\varepsilon$ proton of Arg126^B positioning the Glu105^A carboxylate directly under C-4 of COCHEA-CoA (Figure 5.4C). Presumably, in the absence of substrate lactonization, a catalytic water molecule could be accommodated between C-4 of COCHEA-CoA and Glu105^A (Figure 5.4C, white circle).

In neither complex is there any evidence of an oxyanion hole to accommodate the CoA thioester oxo of the bound COCHEA-CoA. By contrast, an oxyanion hole appears to stabilize the thioester oxo in YdiF and SCOT [175].

5.3.7 Identification of catalytically essential residues

The structural data indicate direct interactions between the acyl-moiety of COCHEA-CoA and each of the side chains of Arg92^B, Glu105^A and Arg126^B such that they may potentially activate the substrate during catalysis. Glu58^B on the other hand is located away from the bound substrate and thus predicted to be non-essential for catalysis. Importantly, these residues are all conserved in IpdABs (Figure 5.1C). To test the catalytic relevance of these residues, each was substituted and the variants purified. CD spectroscopy and SEC MALS indicated that all variants had the same global secondary and tertiary structures as WT IpdAB (Figure 5.5). The E58^{B}A variant showed only a minor reduction in specific activity compared to WT, consistent with its localization and a non-catalytic role. In contrast, the other variants (E105A^A, E105D^A, R92M^B, and R126M^B) had no detectable activity, consistent with these residues playing an essential role in catalysis (Table 5.2).

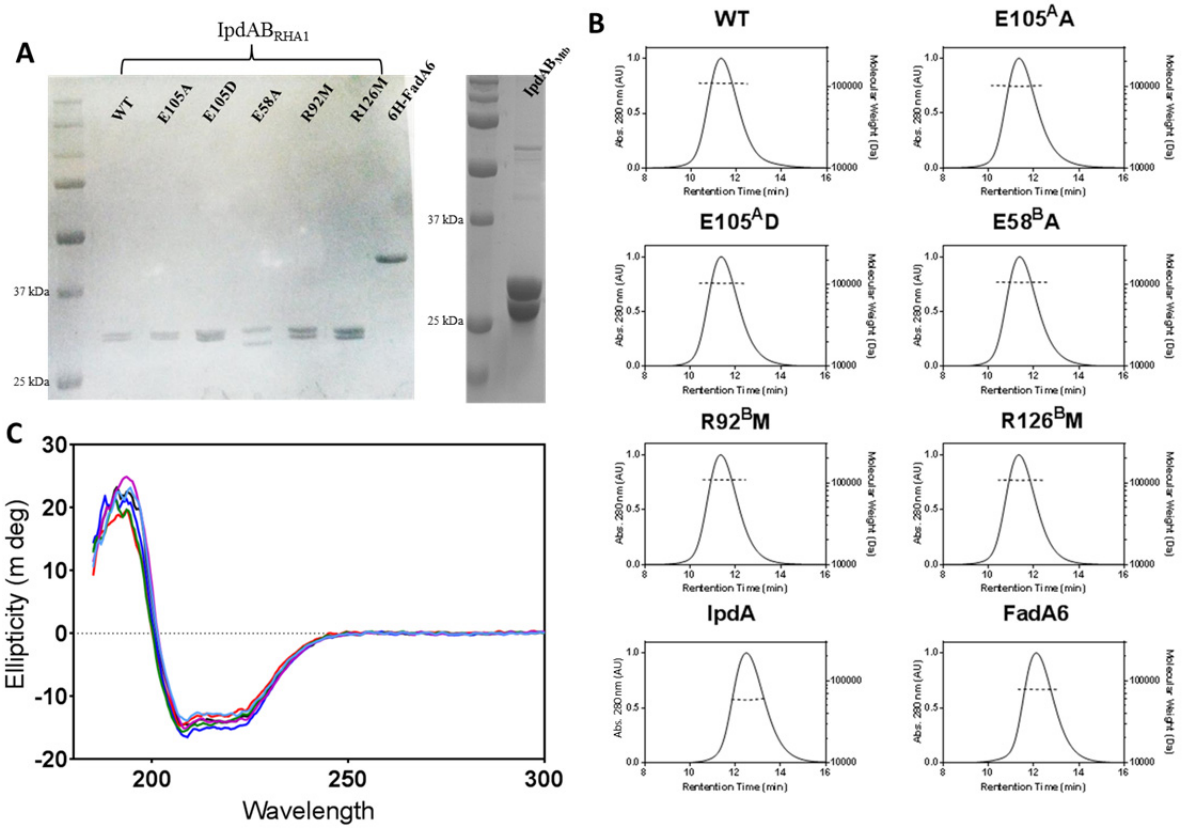


Figure 5.5 Characterization of IpdAB
(**A**) SDS-PAGE showing ~1 µg of indicated protein (**B**) SEC-MALS analysis of IpdAB and variants. Shown is the absorbance at 280 nm and the calculated molecular weight by the ASTRA6 software of proteins eluting from a Superdex 200 5/150 column. (**C**) Circular dichroism spectra of 3 µM IpdAB (blue) or variants E105^{A}A (green), E105^{A}D (black), E58^{B}A (red), R126^{B}M (purple), and R92^{B}M (light blue) in 10 mM sodium phosphate, pH 8.0.

Table 5.2 Specific activity of IpdAB variants

IpdAB variant	Specific Activity (s^{-1})
WT	1.6 (0.4)
E58^{B}A	1.3 (0.3)
E105^{A}A	<0.01
E105^{A}D	<0.01
R126^{B}M	<0.01
R92^{B}M	<0.01

5.3.8 The IpdAB reaction mechanism does not appear to involve a glutamyl-CoA intermediate

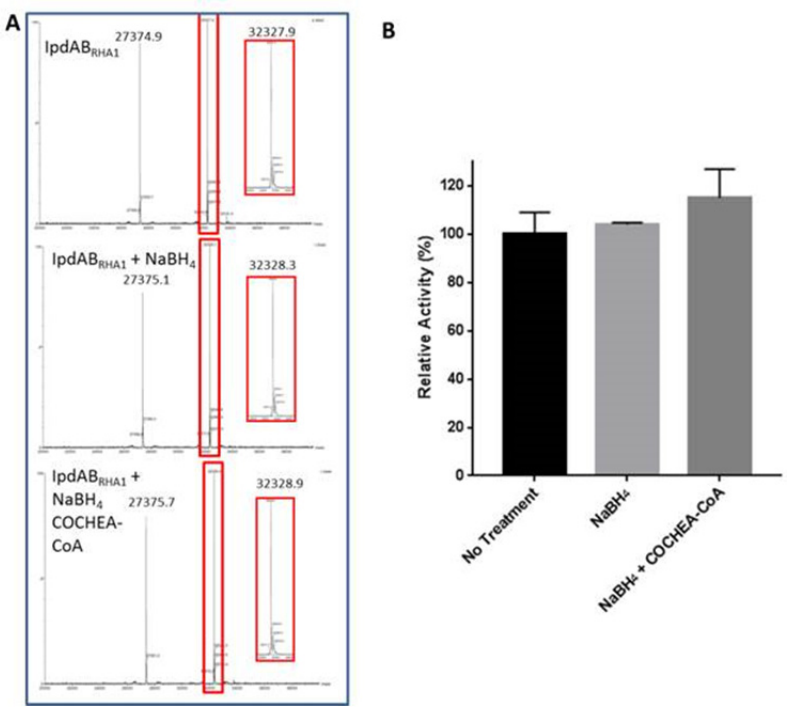


Figure 5.6 IpdAB is not inhibited by sodium borohydride.
IpdAB was incubated in 50 µl 10 mM sodium phosphate at pH 8.0 for 20 min in the presence of the listed compounds then desalted. (**A**) Shown are the mass spectra acquired from intact-protein LC-MS on each treated IpdAB sample. Red squares indicate mass of IpdA. (**B**) The relative specific activity of each $NaBH_4$ treated IpdAB is shown towards 50 µM COCHEA-CoA in the presence of 150 µM CoASH, and 4 µM FadA6 in 200 µl HEPES, pH 7.5 (I= 0.01M). Bars indicate standard deviation (N=3).

In Class I CoTs the glutamyl-CoA intermediate has been trapped by incubating the reaction mixture with sodium borohydride ($NaBH_4$). This reduces the glutamyl-CoA intermediate to a thiohemiacetal which may be observed using LC-MS [227]. This also leads to the enzyme's irreversible inhibition. To test for the occurrence of a glutamyl-CoA intermediate in the turnover of IpdAB, we incubated a mixture of $IpdAB_{RHA1}$ and COCHEA-CoA with $NaBH_4$. LC-MS analysis of the incubated enzyme revealed no significant difference in mass of either the α (32328 Da) or β (27375 Da) subunits with respect to untreated IpdAB (Figure 5.6A). Moreover, treatment with $NaBH_4$, either in the presence or absence of COCHEA-CoA, did not

significantly reduce the specific activity of IpdAB following removal of remaining sodium borohydride (Figure 5.6B). These results strongly indicate that the IpdAB mechanism does not involve a glutamyl-CoA thioester intermediate.

5.3.9 Formation of a β-keto enolate in the E105^{A}A variant

The observation of the lactonized COCHEA-CoA in the IpdAB E105^{A}A active site prompted us to test whether binding to IpdAB variants in solution perturbed the absorption spectrum of COCHEA-CoA. Titration of the E105^{A}A variant with COCHEA-CoA yielded a stable yellow colored species, ES_{yellow} (λ_{max} = 312 nm; Figure 5.7 inset) consistent with an enolate. A dissociation constant (K_D) for COCHEA-CoA of 0.4 ± 0.2 μM was calculated from the titration data. This species was not observed in WT $IpdAB_{RHA1}$ nor either the Arg92^{B} or Arg126^{B} variantsIpdAB catalyzed deuteration of COCHEA-CoA

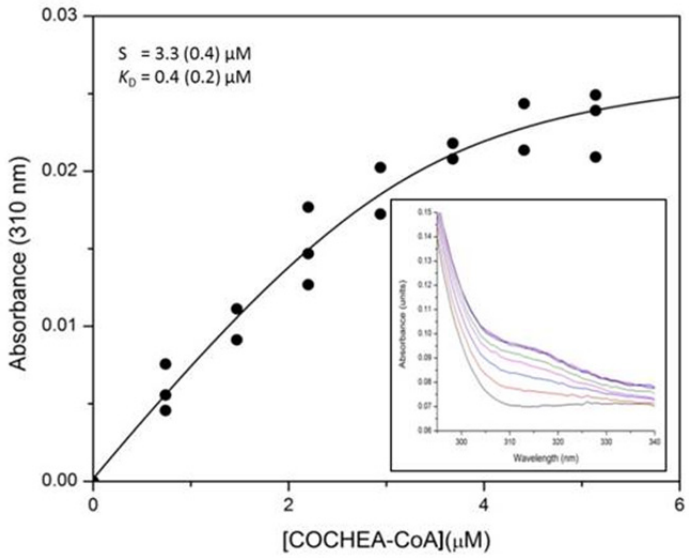


Figure 5.7 IpdAB E105A stabilizes a yellow coloured species.
Shown is the increase in absorbance at 310 nm upon titration of COCHEA-CoA into 3.5 μM $IpdAB_{RHA1}$ E105^{A}A. The curve represents the best fit of the quadratic binding equation. Insert displays spectra recorded at 0 μM (grey) up to 5 μM (purple) COCHEA-CoA.

5.3.10 IpdAB catalyzed deuteration of COCHEA-CoA

Although $IpdAB_{RHA1}$ did not detectably transform COCHEA-CoA ($[M+H]^+$ = 976.196 Da) in the absence of FadA6, the observation of ES_{yellow} in the $E105^AA$ variant suggested that the enzyme might catalyze a reaction that is not detected using LC-MS, such as the rapid interchange between COCHEA-CoA and the ring-opened MeDODA (Figure 5.2F). We hypothesized that such a reaction would be detected by following deuterium incorporation into COCHEA-CoA from deuterium oxide (D_2O). Indeed, incubation of IpdAB and COCHEA-CoA in the presence of D_2O resulted in the formation of a new species with $[M+H]^+$ = 979.216 Da (Figure 5.9B), consistent with deuteration at three positions. This species had the same HPLC retention time as COCHEA-CoA, consistent with the two compounds being structurally identical (data not shown). COCHEA-CoA was not deuterated in the absence of IpdAB (Figure 5.9B), nor was the incorporated deuterium exchanged out when the 979 species was incubated in water (data not shown). The activity appears to be specific to COCHEA-CoA as $IpdAB_{RHA1}$ did not catalyze the deuteration of 5αOH-HIC-CoA or any other CoA thioester tested. MS^3 indicated that all three sites of deuteration were located on the acyl moiety as the characteristic 428 *m/z* fragment ion generated from CoA was observed in both species (Figure 5.9C). In contrast, a diagnostic 181.0865 *m/z* fragment ion assigned as resulting from cleavage of the carboxy thioester in COCHEA-CoA was shifted to 184.1014 *m/z* in ^{2}[H]-COCHEA-CoA (Figure 5.9C). Incubation of $IpdAB_{RHA1}$ and FadA6 with COCHEA-CoA and CoASH in buffered D_2O yielded $^2[H]_3$-MOODA-CoA (*m/z* = 955.2172 Da; Figure 5.9D) and $^2[H]_3$-acetyl-CoA (*m/z* = 813.1544 Da; Figure 5.9E). $IpdAB_{RHA1}$ $E105A^A$ did not catalyze deuteration of COCHEA-CoA (Figure 5.9F).

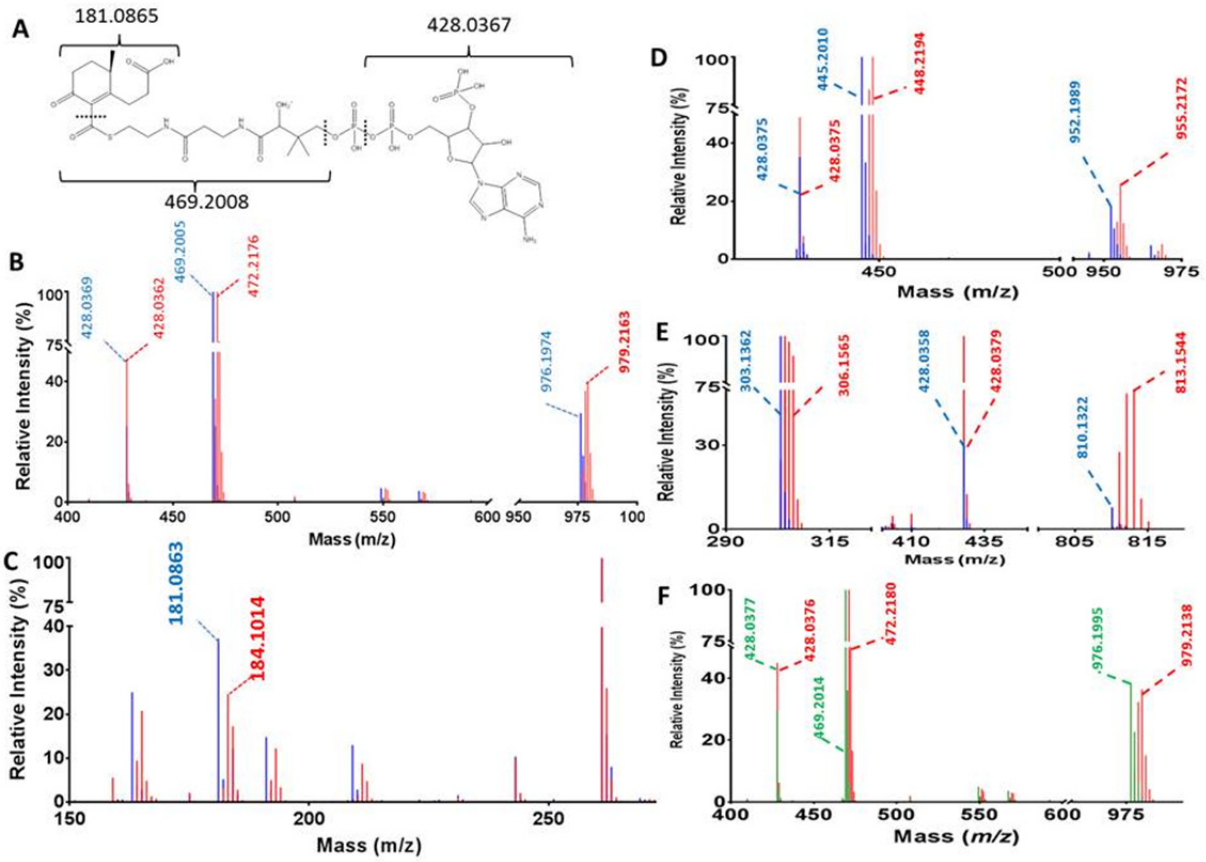


Figure 5.8 IpdAB catalyzes proton exchange on COCHEA-CoA
(**A**) Diagram depicting predicted fragments from MS^3 on COCHEA-CoA. 100 µM COCHEA-CoA was incubated with (red) or without (blue) 2.5 µM $IpdAB_{RHA1}$ in D_2O. Shown are the respective mass spectra for the LC-MS/MS of COCHEA-CoA (**B**); LC-MS^3 (150-275 *m/z*) of COCHEA-CoA (**C**); or LC-MS/MS of MOODA-CoA (**D**) and acetyl-CoA (**E**) when 5 and 150 µM of FadA6 and CoASH, respectively were added . (**E**) LC-MS/MS mass spectra of COCHEA-CoA following incubation with $IpdAB_{RHA1}$ WT (red) or $E105^AA$ (green) in D_2O.

To determine the rate and location of the IpdAB-catalyzed deuteration, the reaction was followed using NMR. A time-dependent decrease in the integration of the C-11 methyl doublet of COCHEA-CoA was observed (Figure 5.10A) indicating exchange at C-5. This loss was not observed in the absence of IpdAB. The initial rate of deuteration at C-5 was calculated using Equation 3 (Figure 5.10B) at 40-50 s^{-1}, although the fit to data beyond 20 minutes was poor, likely due to the competition of ^{2}[H]-COCHEA-CoA for IpdAB. Monitoring the integration of a singlet from a methyl group in the CoA moiety confirmed that COCHEA-CoA did not

significantly degrade during this experiment (Figure 5.10, grey). A concomitant decrease in the C-5 proton signal at 2.78 ppm was observed yielding highly similar results when used for measuring the rate of exchange (data not shown). A time-dependent decrease in proton integration in the 2.4-2.6 ppm range was also observed, consistent with deuteration at C-3. However, LC-MS analysis of the COCHEA-CoA pre- and post-incubation with $IpdAB_{RHA1}$ indicated that M + 1 ^{2}[H] was the most abundant species, suggesting that the rate of exchange at C-3 is significantly slower than at C-5 (data not shown).

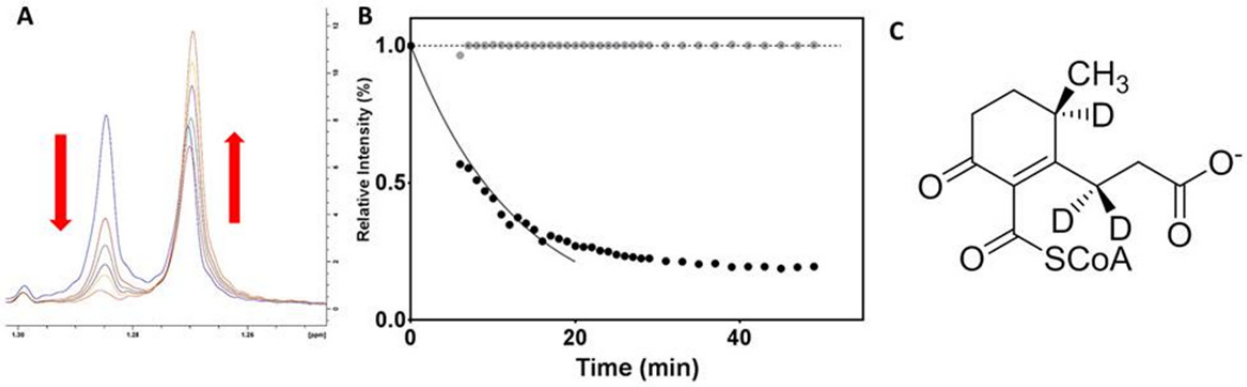


Figure 5.9 NMR characterization of IpdAB catalyzed deuterium incorporation
(**A**) ^{1}H-NMR spectrum displaying the time dependent decrease in the C-10 methyl doublet of COCHEA-CoA upon incubation of $IpdAB_{RHA1}$ in D_2O. Blue, red, green, purple, yellow and orange curves depicts before addition of $IpdAB_{RHA1}$ and 6, 11, 21, and 52 minutes after addition, respectively. (**B**) Time course displaying the relative decrease in proton integration for the C-10 methyl doublet (1.28 ppm; black dots) and the C-36 methyl group (0.75 ppm; grey dots) in the CoA moiety of COCHEA-CoA following addition of $IpdAB_{RHA1.}$ Solid curve displays the best fit of Eq. 2 to the initial 20 min of data. Dotted line indicates relative intensity = 1.0 (**C**) Location of deuteration (D) on COCHEA-CoA as determined by NMR.

5.3.11 ^{18}O is not incorporated into COCHEA-CoA or IpdAB

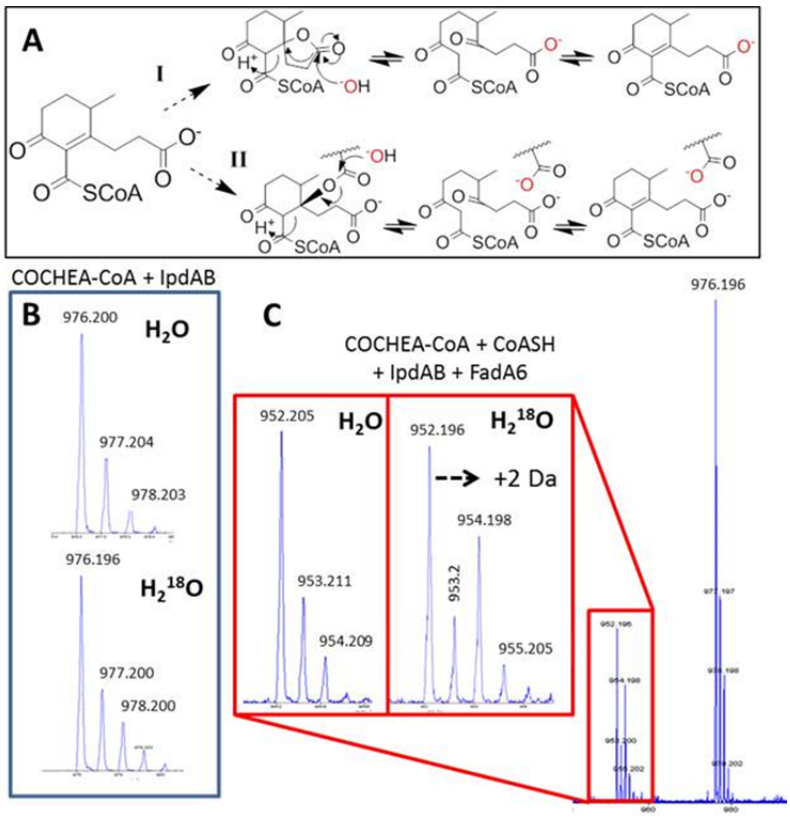


Figure 5.10 Additional $NaBH_4$ and ^{18}O experimental data
(**A**) Diagram of proposed IpdAB reaction mechanisms (I and II). Red atom indicates proposed location of ^{18}O from $H_2{}^{18}O$.^{18}O incorporation into COCHEA-CoA. LC- MS mass spectra of 100 µM COCHEA-CoA following incubation with 5 µM IpdAB (**B**) or (**C**) 2 µM IpdAB, 20 µM FadA6, and 125 µM CoASH in 10 mM sodium phosphate pH 8.0 prepared in H_2O or 97% (^{18}O abundance) $H_2{}^{18}O$. Red box indicates mass spectra of MOODA-CoA.

To test whether the lactonized species observed in the IpdAB·COCHEA-CoA complexes is catalytically relevant, we incubated $IpdAB_{RHA1}$ and COCHEA-COA in the presence of >90% $H_2{}^{18}O$ and assayed for ^{18}O incorporation using LC-MS. More specifically, a retro-Claisen like ring-opening reaction could be facilitated by hydrolysis of the lactone (Reaction I, Figure 5.11A), hydrolysis of a glutamyl ester (Reaction II, Figure 5.11A), or direct hydroxylation at C-4. Hydrolysis of the lactone in Reaction I should result in ^{18}O incorporation at C-1 in the presence of $H_2{}^{18}O$. We did not detect any ^{18}O exchange into COCHEA-CoA using LC-MS (Figure 5.11B). Moreover, intact protein LC-MS indicated that ^{18}O was not detectably

incorporated into IpdAB (data not shown) as would be expected for the hydrolysis of the resulting acyl-enzyme ester linkage. Considering that carboxylates are not readily exchangeable [228, 229], these results indicated that neither the lactonized COCHEA-CoA observed in the crystal structures nor a Glu105^{A}–ester are catalytically relevant. As a positive control, incubation of IpdAB, FadA6, COCHEA-CoA, and CoASH produced MOODA-CoA with a mass of 954.2 Da, consistent with the incorporation of ^{18}O at C-4 (Figure 5.11C).

Although ^{18}O incorporation into MOODA-CoA was observed (Figure 5.11C), ^{16}O - ^{18}O exchange occurred rapidly upon dilution into H_2O + 0.1% formic acid prior to LC-MS. Therefore, we were unable to conclude whether ^{18}O incorporation occurred from the hydrolysis of COCHEA-CoA or oxygen exchange at the C-4 ketone of MOODA-CoA after its production [230].

Chapter 6: **Discussion**

This thesis characterizes aspects of the catabolism of the latter part of the cholesterol molecule by *Mtb*. This work provides insights into cholesterol catabolism in related bacteria and, more generally, microbial steroid catabolism. These studies: (A) establish the order of cholesterol side chain and Rings A/B degradation in Actinobacteria (Chapter 2); (B) provide a structural and biochemical model for the HIP-CoA mediated regulation of the KstR2 regulon, present in steroid-degrading Actinobacteria (Chapter 3); (C) elucidate the catabolic steps involved in steroid Rings C and D opening in all known steroid-degrading bacteria (Chapter 4); and (D) characterize IpdAB, a virulence factor with a novel catalytic function for a member of the CoA Transferase superfamily (Chapter 5). More generally, each chapter demonstrates how CoA thioesters play a vital role in cholesterol catabolism and how improved methods to observe CoA thioesters could advance analyses of bacterial catabolic pathways.

6.1 Cholesterol Rings A and B are degraded prior to the alkyl side chain

The data presented in Chapter 2 establish that in the catabolism of cholesterol by *Mtb* and other Actinobacteria, complete Rings A/B degradation precedes that of the alkyl side chain. Disruption of the genes encoding the last two Rings A/B-degrading enzymes, *hsaC* and *hsaD*, in RHA1 led to the accumulation of metabolites with incompletely degraded side chains when the strains were incubated with cholesterol. Similarly, both $HsaD_{Mtb}$ and $HsaD_{Mab}$ had higher specificity (k_{cat}/K_M values) for substrates with partially degraded side chains. These data extend the findings of Capyk *et al.* (2011) and Casabon *et al.* (2013) in which the alkyl side chain was shown to be degraded somewhere after the KshAB and before the FadD3 reactions, respectively, requiring updates to the cholesterol catabolic pathway [69, 79].

6.1.1 Steroid degradation pathways in *Mab* appear to share a single HsaD

Mab grew on both cholesterol and 4-AD, consistent with containing gene clusters responsible for the catabolism of each steroid. Growth on cholesterol was very similar to what has been reported in *Mtb* and other Actinobacteria [51, 79, 187]. However, *Mab* grew to much lower yields on 4-AD, in contrast to what was reported in *M. smegmatis* [27]. Moreover, although the two strains contain similar 4-AD catabolic genes, apparently under control of a PadR-family transcriptional repressor, growth of *M. smegmatis* on 4-AD was only observed in strains lacking either this repressor or the KstR regulator [27]. The effector of the PadR repressor remains unknown: the 4-AD catabolic genes appear to be constitutively expressed in *M. smegmatis* and were not up-regulated in the presence of either 4-AD or cholesterol [27]. Moreover, the mechanism of uptake of 4-AD by *M. smegmatis* and *Mab* is unclear as actinobacterial Mce4 transporters are specific to steroids with an alkyl side chain [54]. Interestingly, despite the absence of dedicated 4-AD catabolic genes, *Mtb* also appears to grow on this steroid [51].

A unique feature of the cholesterol and 4-AD catabolic pathways of *Mab* is that they likely share a single HsaD homolog. All other steroid catabolic clusters characterized to date in Actinobacteria have a complete set of Rings A/B-degrading enzymes [22, 25, 27]. However, the steady-state kinetic data establish that $HsaD_{Mab}$ transforms both DSHBNC-CoA and DSHA, respectively, with k_{cat}/K_M values similar to those of other MCP hydrolases for their respective physiologically substrates [156, 189]. It is unclear whether the substrate specificity of $HsaD_{Mab}$ reflects that of all Subfamily V MCP hydrolases.

The data do not exclude the possibility that Mab_4366c catalyzes the hydrolysis of steroid-derived MCPs in *M. abscessus*. Although Mab_4366c occurs within a cluster of genes

predicted to encode the catabolism of isopropyl benzene, its phylogeny differs from characterized IpbDs/CumDs (Figure 2.1) [191, 192]. The unique phylogeny of Mab_4366c appears to relate to two regions that share low amino acid identity with characterized MCP hydrolases, Pro76-Val91 and Ala136-Pro163. Based on modeling studies, the latter contributes to the "non-conserved loop" and "non-polar" (NP) subsite in MCP hydrolases [162]. The NP subsite is a major determinant of substrate specificity. This suggests that Mab_4366c may have divergently evolved from a Subfamily IV ancestor to accommodate non-steroidal substrates. Steady-state kinetic analyses of Mab_4366c with DHSA, DSHBNC-CoA and 2-hydroxy-6-oxo-7-methylocta-2,4-dienoate would help evaluate the enzyme's role in steroid catabolism. Similarly, deletion mutants of *hsaD*$_{Mab}$ and *mab_4366c* would be invaluable in validating their respective roles.

6.1.2 HsaD$_{Mtb}$ and HsaD$_{Mab}$ have highest specificity for steroid substrates with incompletely degraded side chains

The specificity of HsaD$_{Mtb}$ for DSHBNC-CoA over DSHA (k_{cat}/K_M values) is greater than in HsaD$_{Mab}$. This is consistent with what has been reported for other cholesterol Rings A/B-degrading enzymes, such as KshA from *Mtb* and KshA1 from *Rhodococcus rhodochrous* [41, 69] as well as the extremely low substrate specificity of HsaAB for a substrate with a completely degraded side chain [76]. Interestingly, in the crystal structure of HsaD$_{Mtb}$·DSHA (PDB# 2WUF), the 17 oxo of DSHA is orientated towards an atypically large, polar pocket at the entrance to the enzyme's active site [77]. Moreover, the constellation of charged and polar residues in this pocket is reminiscent of what has been reported in other CoA-binding enzymes [231]. Residues within this polar pocket are generally conserved across actinobacterial HsaDs, however not in HsaD$_{Mab.}$ Unfortunately, attempts to obtain a structure of HsaD$_{Mtb}$·DSHBNC-

CoA were unsuccessful (data not shown). Preliminary studies using molecular docking of adenine into this pocket implicated Ile144, Asp150, and Trp223 as being significant in enzyme-substrate binding (data not shown). Site-directed mutagenesis of these residues in the $HsaD_{Mtb}$ S114A variant displayed 5-10 nm bathochromic shifts in ES^{red} when titrated with DSHBNC-CoA suggesting perturbation of substrate binding. However, these mutations also destabilized the enzyme's structure making the titration data difficult to interpret (data not shown). Overall, the architecture of $HsaD_{Mtb}$ is consistent with it accommodating a CoA sidechain substrate, however identification of the exact determinants of CoA binding require additional data.

6.1.3 Updates to the pathway of cholesterol degradation

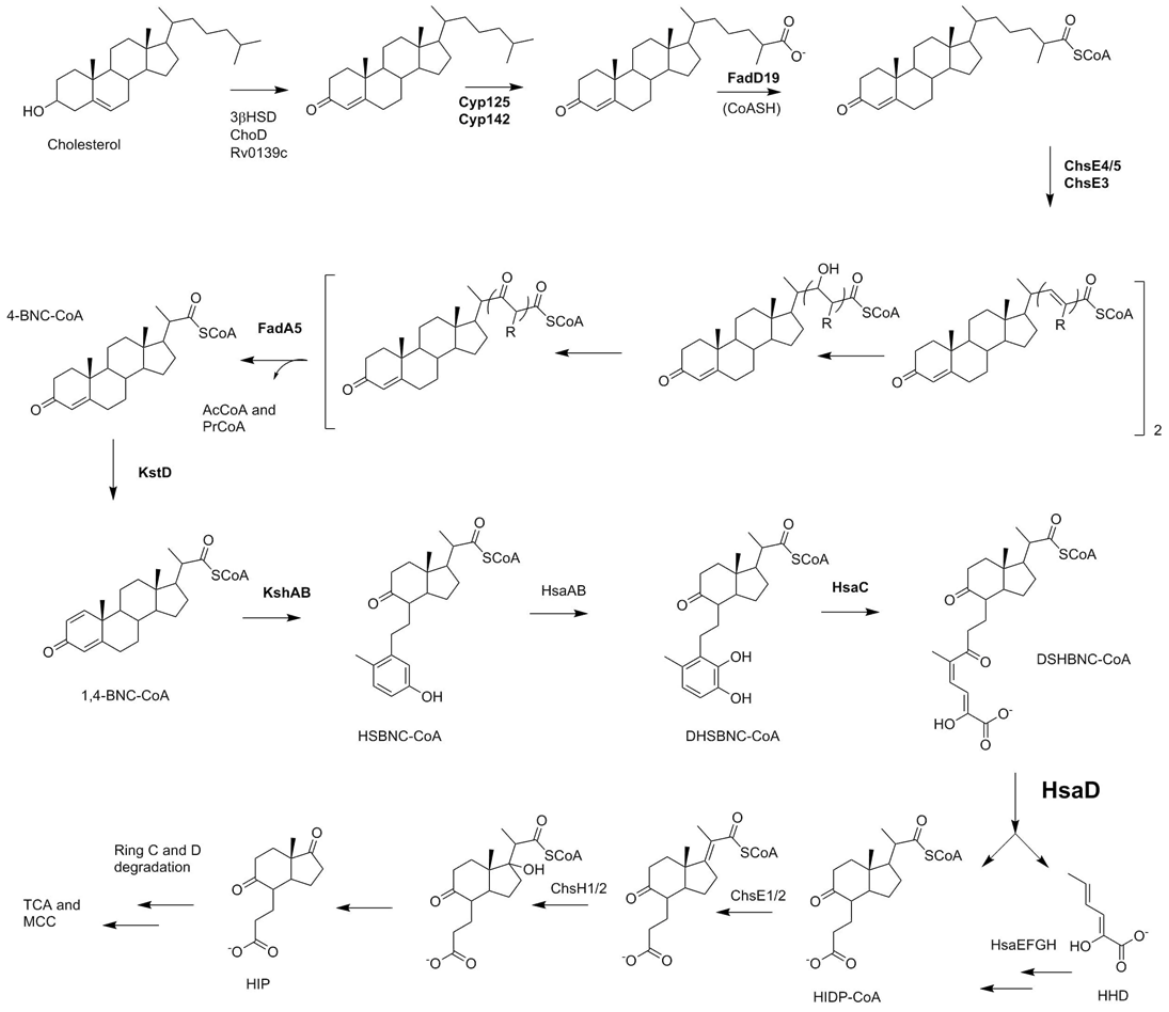


Figure 6.1 Proposed pathway of cholesterol catabolism in actinobacteria.
Enzymes for which *in vitro* activity exists with the substrate depicted are shown in bold. Substrates of Rings A/B catabolic enzymes shown with isopropionyl-CoA moiety. However, the isopentanoyl-CoA or isooctanoyl-CoA analogs cannot be excluded as physiological substrates.

An update to the pathway for cholesterol degradation is proposed (Figure 6.1). Following cholesterol uptake [33], C-3 dehydrogenation [72, 232], C-27 oxidation [58], and CoA thioesterification at C-27 [179] the resulting, 3-oxo-4-cholestenoyl-CoA is degraded via two rounds of β-oxidation involving ChsE4-ChsE5, ChsE3, and FadA5, yielding a propionyl-CoA and an acetyl-CoA molecule [47, 63, 67, 233]. Prior to removal of the C-17 isopropionyl-CoA

moiety, KstD forms 1,4-BNC-CoA and cholesterol Ring A is aromatized by KshAB yielding 3-hydroxy-9-oxo-9,10-seco-23,24-bisnorchola-1,3,5(10)-trien-22-oyl-CoA (HSBNC-CoA) [69]. HsaAB presumably hydroxylates HSBNC-CoA to DHSBNC-CoA which undergoes *meta*-cleavage by HsaC forming DSHBNC-CoA [76]. DSHBNC-CoA is cleaved into HHD and 3aα-H-4α(3'-propanoate)-7aβ-methylhexahydro-5-indanone-1-propionyl-CoA (HIDP-CoA) where side-chain degradation continues likely involving ChsE1-ChsE2 and ChsH1-ChsH2 [63, 67]. Finally, a currently unidentified thiolase or aldolase removes the final propionyl-CoA moiety generating HIP which feeds into Rings C and D degradation.

Consistent with the updated pathway, deletion of the intracellular growth (*igr*) operon in *Mtb*, containing *chsE1, chsE2, chsH1, chsH2* and *ltp2*, yields extracellular HIDP accumulation in the presence of cholesterol [48, 66]. Furthermore, HSBNC accumulates when *Mtb* is treated with an HsaA inhibitor in the presence of cholesterol [134].

6.2 HIP-CoA binding to KstR2 regulates catabolism of cholesterol Rings C and D

Chapter 3 provides the first molecular insights into the KstR2-mediated regulation of the expression of steroid catabolic genes in *Mtb* and other actinobacteria. The structural and titration data establish that the KstR2 dimer binds 2 molecules of HIP-CoA. Each HIP-CoA molecule binds in a deep cleft that spans the KstR2 dimer with the adenosine and HIP moieties bound by separate protomers. The extensive electrostatic and hydrophobic interactions that mediate the binding of HIP-CoA to $KstR2_{Mtb}$ are corroborated by the high affinity of the regulator for its effector molecule. The functional significance of the $KstR2_{Mtb}$·HIP-CoA structure was further validated by directed mutagenesis, which established that Arg162 and Trp166 contribute significantly to the binding of HIP-CoA while minimally affecting the binding of the operator DNA. Finally, comparison of the $KstR2_{Mtb}$·HIP-CoA, $KstR2_{RHA1}$ and SlmA·DNA structures

suggests how effector binding alters the conformation of the regulator to relieve binding of the operator DNA.

6.2.1 Comparisons to TetR family of transcriptional repressors

The amino acid sequence conservation among KstR2 orthologs in steroid-degrading Actinobacteria further validates the functional importance of the residues identified in the $KstR2_{Mtb}$·HIP-CoA structure. These orthologs share ~50% amino acid sequence identity with higher conservation among DBD residues predicted to bind the operator DNA and EBD residues that bind HIP-CoA. Specifically, the residues located on helices α2 and 3 in $KstR2_{Mtb}$ are all conserved with the exception of Gly39. This is consistent with the conserved nucleotide sequence of the operator across Actinobacteria [83]. Similarly, 19 of 23 residues that contact the effector in the $KstR2_{Mtb}$·HIP-CoA complex, located between residues 138 to 195, are conserved in KstR2 orthologs.

Comparison of the structures of FadRs from *Thermus thermophilus* (3ANP, 3ANG; [88]) and *Bacillus subtilis* (1VI0; [234]) in complex with fatty acyl-CoA are strikingly similar to the $KstR2_{Mtb}$·HIP-CoA complex: in each, the ligand is bound in a cleft that spans the two protomers with the fatty acid and adenosine moieties bound to separate chains. Conserved residues include Arg159 and Arg173 (KstR2 numbering) that interact with the adenine and Arg162' which hydrogen bonds with the diphosphate moiety of CoA. Nevertheless, the adenine ring in the FadR·fatty acyl-CoA complexes is flipped 180^{o} with respect to that in the $KstR2_{Mtb}$·HIP-CoA complex and oriented perpendicular to the dimer's rotational axis. By contrast, the fatty acyl-CoA binds in a different way in the DesT complexes, with the CoA moiety at the top of the EBD and the acyl chain extending down a channel, parallel to the α-helices of the EBD [235]. Consistent with the different binding mode, none of the Arg residues are conserved in DesT. The

three conserved Arg residues are also not conserved in TFRs that bind smaller effectors, such as QacR of *Staphylococcus aureus* (PDB 1JT6), *Pseudomonas putida* TtgR (PDB 2UXI), and *Streptomyces coelicolor* ActR (PDB 3B6A).

Conservation of the CoA-binding residues in the KstR2 and FadR may be extended to other TFRs to gain insight into their respective effectors. Comparison of sequence topology maps of 48 structures of TFRs identified in Yu *et al.* (2010) indicate that 13 contain at least two of the three conserved basic residues in orientations permissive to CoA binding [165]. Of note, Fad35R of *M. tuberculosis* (PDB 4G12) regulates the expression of *fad35,* which encodes an acetyl-CoA synthetase involved in fatty acid degradation [236]. Although Fad35R binds tetracycline, other evidence suggested that a fatty acyl-CoA could be the physiological effector [236]. The presence of Lys184 on the α8-α9 loop, Arg166 on the apical end of helix α8, and His-170 in the middle of helix α8 supports this hypothesis. *T. thermophilus* HB8 PfmR (PDB 3VPR), which is predicted to regulate PAA or fatty acid catabolism possess basic residues on the α8-α9 loop and middle of helix α8 (Arg163 and Arg149, respectively) but lacks a basic residue at the top of helix α8. Although the respective effectors of these 13 TFRs have yet to be identified, it appears that a significant subset of TFRs bind CoA thioesters.

The proposed mechanism of response of $KstR2_{Mtb}$ to HIP-CoA is similar to what has been proposed for other TFRs where the binding of a small molecule effector induces conformational changes that abrogate the regulator's interactions with its operator DNA [164, 165]. In the absence of a $KstR2_{Mtb}$·DNA structure, the comparative analysis between the $KstR2_{Mtb}$·HIP-CoA and SlmA·DNA complexes unveiled differences in the relative conformation of DBD domains in the TFRs that, in case of $KstR2_{Mtb}$, are likely triggered by HIP-CoA binding. More specifically, HIP-CoA binding to $KstR2_{Mtb}$ repositions helices α4 and α6 causing a net 15^{o} outward rotation

of DBD helix α1, displacing helices α2 and α3 compared to corresponding elements in SlmA·DNA complex (Figure 3.4). Interestingly, Tyr48 and His50 on DBD helix α3 in $KstR2_{Mtb}$·HIP-CoA are rotated outward by ~10 Å as compared to the ligand-free structure of $KstR2_{RHA1}$ (Figure 3.3). These residues are highly conserved in TFRs and, in *E. coli* TetR, directly interact with the bases in the major groove of the DNA [165]. Lastly, Lys54, which is also highly conserved in TFRs [165], is displaced by ~7 Å in $KstR2_{Mtb}$·HIP-CoA compared to $KstR2_{RHA1}$. The effect of HIP-CoA on KstR2 is also remarkably similar to the displacement of helix α4 and outward rotation of helix α3 observed in apo- *vs*. ligand bound *B. subtilis* FadR [87]. In TetR, effector binding thermodynamically stabilizes the DBD in a conformation that is incompatible with DNA binding, preventing the DBD from assuming a conformation that is competent for DNA binding [237]. The inability of $KstR2_{Mtb}$·HIP-CoA to bind DNA likely has the same mechanistic origin.

The binding of two $KstR2_{Mtb}$ dimers to its operator was somewhat unexpected considering that the KstR box of 14 bp [83] is too short to accommodate four recognition helices in the major groove. Typically the TFR operators that bind two dimers are at least 22 bp in length [205, 238-240]. By contrast, TFRs that bind operator sequences of less than 17 bp bind as a single dimer [239]. It is possible that the KstR2 box extends beyond the 14 bp identified by Kendall *et al.* [83]. Alternatively, the KstR2 dimers may bind opposite sides of the DNA helix, as in the SlmA·DNA complex (PDB: 4CGT). In such a scenario, the first KstR dimer induces a conformational change in the DNA that facilitates the binding of the second dimer to a non-canonical sequence. Intriguingly, $KstR2_{Mtb}$ contains the Arg-X-Thr motif that is present in SlmA and induces a kink in the DNA [205]. Similar models have been invoked to explain cooperative binding in each of two other TFRs, QacR and CprB [239, 240], although QacR lacks the Arg

residue and does not induce a kink in the DNA. Interestingly, single KstR2 boxes occur between divergently transcribed promoters [83] and gel shift assays indicate that KstR2 binds to each of the three boxes with the same stoichiometry [91]. The binding of two KstR2 dimers to opposite sides of the DNA helix may enable the repressor to act at both promoters. Additional experiments and structural data are required to determine the precise architecture of the $KstR2_{Mtb}$·operator complexes and how this regulates divergently transcribed promoters.

6.2.2 Insights into the inducer of KstR

The KstR2·HIP-CoA model predicted that the effector of *M. tuberculosis* KstR (PDB# 3MNL) is also a CoA thioester. $KstR2_{Mtb}$ and $KstR_{Mtb}$ share only 18% amino acid sequence identity. However, the conserved residues include four that mediate CoA binding in KstR2 (Arg162, Asp163, Trp171 and Arg173) and two that mediate steroid binding (Trp166 and Phe70). Since the publication of Chapter 3, 3-oxo-cholest-4-en-26-oyl-CoA was confirmed to be the effector of $KstR_{Mtb}$ in collaboration with the Lott Lab at the University of Auckland [86]. Therefore, the effector of $KstR_{Mtb}$ mirrors the regulatory logic of the KstR2 regulon in *M. tuberculosis* in that the catabolic genes are induced by the first CoA thioester catabolite produced. More particularly, C-27 of cholesterol's alkyl side chain is oxidized to a carboxylate by either Cyp125 or 142 [58], and thioesterified by FadD19 [179]. Subsequent structural and biochemical analysis demonstrated a mechanism of gene repression analogous to that described for KstR2 by HIP-CoA. Overall, the identification of the effector of KstR demonstrates how structural features of $KstR2_{Mtb}$·HIP-CoA can be applied to guide the identification of physiological effectors for other TetR family transcriptional repressors.

6.3 Elucidation of the catabolism of cholesterol Rings C and D

Until this work, the pathway by which the last two steroid rings, contained in HIP, are degraded in bacteria was unknown. By profiling the culture supernatants and CoA metabolomes of deletion mutants of HIP catabolic genes, *ipdAB*, *ipdC*, *ipdF*, *echA20*, and *fadE32* in *Mtb, M. smegmatis,* and/or RHA1, three characteristic HIP catabolites were identified: 5αOH-HIC-CoA, COCHEA-CoA, and MOODA. Enzymological experiments further identified HIEC-CoA and MOODA-CoA. When viewed together, the mutant data, enzymological transformations, and bioinformatics analyses, elaborate a pathway for HIP catabolism.

6.3.1 The catabolic pathway of cholesterol Rings C and D in actinobacteria

A model for HIP degradation was proposed in which Ring D cleavage precedes that of Ring C (Figure 6.2). In the proposed pathway, the isopropionyl side chain of HIP is first degraded via β-oxidation to yield 5-OH HIC-CoA. This is then transformed to HIEC-CoA by IpdF and IpdC, before undergoing two successive ring-cleavage reactions: EchA20-catalyzed hydrolysis of Ring D followed by IpdAB-catalyzed hydrolysis of Ring C. Thiolysis of the Ring C-opened product, potentially by FadA6 or another thiolase, yields MOODA-CoA which is then oxidized to $^{2}\Delta$-MOODA-CoA by an ACAD comprised in whole or in part by FadE32. Although the fate of $^{2}\Delta$-MOODA-CoA is unclear, a final round of β-oxidation to yield 2-methyl-β-ketoadipyl-CoA (MβKA-CoA) is proposed. This could then be cleaved to propionyl-CoA and succinyl-CoA in a manner analogous to the cleavage of β-ketoadipyl-CoA to succinyl-CoA and acetyl-CoA in the final step of the β-ketoadipate pathway used in the bacterial catabolism of aromatic compounds [241]. While several aspects of the HIP pathway have yet to be elucidated, the data support the proposed physiological roles of four enzymes: IpdF, IpdC, EchA20, and IpdAB.

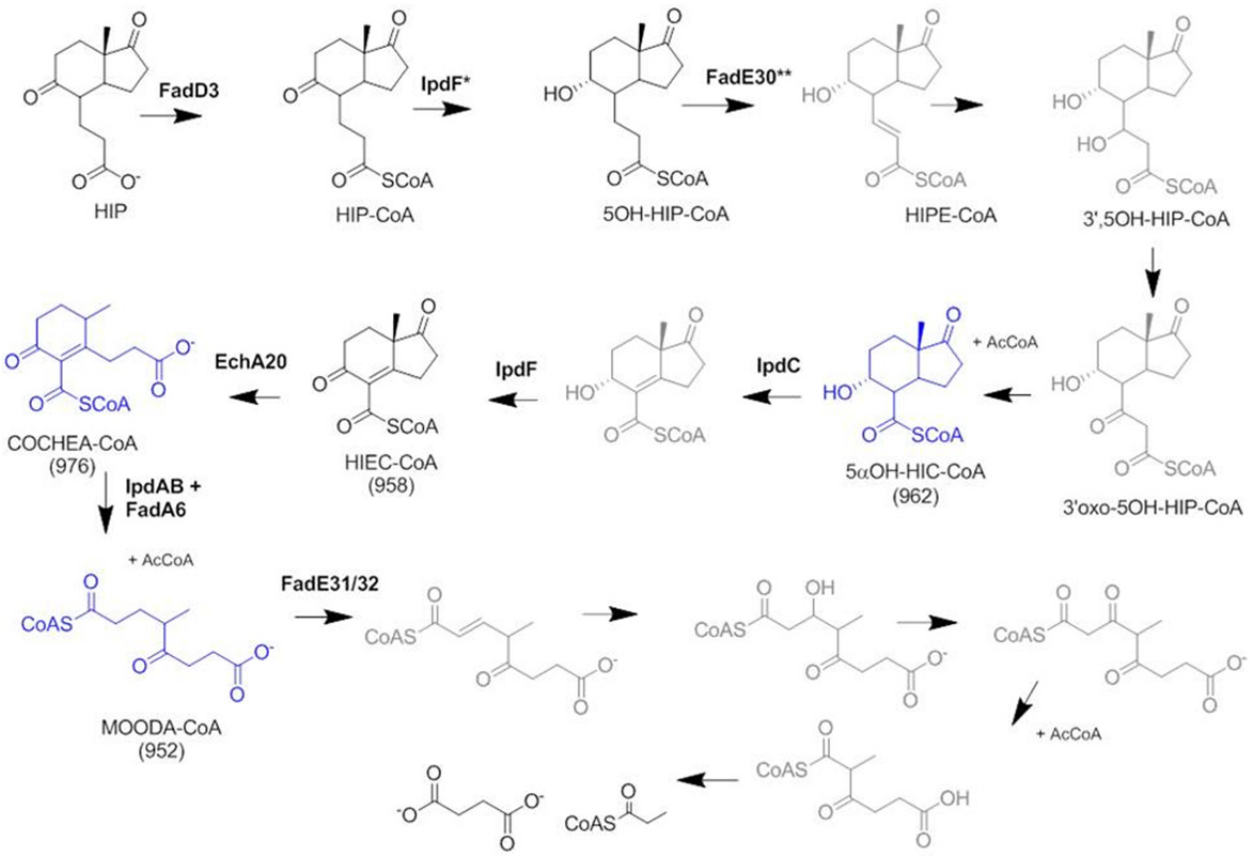


Figure 6.2 Proposed HIP catabolic pathway.
NMR-confirmed metabolites are in blue. Metabolites for which MS data were obtained are in black. Other metabolites are in grey. *The current study established that IpdF has this activity, but its physiological relevance is unclear. **Role of FadE30 assigned previously [23]

The proposed pathway provides an important framework to further characterize various aspects of steroid metabolism, including the identity of specific metabolites, such as HIEC-CoA and MβKA-CoA, as well as enzymatic steps, such as those catalyzed by Rv3548c and Rv3549c, encoded by the KstR2 regulon. The model further suggests the identities of the *fadE*-encoded ACADs that act on HIP-CoA and MOODA-CoA, respectively. Another unknown aspect of the pathway is the significance of the IpdF-catalyzed reaction: it is unclear why the 5-oxo group would be reduced and then reoxidized. Anecdotally, chemical synthesis of HIC-CoA proved difficult due to the propensity of HIC-CoA to undergo acid catalyzed ring opening whereas 5OH-HIC-CoA was stable (Hurst, unpublished). Finally, this pathway also predicts that cholesterol feeds into central metabolism via four propionyl-CoAs, four acetyl-CoAs, one

pyruvate and one succinyl-CoA. Notably, propionyl-CoA, a potentially toxic metabolite [49], is derived from all three parts of cholesterol: the side chain, Rings A/B and Rings C/D.

6.4 IpdAB is a ring-cleaving hydrolase

Among the enzymes whose functions were assigned, that of IpdAB was unexpected based on the bioinformatic analyses (Table 6.2). More specifically, no Class I CoA transferase has been reported to catalyze a retro-aldol hydrolysis**.** This study provides evidence that IpdAB transforms COCHEA-CoA, thereby catalyzing the hydrolytic opening of the last ring in the bacterial catabolism of steroids. The data further indicate that this reaction does not involve CoA transfer despite IpdAB's striking similarity to Class I CoTs. First, $IpdAB_{RHA1}$ did not catalyze the transfer of CoA from small acyl-CoA donors to small acids as is typical for Class I CoT. Second, the IpdAB active site lacks key elements that are conserved in CoTs including the catalytically essential glutamate and an oxyanion hole to accommodate the CoA thioester [175]. Third, $NaBH_4$ did not detectably inhibit IpdAB, nor did it trap a glutamyl-CoA thioester intermediate. Finally, the position of the CoA in MOODA-CoA, the product of the $IpdAB_{RHA1}$/FadA6 reaction, indicates that the CoA isn't transferred from one carbon to another prior to thiolysis. Therefore, the data support IpdAB as transforming COCHEA-CoA into 6-methyl-3,7-dioxodecanedioyl-CoA (MeDODA-CoA).

6.4.1 Proposed mechanism of IpdAB

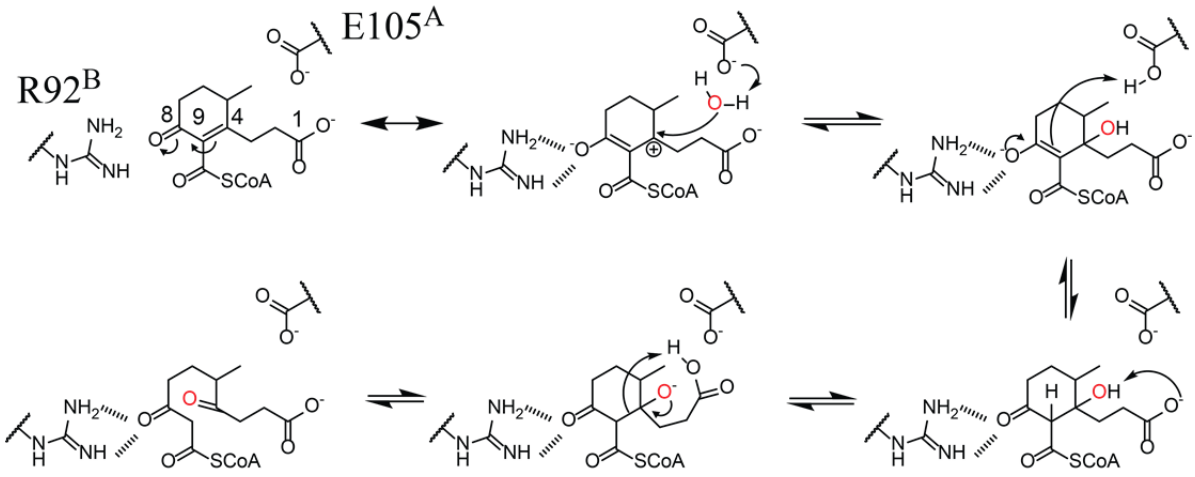


Figure 6.3 Proposed mechanism of IpdAB.
Location of O atom from H_2O is identified in red.

Based on the presented data, a mechanism for IpdAB involving a retro-Claisen like hydrolysis (Figure 6.3) is proposed. In this mechanism, binding of COCHEA-CoA to the IpdAB active site stabilizes the C-8 enolate resonant form of COCHEA-CoA via hydrogen bonding from $Arg92^B$ and $Asn131^B$. A solvent species, activated by either $Glu105^A$ or the C-1 carboxylate, attacks at C-4, which is more electrophilic due to stabilization of the enolate. Tautomerization of the C-8 enolate protonates C-9. Deprotonation of the resulting C-4 hydroxyl yields an alkoxide anion, whose ketonization and concerted protonation at C-9 permits the cleavage of the C-4/C-9 bond and Ring C opening.

The proposed mechanism is supported by several lines of evidence. First, the $E105^AA$ variant stabilizes a species whose spectrum is consistent with a CoA-enolate, as exemplified by the acetoacetyl-CoA-Mg^{2+} complex (λ_{max} = 310 nm [242]). The structure of the IpdAB·COCHEA-CoA complex suggests that $Arg92^B$ stabilizes the C-8 enolate. Second, deuteration at C-5 and C-3 is consistent with the increased electrophilicity of C-4. The failure of the $E105A^A$ variant to catalyze deuteration of COCHEA-CoA indicates that deuteration occurs

concerted with or after hydroxylation at C-4. Further, the rate of deuteration is an order of magnitude faster than COCHEA-CoA turnover and is thus not rate-limiting. Finally, the lack of ^{18}O incorporation at C-1 indicates that the lactonized COCHEA-CoA is not catalytically relevant although the nucleophile could be activated by the C-1 carboxylate.

Several aspects of the mechanism remain to be elucidated. First, although the proposed mechanism posits $Glu105^{A}$ as the base that activates the nucleophile, the structural data indicate that both this residue and the C-1 carboxylate may be positioned to activate water and/or deprotonate the C-4 hydroxyl. Nevertheless, it is unclear whether the C-1 carboxylate is catalytically essential. Additional mechanistic studies involving the substrate were complicated due to limited quantities of COCHEA-CoA and the inability to directly observe its hydrolysis product despite various attempts to chemically trap MeDODA-CoA (data not shown). It is possible that MeDODA-CoA was not observed because the equilibrium favours COCHEA-CoA and thiolysis of MeDODA-CoA drives the hydrolytic reaction forward. Indeed, the possibility that IpdAB catalyzes an intramolecular CoA transfer from C-11 of COCHEA-CoA to C-11 in MeDODA-CoA cannot be excluded. However, this is not supported by the results of the CoA transferase studies and is thermodynamically unfavourable as formation of a carboxylic acid or ester would increase the pK_a of hydrogens at C-9, increasing the activation energy for C-4/C-9 hydrolysis.

6.5 Broader implications from the elucidation of HIP catabolism

6.5.1 The HIP catabolic pathway in other bacteria

The occurrence of a single KstR2 regulon in Actinobacteria in strains that contain several distinct steroid catabolic pathways [22, 25] suggests that hydroxylated HIP-CoA can act as the effector of at least some KstR2 orthologs. For example, RHA1 possesses at least three distinct

pathways that converge on the HIP catabolic pathway, two of which are responsible for cholesterol and bile acid catabolism, respectively [22]. The catabolism of bile acids such as cholate results in the production of 3'OH HIP [79], suggesting that 3'OH HIP-CoA would be an effector of $KstR2_{RHA1}$. Inspection of the $KstR2_{Mtb}$·HIP-CoA structure indicates that there is sufficient space adjacent to C-3' of the HIP moiety to accommodate a hydroxyl group. Moreover, all of the residues within a radius of 6 Å of C3' are conserved in $KstR2_{RHA1}$ (Figure 3.3). By contrast, the binding of 7β-OH HIP-CoA is predicted to be sterically hindered by Tyr-108 (Tyr-112 in $KstR2_{RHA1}$). This is consistent with the finding that C-12 hydroxyl groups of bile acids (corresponding to C-7 of HIP) are removed prior to Rings C/D degradation in Actinobacteria [28, 79]. By contrast, 7β-OH HIP is produced in steroid-degrading Gram negative bacteria such as *Pseudomonas putida* DOC21 [21]. However, the HIP catabolic genes in these strains appear to be regulated by a LuxR-type transcriptional repressor [18, 21].

The HIP catabolic genes are conserved in steroid-degrading bacteria for which genome sequence data are available, suggesting that the pathway is not only employed in the degradation of steroids other than cholesterol [18, 22, 25], but also in the anaerobic degradation of steroids [20]. Indeed, differences in the HIP catabolic gene cluster in diverse bacteria appear to reflect the different steroid-catabolizing capabilities of the strains. For example, in bacteria that catabolize cholate or other bile acids, the HIP catabolic gene cluster contains *echA13* (RHA1_RS22405 in RHA1) [17, 22, 91]. A homolog of EchA20, EchA13 is proposed to remove the hydroxyl of 7β-OH HIP, generated from cholate degradation [21, 91], and is not present in *Mtb*, which does not degrade cholate. Similarly, the HIP catabolic gene cluster of *S. denitrificans* DSM 18526, which is up-regulated during the aerobic and anaerobic catabolism of testosterone [20], lacks a homolog

of *fadD3*/*stdA3*. Instead, this strain contains a homolog elsewhere in the genome: ACG33_09100 shares 53% amino acid sequence identity with StdA3 of *P. putida* DOC21 [21]. The genomic context of *fadD3* in *S. denitrificans* DSM 18526 may reflect the possibility that the β-oxidation of steroid Rings A and B yields HIP-CoA directly, obviating the need for FadD3 in anaerobic steroid catabolism.

6.5.2 Cleavage of MeDODA-CoA likely requires an unidentified thiolase

The low turnover rate of FadA6 in the coupled reactions indicates that it may not be the physiological thiolase responsible for MOODA-CoA formation. Indeed, $FadA5_{Mtb}$, which acts on the alkyl side chain of cholesterol [47], turned over MeDODA-CoA at the same rate as FadA6 in the coupled reaction (data not shown). FadA6 may be required in the cleavage of 3'oxo- 5-OH-HIP-CoA to 5αOH-HIC-CoA, upstream of COCHEA-CoA in the HIP catabolic pathway, and the enzyme acting on MeDODA-CoA may be encoded outside of the KstR2 regulon (Figure 6.2).

6.5.3 HIP catabolism displays similarities with anaerobic aromatic catabolism

A reoccurring theme in Rings C/D degradation is the similarities with bacterial aromatic catabolism. The hydrolytic cleavage of the cyclohexenone ring by IpdAB is similar to that catalyzed by Oah, a crotonase involved in the anaerobic degradation of benzoate in *Thauera aromatic* [243]. Oah hydrolyzes the cyclohexene ring of 6-oxocyclohex-1-ene-1 carbonyl-CoA, a structural analog of COCHEA-CoA, by hydroxylating a double bond. The ensuing retro-Claisen opening of the cyclohexanone ring is comparable to that of IpdAB (Figure 6.3) however, carbon-carbon cleavage occurs adjacent to the carbon double bond yielding 3-oxopimeloyl-CoA, a β-keto CoA thioester analogous to MeDODA-CoA [243]. Interestingly, the subsequent catabolism of 3-oxopimeloyl-CoA and MOODA-CoA are predicted to be similar [178, 244].

The phylogeny of *ipdAB* implies divergent evolution from ancestral genes encoding a Class I β-keto-CoA transferase to facilitate the opening of steroid Ring C. Interestingly, the degradation of HIP is predicted to yield methyl-β-ketoadipyl CoA (Figure 6.2). This metabolite is similar to β-ketoadipyl-CoA, the product of the Class I β-keto- CoA transferase, PcaJI, involved in benzoate degradation [245]. HIP catabolism is not predicted to require a β-ketoadipate CoA transferase like PcaJI. However, the subsequent catabolic step in both pathways involves a thiolase, and generates succinyl-CoA and either acetyl-CoA or propionyl-CoA and [245]. It is possible that duplication of the β-ketoadipate pathway may have facilitated the repurposing and divergent evolution of an ancestral PcaJI homolog to generate IpdAB.

6.6 Insights for *Mtb* pathogenicity and therapeutic development

6.6.1 Disruption of KstR2 genes yield a 'cholesterol-dependent-toxicity' in *Mtb*

Deletion of *ipdAB* and *ipdC* in *Mtb* yielded strains that failed to grow on glycerol in the presence of cholesterol and, in the case of *ipdAB*, significantly slowed the growth in THP-1 derived macrophages. These strains displayed distinct differences in the concentration and identity of CoA thioesters and CoASH. This suggests that 'cholesterol-dependent toxicity' of the mutants may be due to the sequestration of CoASH, making it unavailable for other cellular processes. Consistent with the sequestration hypothesis, strains which grow in the presence of cholesterol (*ΔfadD3* RHA1, *ΔfadE32 M. smegmatis*, and *ΔipdF M. smegmatis*) failed to accumulate significant levels of unique, cholesterol-derived CoA thioesters [79]. This implies that low intracellular CoASH concentrations in cholesterol catabolic gene mutants in *Mtb* may be predictive of cholesterol-dependent toxicity. The lack of toxicity in the *ΔipdF* and *ΔfadE32* mutants could reflect the ease of hydrolysis of the corresponding CoA thioesters (*e.g.*, MOODA-

CoA to MOODA) and/or the presence of compensatory enzymatic activities (*e.g.*, other cellular dehydrogenases could catalyze oxidation of the 5-OH HIC-CoA hydroxyl).

Disruption of cholesterol catabolic genes is reported to generate both attenuated and avirulent strains of *Mtb* due to the predicted accumulation of toxic cholesterol-derived metabolites [39, 47, 51]. For example, the reduced virulence of Δ*hsaC Mtb* in guinea pig infection models has been attributed to the accumulation of toxic catechols resulting from DHSA accumulation [39]. However, DHSBNC-CoA accumulation, implied by the data in Chapter 2, suggests CoA sequestration may also contribute to this attenuated phenotype. Interestingly, disruption of the ratio between acetyl-CoA and propionyl-CoA in *Mtb* during growth on cholesterol has been reported to result in a toxic phenotype [112]. Indeed, reduction in CoASH in strains displaying cholesterol-dependent toxicity typically coincided with reduced acetyl-CoA levels in the CoA metabolic data (data not shown). Although additional data are required to test the CoASH sequestration hypothesis, the data rationalizes the effectiveness of the *ipdAB* mutant as a vaccine in *R. equi* and indicates which HIP degradation enzymes would be good targets for future therapeutic development.

6.6.2 Therapeutics targeting *Mab*

The role of steroid catabolism in *Mab* infection is unclear. Resolving this question is complicated by a scarcity of effective genetic tools and infection models [246]. *Mab* was recently shown to replicate in bone marrow-derived murine macrophages [247]. However, it has not yet been established whether *Mab* is an obligate intracellular pathogen. Developing cholesterol catabolism as a target for novel therapeutics in *Mab* is dependent on the development of reliable infection models. By analogy, inhibition of cholesterol utilization by *Mtb* was

achieved using small molecules targeting HsaAB, HsaD, an adenylate cyclase as well as other undetermined cholesterol catabolic genes [53, 134, 136]. Although it is unclear whether any of these small molecules will be effective towards inhibiting cholesterol catabolism in *Mab*, the occurrence of CoA thioester metabolites in the degradation of cholesterol Rings A/B increases the value of these enzymes as targets for inhibitors given the proposed toxicity of CoA sequestration [178].

6.6.3 Therapeutics targeting IpdAB

The data presented herein corroborates previous work which identified IpdAB as essential for virulence [23, 52] as Δ*ipdAB Mtb* was attenuated in MΦ and failed to grow on cholesterol in the presence of another carbon source (Chapter 4). The characterization of IpdAB provides insights for the rational design of IpdAB inhibitors. It is conceivable that a non-hydrolyzable cyclopentane ring structure about C-4 of COCHEA-CoA, or any lactonized COCHEA-CoA analog thereof, could effectively bind the active site, inhibiting IpdAB. Similarly, the proposed mechanism requires a carbon-carbon double bond at C-4/C-5 of COCHEA-CoA for its hydrolysis. Presumably, saturation of this bond should not significantly alter binding, but would prevent hydroxylation at C-4.

Currently, no inhibitors of $IpdAB_{Mtb}$ have been reported in the literature. Unfortunately, the cholesterol-dependent-toxicity observed in Δ*ipdAB RHA1* is suppressed by the deletion of *fadD3* (Figure 4.3). Therefore, spontaneous suppressor mutants could diminish the effectiveness of therapeutics targeting IpdAB. Additionally, $FadD3_{Mtb}$ can be reversibly inhibited via post translational modification (Brown, unpublished) permitting the bacteria to arrest HIP catabolism, alleviating toxicity related to IpdAB inhibition. These processes may explain why Vanderven *et al.* failed to observe small molecules that target IpdAB or any cholesterol Rings C and D

degradation enzyme [134]. Although this diminishes the hope of IpdAB inhibitors as monotherapies, they may still prove effective in combination therapies.

6.7 Remaining questions and future directions

6.7.1 Cholesterol side chain and Rings A and B degradation

The identification of concurrent side chain and Rings A and B degradation implies that HsaAB acts on a substrate with a partially degraded side chain. Unfortunately, the specificity of HsaAB for HSBNC-CoA could not be reliably determined (data not shown). This suggests that either HSBNC-CoA is not the physiological substrate of HsaAB or that the assay conditions for HsaAB do not reflect its physiological activity. Due to the identification of multiple inhibitors of HsaAB that are active *in vivo* [134] and the poor substrate specificity towards 3-HSA [76], additional investigations of HsaAB with HSBNC-CoA are warranted.

Due to the limited availability of isopropionyl-CoA derivatives of steroid metabolites, I was unable to ascertain the precise alkyl-side chain intermediate for which $HsaD_{Mtb}$ and $HsaD_{Mab}$ have highest specificities. The physiological substrate of HsaD may contain an isopropionyl-, isopentanoyl, or isooctanoyl-CoA moiety with an array of different β-oxidative desaturation or hydroxylations thereof. Interestingly, the large entrance to the active site in $HsaD_{Mtb}$ may permit promiscuity towards a wide range side chain intermediates [77]. However, only metabolites with isopropionyl side chains were observed in Δ*hsaC* and Δ*hsaD* RHA1, suggesting that substrates with an isopropionyl-CoA side chain may be preferred by HsaD.

6.7.2 Regulation of cholesterol catabolism

In addition to the transcriptional regulation of cholesterol and HIP catabolism by KstR and KstR2, respectively, this catabolism is also regulated at the post-translational level. Acyl-CoA synthetases (FadD) catalyze the production of the respective effectors for KstR and KstR2

[79, 179]. FadDs are regulated via the reversible acetylation of their catalytic lysine, which inhibits catalytic activity. Acetylation is effected by protein lysine acetyltransferases (PAT) which, in turn, are regulated allosterically by cAMP [248-250]. Deacetylation of FadDs is catalyzed by deacetylases [250]. Recently, a number of small molecules that inhibit growth of *Mtb* on cholesterol and in MΦ were discovered that target an adenylate cyclase encoded by *rv1625c* [134, 136], implying a regulatory role for cAMP in cholesterol and/or HIP catabolism. To date, the precise involvement of cAMP is unknown in these pathways, however FadD3 is acetylated by a PAT in *Mtb* (Mt-PAT), suggesting Mt-PAT and cAMP mediated regulation (K. Brown, unpublished). Furthermore, KstR-regulated genes are up-regulated during overexpression of hypoxia related regulators, Rv0081 and DosR [188]. Although the Mce4 genes are up-regulated in the presence of cholesterol and are predicted to be KstR-regulated [251], overexpression of KstR did not lead to repression of the Mce4 operon [252, 253]. Similarly, chromosome immuno-precipitation (ChIP) of KstR failed to identify a KstR-binding site upstream of the *mce4* genes [252]. Overall, the additional regulatory mechanisms of cholesterol and HIP catabolism warrant future investigation.

6.7.3 Elucidation of HIP catabolism

The metabolic steps between 5OH-HIP-CoA to 5OH-HIC-CoA and MOODA-CoA to central metabolism were not elucidated (Figure 6.2) in Chapter 4. Presumably, 5αOH-HIP-CoA undergoes β-oxidation to form 5OH-HIC-CoA likely involving the short chain dehydrogenases, Rv3548c and Rv3549c, and/or the acyl-CoA dehydrogenases, FadE30 and FadE33 [25]. Preliminary characterizations of *M. smegmatis* with respective gene deletions of *rv3548c, rv3549c, fadE30,* and *fadE33* support our current functional assignment (data not shown). In contrast, *ΔfadA6 M. smegmatis* accumulated metabolites upstream of COCHEA-CoA

corroborating its low specific activity towards MeDODA-CoA. Overall, validation of these mutants was hindered by our inability to produce HIPE-CoA (Δ3'-5OH-HIP-CoA, Figure 6.2) for use in *in vitro* enzyme studies. Future studies should focus on producing FadE30 for the production of this metabolite.

6.7.4 Characterization of IpdAB

IpdABs make up a subfamily with a distinct catalytic function within the CoA transferase superfamily. However, it is unclear whether such enzymes are unique to steroid catabolism. Preliminary searches of actinobacterial genomes failed to identify annotated CoA transferases with conserved IpdAB characteristic residues such as $Glu105^{A}$, $Gly57^{B}$, $Arg126^{B}$, or $Arg92^{B}$. Unfortunately, amino acid sequence conservation with CoA transferases limit the effectiveness of BLAST searches for IpdAB paralogs in bacterial genomes. Given the ubiquity of steroid catabolism in bacteria, and the observation of COCHEA-CoA like catabolites in non-steroidal metabolic pathways [243], the preponderance of evidence would point to there being unidentified members of the IpdAB subfamily in non-steroidal catabolic pathways.

Bibliography

1. Makin, H.L.J.a.G., D.B. (2009): *Springer.*

2. Singh, P. (2016) *Frontiers in Cell and Developmental Biology*. **4**: p. 156.

3. Miras-Moreno, B., Sabater-Jara, A.B., Pedreno, M.A. and Almagro, L. (2016) *Journal of Agricultural Food Chemistry*. **64**(38): p. 7049-58.

4. Goluszko, P. and Nowicki, B. (2005) *Infection and Immunity*. **73**(12): p. 7791-6.

5. Chyu, K.-Y., Lio, W.M., Dimayuga, P.C., Zhou, J., Zhao, X., Yano, J., Trinidad, P., Honjo, T., et al. (2014) *PLoS One*. **9**(3): p. e92095.

6. Jin, S., Zhou, F., Katirai, F. and Li, P.-L. (2011) *Antioxidants & Redox Signaling*. **15**(4): p. 1043-1083.

7. Kabouridis, P.S., Janzen, J., Magee, A.L. and Ley, S.C. (2000) *European Journal of Immunology*. **30**(3): p. 954-63.

8. Barman, S. and Nayak, D.P. (2007) *Journal of Virology*. **81**(22): p. 12169-12178.

9. Buschiazzo, J., Ialy-Radio, C., Auer, J., Wolf, J.-P., Serres, C., Lefèvre, B. and Ziyyat, A. (2013) *PLoS One*. **8**(4): p. e62919.

10. Piironen, V., Lindsay, D.G., Miettinen, T.A., Toivo, J. and Lampi, A.-M. (2000) *Journal of the Science of Food and Agriculture*. **80**(7): p. 939-966.

11. Charlton-Menys, V. and Durrington, P.N. (2008) *Experimental Physiology*. **93**(1): p. 27-42.

12. Pearson, A., Budin, M. and Brocks, J.J. (2003) *Proceedings of the National Academy of Science* **100**(26): p. 15352-7.

13. Bird, C.W., Lynch, J.M., Pirt, F.J., Reid, W.W., Brooks, C.J.W. and Middleditch, B.S. (1971) *Nature*. **230**(5294): p. 473-474.

14. Bode, H.B., Zeggel, B., Silakowski, B., Wenzel, S.C., Reichenbach, H. and Muller, R. (2003) *Molecular Microbiology*. **47**(2): p. 471-81.

15. Wei, J.H., Yin, X. and Welander, P.V. (2016) *Frontiers in Microbiology*. **7**: p. 990.

16. Volkman, J.K. (2003) *Applied Microbiology and Biotechnology*. **60**(5): p. 495-506.

17. Bergstrand, L.H., Cardenas, E., Holert, J., Van Hamme, J.D. and Mohn, W.W. (2016) *MBio*. **7**(2).

18. Horinouchi, M., Hayashi, T. and Kudo, T. (2012) *Journal of Steroid Biochemistry and Molecular Biology*. **129**(1-2): p. 4-14.

19. Tai, H.H. and Sih, C.J. (1970) *Journal of Biological Chemistry*. **245**(19): p. 5062-71.

20. Yang, F.C., Chen, Y.L., Tang, S.L., Yu, C.P., Wang, P.H., Ismail, W., Wang, C.H., Ding, J.Y., et al. (2016) *ISME Journal*. **10**(8): p. 1967-83.

21. Barrientos, A., Merino, E., Casabon, I., Rodriguez, J., Crowe, A.M., Holert, J., Philipp, B., Eltis, L.D., et al. (2015) *Environmental Microbiology*. **17**(1): p. 47-63.

22. Mohn, W.W., Wilbrink, M.H., Casabon, I., Stewart, G.R., Liu, J., van der Geize, R. and Eltis, L.D. (2012) *Journal of Bacteriology*. **194**(24): p. 6712-9.

23. van der Geize, R., Grommen, A.W., Hessels, G.I., Jacobs, A.A. and Dijkhuizen, L. (2011) *PLoS Pathogens*. **7**(8): p. e1002181.

24. Petrusma, M., Hessels, G., Dijkhuizen, L. and van der Geize, R. (2011) *Journal of Bacteriology*. **193**(15): p. 3931-3940.

25. van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., et al. (2007) *Proceedings of the National Academy of Science*. **104**(6): p. 1947-1952.

26. García, J.L., Uhía, I. and Galán, B. (2012) *Microbial biotechnology*. **5**(6): p. 679-699.

27. Fernandez-Cabezon, L., Garcia-Fernandez, E., Galan, B. and Garcia, J.L. (2017) *Environmental Microbiology*.

28. Lee, S.S. and Sih, C.J. (1967) *Biochemistry*. **6**(5): p. 1395-1403.

29. Gibson, D.T., Wang, K.C., Sih, C.J. and Whitlock, H., Jr. (1966) *Journal of Biological Chemistry*. **241**(3): p. 551-9.

30. Sih, C.J. (1962) *Biochimica et Biophysica Acta*. **62**(3): p. 541-547.

31. Horinouchi, M., Yamamoto, T., Taguchi, K., Arai, H. and Kudo, T. (2001) *Microbiology*. **147**(Pt 12): p. 3367-75.

32. Kendall, S.L., Withers, M., Soffair, C.N., Moreland, N.J., Gurcha, S., Sidders, B., Frita, R., ten Bokum, A., et al. (2007) *Molecular Microbiology*. **65**(3): p. 684-699.

33. Pandey, A.K. and Sassetti, C.M. (2008) *Proceedings of the National Academy of Science*. **105**(11): p. 4376-80.

34. Wang, P.H., Leu, Y.L., Ismail, W., Tang, S.L., Tsai, C.Y., Chen, H.J., Kao, A.T. and Chiang, Y.R. (2013) *Journal of Lipid Research*. **54**(5): p. 1493-504.

35. Chiang, Y.R., Fang, J.Y., Ismail, W. and Wang, P.H. (2010) *Microbiology*. **156**(Pt 7): p. 2253-9.

36. Wang, P.H., Yu, C.P., Lee, T.H., Lin, C.W., Ismail, W., Wey, S.P., Kuo, A.T. and Chiang, Y.R. (2014) *Applied Environmental Microbiology*. **80**(11): p. 3442-52.

37. Horinouchi, M., Hayashi, T., Koshino, H., Kurita, T. and Kudo, T. (2005) *Applied Environmental Microbiology*. **71**(9): p. 5275-81.

38. Sih, C.J., Lee, S.S., Tsong, Y.Y. and Wang, K.C. (1966) *Journal of Biological Chemistry*. **241**(3): p. 540-550.

39. Yam, K.C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.X., Snieckus, V., Ly, L.H., Converse, P.J., et al. (2009) *PLoS Pathogens*. **5**(3): p. e1000344.

40. Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M. (2011) *PLoS Pathogens*. **7**(9): p. e1002251.

41. Penfield, J.S., Worrall, L.J., Strynadka, N.C. and Eltis, L.D. (2014) *Journal of Biological Chemistry*. **289**(37): p. 25523-36.

42. Wilbrink, M.H., Petrusma, M., Dijkhuizen, L. and van der Geize, R. (2011) *Applied and Environmental Microbiology*. **77**(13): p. 4455-4464.

43. Frank, D.J., Madrona, Y. and Ortiz de Montellano, P.R. (2014) *Journal of Biological Chemistry*. **289**(44): p. 30417-25.

44. Lu, R., Schmitz, W. and Sampson, N.S. (2015) *Biochemistry*. **54**(37): p. 5669-72.

45. Rohde, K.H., Abramovitch, R.B. and Russell, D.G. (2007) *Cell Host Microbe*. **2**(5): p. 352-64.

46. Homolka, S., Niemann, S., Russell, D.G. and Rohde, K.H. (2010) *PLoS Pathogens*. **6**(7): p. e1000988.

47. Nesbitt, N.M., Yang, X., Fontan, P., Kolesnikova, I., Smith, I., Sampson, N.S. and Dubnau, E. (2010) *Infection and Immunity*. **78**(1): p. 275-82.

48. Thomas, S.T., VanderVen, B.C., Sherman, D.R., Russell, D.G. and Sampson, N.S. (2011) *Journal of Biological Chemistry*. **286**(51): p. 43668-78.

49. Griffin, Jennifer E., Pandey, Amit K., Gilmore, Sarah A., Mizrahi, V., McKinney, John D., Bertozzi, Carolyn R. and Sassetti, Christopher M. (2012) *Chemistry & Biology*. **19**(2): p. 218-227.

50. Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J. and Dziadek, J. (2007) *FEMS Microbiology Letters*. **275**(1): p. 106-12.

51. Hu, Y., van der Geize, R., Besra, G.S., Gurcha, S.S., Liu, A., Rohde, M., Singh, M. and Coates, A. (2010) *Molecular Microbiology*. **75**(1): p. 107-21.

52. Rengarajan, J., Bloom, B.R. and Rubin, E.J. (2005) *Proceedings of the National Academy of Science*. **102**(23): p. 8327-32.

53. Ryan, A., Polycarpou, E., Lack, N.A., Evangelopoulos, D., Sieg, C., Halman, A., Bhakta, S., Eleftheriadou, O., et al. (2017) *British Journal of Pharmacology*.

54. Mohn, W.W., van der Geize, R., Stewart, G.R., Okamoto, S., Liu, J., Dijkhuizen, L. and Eltis, L.D. (2008) *Journal of Biological Chemistry*. **283**(51): p. 35368-74.

55. Ekiert, D.C., Bhabha, G., Isom, G.L., Greenan, G., Ovchinnikov, S., Henderson, I.R., Cox, J.S. and Vale, R.D. (2017) *Cell*. **169**(2): p. 273-285.e17.

56. Klepp, L.I., Forrellad, M.A., Osella, A.V., Blanco, F.C., Stella, E.J., Bianco, M.V., Santangelo Mde, L., Sassetti, C., et al. (2012) *Microbes and Infection*. **14**(7-8): p. 590-9.

57. Johnston, J.B., Ouellet, H. and de Montellano, P.R.O. (2010) *Journal of Biological Chemistry*. **285**(47): p. 36352-36360.

58. Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W.R., Jr., et al. (2009) *Journal of Biological Chemistry*. **284**(51): p. 35534-35542.

59. Capyk, J.K., D'Angelo, I., Strynadka, N.C. and Eltis, L.D. (2009) *Journal of Biological Chemistry*. **284**(15): p. 9937-46.

60. Ouellet, H., Guan, S., Johnston, J.B., Chow, E.D., Kells, P.M., Burlingame, A.L., Cox, J.S., Podust, L.M., et al. (2010) *Molecular Microbiology*. **77**(3): p. 730-42.

61. Johnston, J.B., Singh, A.A., Clary, A.A., Chen, C.-K., Hayes, P.Y., Chow, S., De Voss, J.J. and Ortiz de Montellano, P.R. (2012) *Bioorganic & Medicinal Chemistry*. **20**(13): p. 4064-4081.

62. Otani, H., Lee, Y.-E., Casabon, I. and Eltis, L.D. (2014) *Journal of Bacteriology*. **196**(24): p. 4293-4303.

63. Yang, M., Lu, R., Guja, K.E., Wipperman, M.F., St. Clair, J.R., Bonds, A.C., Garcia-Diaz, M. and Sampson, N.S. (2015) *ACS Infectious Disease*. **1**(2): p. 110-125

64. Wipperman, M.F., Sampson, N.S. and Thomas, S.T. (2014) *Critical Reviews in Biochemistry and Molecular Biology*. **49**(4): p. 269-293.

65. Schaefer, C.M., Lu, R., Nesbitt, N.M., Schiebel, J., Sampson, N.S. and Kisker, C. (2015) *Structure*. **23**(1): p. 21-33.

66. Chang, J.C., Miner, M.D., Pandey, A.K., Gill, W.P., Harik, N.S., Sassetti, C.M. and Sherman, D.R. (2009) *Journal of Bacteriology*. **191**(16): p. 5232-9.

67. Yang, M., Guja, K.E., Thomas, S.T., Garcia-Diaz, M. and Sampson, N.S. (2014) *ACS Chemical Biology*. **9**(11): p. 2632-2645.

68. Wilbrink, M.H., van der Geize, R. and Dijkhuizen, L. (2012) *Microbiology*. **158**(12): p. 3054-3062.

69. Capyk, J.K., Casabon, I., Gruninger, R., Strynadka, N.C. and Eltis, L.D. (2011) *Journal of Biological Chemistry*. **286**(47): p. 40717-24.

70. Yang, X., Gao, J., Smith, I., Dubnau, E. and Sampson, N.S. (2011) *Journal of Bacteriology*. **193**(6): p. 1473-6.

71. Kreit, J. and Sampson, N.S. (2009) *The FEBS Journal*. **276**(23): p. 6844-6856.

72. Brzostek, A., Rumijowska-Galewicz, A., Dziadek, B., Wojcik, E.A. and Dziadek, J. (2013) *Journal of Steroid Biochemistry and Molecular Biology*. **134**: p. 1-7.

73. Brzostek, A., Sliwinski, T., Rumijowska-Galewicz, A., Korycka-Machala, M. and Dziadek, J. (2005) *Microbiology*. **151**(Pt 7): p. 2393-402.

74. Knol, J., Bodewits, K., Hessels, G.I., Dijkhuizen, L. and van der Geize, R. (2008) *Biochemistry Journal*. **410**(2): p. 339-46.

75. van der Geize, R., Hessels, G.I., van Gerwen, R., van der Meijden, P. and Dijkhuizen, L. (2002) *Molecular Microbiology*. **45**(4): p. 1007-18.

76. Dresen, C., Lin, L.Y., D'Angelo, I., Tocheva, E.I., Strynadka, N. and Eltis, L.D. (2010) *Journal of Biological Chemistry*. **285**(29): p. 22264-22275.

77. Lack, N.A., Yam, K.C., Lowe, E.D., Horsman, G.P., Owen, R.L., Sim, E. and Eltis, L.D. (2010) *Journal of Biological Chemistry*. **285**(1): p. 434-443.

78. Carere, J., McKenna, S.E., Kimber, M.S. and Seah, S.Y. (2013) *Biochemistry*. **52**(20): p. 3502–3511.

79. Casabon, I., Crowe, A.M., Liu, J. and Eltis, L.D. (2013) *Molecular Microbiology*. **87**(2): p. 269-283.

80. Yam, K.C., Okamoto, S., Roberts, J.N. and Eltis, L.D. (2011) *Canadian Journal of Microbiology*. **57**(3): p. 155-68.

81. Seah, S.Y., Labbe, G., Kaschabek, S.R., Reifenrath, F., Reineke, W. and Eltis, L.D. (2001) *Journal of Bacteriology*. **183**(5): p. 1511-6.

82. Eltis, L.D., Hofmann, B., Hecht, H.J., Lunsdorf, H. and Timmis, K.N. (1993) *Journal of Biological Chemistry*. **268**(4): p. 2727-32.

83. Kendall, S.L., Burgess, P., Balhana, R., Withers, M., ten Bokum, A., Lott, J.S., Gao, C., Uhia-Castro, I., et al. (2010) *Microbiology*. **156**(Pt 5): p. 1362-1371.

84. Hashimoto, S. and Hayakawa, S. (1977) *Biochemistry Journal*. **164**(3): p. 715-26.

85. Capyk, J. (2012). Doctoral Dissertation. *The University of British Columbia*

86. Ho, N.A., Dawes, S.S., Crowe, A.M., Casabon, I., Gao, C., Kendall, S.L., Baker, E.N., Eltis, L.D., et al. (2016) *Journal of Biological Chemistry*. **291**(14): p. 7256-66.

87. Fujihashi, M., Nakatani, T., Hirooka, K., Matsuoka, H., Fujita, Y. and Miki, K. (2014) *Proteins*. **82**(7): p. 1301-10.

88. Agari, Y., Agari, K., Sakamoto, K., Kuramitsu, S. and Shinkai, A. (2011) *Microbiology*. **157**(Pt 6): p. 1589-601.

89. Luengo, J.M., Garcia, J.L. and Olivera, E.R. (2001) *Molecular Microbiology*. **39**(6): p. 1434-42.

90. Otani, H., Stogios, P.J., Xu, X., Nocek, B., Li, S.N., Savchenko, A. and Eltis, L.D. (2016) *Nucleic Acids Research*. **44**(2): p. 595-607.

91. Casabon, I., Zhu, S.H., Otani, H., Liu, J., Mohn, W.W. and Eltis, L.D. (2013) *Molecular Microbiology*. **89**(6): p. 1201-1212.

92. Pruneda-Paz, J.L., Linares, M., Cabrera, J.E. and Genti-Raimondi, S. (2004) *Journal of Bacteriology*. **186**(5): p. 1430-7.

93. Holert, J., Kulic, Z., Yucel, O., Suvekbala, V., Suter, M.J., Moller, H.M. and Philipp, B. (2013) *Journal of Bacteriology*. **195**(3): p. 585-95.

94. WHO. (2016). *Global Tuberculosis Report.*URL: www.who.int/tb

95. Russell, D.G., Barry, C.E. and Flynn, J.L. (2010) *Science*. **328**(5980): p. 852-856.

96. Schatz, A., Bugle, E. and Waksman, S.A. (1944) *Proceedings of the Society for Experimental Biology and Medicine*. **55**(1): p. 66-69.

97. Zumla, A., Nahid, P. and Cole, S.T. (2013) *Nature Reviews Drug Discovery*. **12**(5): p. 388-404.

98. Johnson, R., Streicher, E.M., Louw, G.E., Warren, R.M., van Helden, P.D. and Victor, T.C. (2006) *Current Issues in Molecular Biology*. **8**(2): p. 97-111.

99. Udwadia, Z.F., Amale, R.A., Ajbani, K.K. and Rodrigues, C. (2012) *Clinical Infectious Disease*. **54**(4): p. 579-81.

100. Marrakchi, H., Lanéelle, M.-A. and Daffé, M. (2014) *Chemistry & Biology*. **21**(1): p. 67-85.

101. Jarlier, V. and Nikaido, H. (1990) *Journal of Bacteriology*. **172**(3): p. 1418-1423.

102. Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.P., Clement, C., Ouhdouch, Y., et al. (2016) *Microbiology and Molecular Biology Reviews*. **80**(1): p. 1-43.

103. Mahato, S.B. and Garai, S. (1997) *Steroids*. **62**(4): p. 332-345.

104. Fernandes, P., Cruz, A., Angelova, B., Pinheiro, H.M. and Cabral, J.M.S. (2003) *Enzyme and Microbial Technology*. **32**(6): p. 688-705.

105. van der Geize, R. and Dijkhuizen, L. (2004) *Current Opinion in Microbiology*. **7**(3): p. 255-261.

106. Lee, M.-R., Sheng, W.-H., Hung, C.-C., Yu, C.-J., Lee, L.-N. and Hsueh, P.-R. (2015) *Emerging Infectious Diseases*. **21**(9): p. 1638-1646.

107. Bryant, J.M., Grogono, D.M., Rodriguez-Rincon, D., Everall, I., Brown, K.P., Moreno, P., Verma, D., Hill, E., et al. (2016) *Science*. **354**(6313): p. 751-757.

108. Jeon, K., Kwon, O.J., Lee, N.Y., Kim, B.J., Kook, Y.H., Lee, S.H., Park, Y.K., Kim, C.K., et al. (2009) *American Journal of Respiratory and Critical Care Medicine*. **180**(9): p. 896-902.

109. Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. and Jacobs, W.R., Jr. (1990) *Molecular Microbiology*. **4**(11): p. 1911-9.

110. Seto, M., Kimbara, K., Shimura, M., Hatta, T., Fukuda, M. and Yano, K. (1995) *Applied Environmental Microbiology*. **61**(9): p. 3353-3358.

111. McLeod, M.P., Warren, R.L., Hsiao, W.W.L., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., et al. (2006) *Proceedings of the National Academy of Sciences*. **103**(42): p. 15582-15587.

112. Lee, W., VanderVen, B.C., Fahey, R.J. and Russell, D.G. (2013) *Journal of Biological Chemistry*. **288**(10): p. 6788-800.

113. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., et al. (1998) *Nature*. **393**(6685): p. 537-44.

114. Sassetti, C.M. and Rubin, E.J. (2003) *Proceedings of the National Academy of Sciences*. **100**(22): p. 12989-94.

115. Lovewell, R.R., Sassetti, C.M. and VanderVen, B.C. (2016) *Current Opinions in Microbiology*. **29**: p. 30-6.

116. Savvi, S., Warner, D.F., Kana, B.D., McKinney, J.D., Mizrahi, V. and Dawes, S.S. (2008) *Journal of Bacteriology*. **190**(11): p. 3886-95.

117. McKinney, J.D., Honer zu Bentrup, K., Munoz-Elias, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., et al. (2000) *Nature*. **406**(6797): p. 735-8.

118. Daniel, J., Maamar, H., Deb, C., Sirakova, T.D. and Kolattukudy, P.E. (2011) *PLoS Pathogens*. **7**(6): p. e1002093.

119. Munoz-Elias, E.J., Upton, A.M., Cherian, J. and McKinney, J.D. (2006) *Molecular Microbiology*. **60**(5): p. 1109-22.

120. Quadri, L.E. (2014) *Critical Reviews in Biochemistry and Molecular Biology*. **49**(3): p. 179-211.

121. Carroll, P. and Parish, T. (2015) *PLoS One*. **10**(7): p. e0133129.

122. Kim, M.J., Wainwright, H.C., Locketz, M., Bekker, L.G., Walther, G.B., Dittrich, C., Visser, A., Wang, W., et al. (2010) *EMBO Molecular Medicine*. **2**(7): p. 258-74.

123. Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffe, M., Emile, J.F., et al. (2008) *PLoS Pathogens*. **4**(11): p. e1000204.

124. Martens, G.W., Arikan, M.C., Lee, J., Ren, F., Vallerskog, T. and Kornfeld, H. (2008) *Infection and Immunity*. **76**(8): p. 3464-72.

125. Schafer, G., Guler, R., Murray, G., Brombacher, F. and Brown, G.D. (2009) *PLoS One*. **4**(12): p. e8448.

126. Su, V.Y., Su, W.J., Yen, Y.F., Pan, S.W., Chuang, P.H., Feng, J.Y., Chou, K.T., Yang, K.Y., et al. (2017) *Chest*.

127. Parihar, S.P., Guler, R., Khutlang, R., Lang, D.M., Hurdayal, R., Mhlanga, M.M., Suzuki, H., Marais, A.D., et al. (2014) *Journal of Infectious Disease*. **209**(5): p. 754-63.

128. Munoz, S., Rivas-Santiago, B. and Enciso, J.A. (2009) *Scandinavian Journal of Immunology*. **70**(3): p. 256-63.

129. Gatfield, J. and Pieters, J. (2000) *Science*. **288**(5471): p. 1647-50.

130. Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E.J., Leys, D., et al. (2010) *Journal of Biological Chemistry*. **285**(49): p. 38270-38282.

131. Salamon, H., Bruiners, N., Lakehal, K., Shi, L., Ravi, J., Yamaguchi, K.D., Pine, R. and Gennaro, M.L. (2014) *Journal of Immunology* . **193**(1): p. 30-34.

132. Marques, M.A., Berredo-Pinho, M., Rosa, T.L., Pujari, V., Lemes, R.M., Lery, L.M., Silva, C.A., Guimaraes, A.C., et al. (2015) *Journal of Bacteriology*. **197**(23): p. 3698-707.

133. Norlin, M., von Bahr, S., Bjorkhem, I. and Wikvall, K. (2003) *Journal of Lipid Research*. **44**(8): p. 1515-22.

134. VanderVen, B.C., Fahey, R.J., Lee, W., Liu, Y., Abramovitch, R.B., Memmott, C., Crowe, A.M., Eltis, L.D., et al. (2015) *PLoS Pathogens*. **11**(2): p. e1004679.

135. Ryan, A., Keany, S., Eleftheriadou, O., Ballet, R., Cheng, H.Y. and Sim, E. (2014) *FEMS Microbiology Letters*. **350**(1): p. 42-7.

136. Johnson, R.M., Bai, G., DeMott, C.M., Banavali, N.K., Montague, C.R., Moon, C., Shekhtman, A., VanderVen, B., et al. (2017) *Molecular Microbiology*. **105**(2): p. 294-308

137. Kumar, P., Chhibber, M. and Surolia, A. (2007) *Biochemical and Biophysical Research Communications*. **361**(4): p. 903-909.

138. Ambady, A., Awasthy, D., Yadav, R., Basuthkar, S., Seshadri, K. and Sharma, U. (2012) *Tuberculosis*. **92**(6): p. 521-528.

139. Strauss, E., Kinsland, C., Ge, Y., McLafferty, F.W. and Begley, T.P. (2001) *Journal of Biological Chemistry*. **276**(17): p. 13513-6.

140. Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) *Molecular Microbiology*. **48**(1): p. 77-84.

141. Evans, J.C., Trujillo, C., Wang, Z., Eoh, H., Ehrt, S., Schnappinger, D., Boshoff, H.I., Rhee, K.Y., et al. (2016) *ACS Infectious Disease*. **2**(12): p. 958-968.

142. Nelson, D.L. 2005: Fourth edition. New York : *W.H. Freeman*.

143. White, H. and Jencks, W.P. (1976) *Journal of Biological Chemistry*. **251**(6): p. 1688-99.

144. Modis, Y. and Wierenga, R.K. (1999) *Structure*. **7**(10): p. 1279-1290.

145. Luo, L., Taylor, K.L., Xiang, H., Wei, Y., Zhang, W. and Dunaway-Mariano, D. (2001) *Biochemistry*. **40**(51): p. 15684-92.

146. Whitty, A., Fierke, C.A. and Jencks, W.P. (1995) *Biochemistry*. **34**(37): p. 11678-89.

147. Fraser, M.E., Hayakawa, K. and Brown, W.D. (2010) *Biochemistry*. **49**(48): p. 10319-28.

148. Haynes, C.A. (2011) *Biochimica et Biophysica Acta*. **1811**(11): p. 663-8.

149. Dalluge, J.J., Gort, S., Hobson, R., Selifonova, O., Amore, F. and Gokarn, R. (2002) *Analytical and Bioanalytical Chemistry*. **374**(5): p. 835-840.

150. Li, Q., Zhang, S., Berthiaume, J.M., Simons, B. and Zhang, G.F. (2014) *Journal of Lipid Research*. **55**(3): p. 592-602.

151. Basu, S.S. and Blair, I.A. (2012) *Nature Protocols*. **7**(1): p. 1-12.

152. Alcaide, M., Tornes, J., Stogios, P.J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., et al. (2013) *Biochemistry Journal*. **454**(1): p. 157-66.

153. Kuatsjah, E., Chen, H.M., Withers, S.G. and Eltis, L.D. (2017) *FEBS Letters*. **591**(7): p. 1001-1009.

154. Ghosal, D., Ghosh, S., Dutta, T.K. and Ahn, Y. (2016) *Frontiers in Microbiology*. **7**: p. 1369.

155. Díaz, E., Jiménez, J.I. and Nogales, J. (2013) *Current Opinion in Biotechnology*. **24**(3): p. 431-442.

156. Horsman, G.P., Ke, J., Dai, S., Seah, S.Y.K., Bolin, J.T. and Eltis, L.D. (2006) *Biochemistry*. **45**(37): p. 11071-11086.

157. Fleming, S.M., Robertson, T.A., Langley, G.J. and Bugg, T.D. (2000) *Biochemistry*. **39**(6): p. 1522-31.

158. Ruzzini, A.C., Ghosh, S., Horsman, G.P., Foster, L.J., Bolin, J.T. and Eltis, L.D. (2012) *Journal of the American Chemical Society*. **134**(10): p. 4615-4624.

159. Ruzzini, A.C., Bhowmik, S., Ghosh, S., Yam, K.C., Bolin, J.T. and Eltis, L.D. (2013) *Biochemistry*. **52**(42): p. 7428-38.

160. Díaz, E. and Timmis, K.N. (1995) *Journal of Biological Chemistry*. **270**(11): p. 6403-6411.

161. Dunn, G., Montgomery, M.G., Mohammed, F., Coker, A., Cooper, J.B., Robertson, T., Garcia, J.L., Bugg, T.D., et al. (2005) *Journal of Molecular Biology*. **346**(1): p. 253-65.

162. Ruzzini, A.C., Bhowmik, S., Yam, K.C., Ghosh, S., Bolin, J.T. and Eltis, L.D. (2013) *Biochemistry*. **52**(33): p. 5685-5695.

163. Hillen, W., Klock, G., Kaffenberger, I., Wray, L.V. and Reznikoff, W.S. (1982) *Journal of Biological Chemistry*. **257**(11): p. 6605-13.

164. Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., et al. (2005) *Microbiology and Molecular Biology Reviews*. **69**(2): p. 326-56.

165. Yu, Z., Reichheld, S.E., Savchenko, A., Parkinson, J. and Davidson, A.R. (2010) *Journal of Molecular Biology*. **400**(4): p. 847-64.

166. Cuthbertson, L. and Nodwell, J.R. (2013) *Microbiology and Molecular Biology Reviews*. **77**(3): p. 440-75.

167. Heider, J. (2001) *FEBS Letters*. **509**(3): p. 345-9.

168. Buckel, W., Dorn, U. and Semmler, R. (1981) *European Journal of Biochemistry*. **118**(2): p. 315-21.

169. Jacob, U., Mack, M., Clausen, T., Huber, R., Buckel, W. and Messerschmidt, A. (1997) *Structure*. **5**(3): p. 415-26.

170. Selmer, T. and Buckel, W. (1999) *Journal of Biological Chemistry*. **274**(30): p. 20772-8.

171. Dimroth, P., Loyal, R. and Eggerer, H. (1977) *European Journal of Biochemistry*. **80**(2): p. 479-88.

172. Buckel, W. and Bobi, A. (1976) *European Journal of Biochemistry*. **64**(1): p. 255-62.

173. Berthold, C.L., Toyota, C.G., Richards, N.G. and Lindqvist, Y. (2008) *Journal of Biological Chemistry*. **283**(10): p. 6519-29.

174. Leutwein, C. and Heider, J. (2001) *Journal of Bacteriology*. **183**(14): p. 4288-95.

175. Rangarajan, E.S., Li, Y., Ajamian, E., Iannuzzi, P., Kernaghan, S.D., Fraser, M.E., Cygler, M. and Matte, A. (2005) *Journal of Biological Chemistry*. **280**(52): p. 42919-28.

176. Murphy, J.R., Mullins, E.A. and Kappock, T.J. (2016) *Frontiers in Chemistry*. **4**: p. 23.

177. Crowe, A.M., Stogios, P.J., Casabon, I., Evdokimova, E., Savchenko, A. and Eltis, L.D. (2015) *Journal of Biological Chemistry*. **290**: p. 872-882.

178. Crowe, A.M., Casabon, I., Brown, K.L., Liu, J., Lian, J., Rogalski, J.C., Hurst, T.E., Snieckus, V., et al. (2017) *mBio*. **8**(2).

179. Casabon, I., Swain, K., Crowe, A.M., Eltis, L.D. and Mohn, W.W. (2014) *Journal of Bacteriology*. **196**(3): p. 579-587.

180. Sambrook, J. and Russell, D.W. (2001) *Cold Spring Harbor Laboratory Press.*

181. Dereeper, A., Audic, S., Claverie, J.M. and Blanc, G. (2010) *BMC Evolutionary Biology*. **10**: p. 8.

182. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., et al. (2008) *Nucleic Acids Research*. **36**: p. W465-9.

183. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. and Notredame, C. (2006) *Nucleic Acids Research*. **34**: p. W604-8.

184. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F. and Notredame, C. (2011) *Nucleic Acids Research*. **39**: p. W13-7.

185. Bauchop, T. and Elsden, S.R. (1960) *Journal of General Microbiology*. **23**: p. 457-469.

186. Cornish-Bowden, A. (2014) *Perspectives in Science*. **1**(1–6): p. 121-125.

187. Uhia, I., Galan, B., Kendall, S.L., Stoker, N.G. and Garcia, J.L. (2012) *Environmental Microbiology Reports*. **4**(2): p. 168-82.

188. Galagan, J.E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., et al. (2013) *Nature*. **499**(7457): p. 178-183.

189. Seah, S.Y., Ke, J., Denis, G., Horsman, G.P., Fortin, P.D., Whiting, C.J. and Eltis, L.D. (2007) *Journal of Bacteriology*. **189**(11): p. 4038-45.

190. Fernández-Cabezón, L., Galán, B. and García, J.L. (2017) *Microbial Biotechnology*. **10**(1): p. 151-161.

191. Habe, H., Kasuga, K., Nojiri, H., Yamane, H. and Omori, T. (1996) *Applied Environmental Microbiology*. **62**(12): p. 4471-7.

192. Dabrock, B., Kesseler, M., Averhoff, B. and Gottschalk, G. (1994) *Applied Environmental Microbiology*. **60**(3): p. 853-60.

193. Stecker, C., Johann, A., Herzberg, C., Averhoff, B. and Gottschalk, G. (2003) *Journal of Bacteriology*. **185**(17): p. 5269-74.

194. Dagley, S., Evans, W.C. and Ribbons, D.W. (1960) *Nature*. **188**: p. 560-6.

195. Kapust, R.B., Tözsér, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D. and Waugh, D.S. (2001) *Protein Engineering*. **14**(12): p. 993-1000.

196. Minor, W., Cymborowski, M., Otwinowski, Z. and Chruszcz, M. (2006) *Acta Crystallographica D Biological Crystallography*. **62**(Pt 8): p. 859-66.

197. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., et al. (2011) *Acta Crystallographica D Biological Crystallography*. **67**(Pt 4): p. 235-42.

198. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., et al. (2010) *Acta Crystallographica D Biological Crystallography*. **66**(Pt 2): p. 213-21.

199. Emsley, P. and Cowtan, K. (2004) *Acta Crystallographica D Biological Crystallography*. **60**(Pt 12 Pt 1): p. 2126-32.

200. Krissinel, E. and Henrick, K. (2007) *Journal of Molecular Biology*. **372**(3): p. 774-97.

201. Holm, L. and Rosenstrom, P. (2010) *Nucleic Acids Research*. **38**: p. W545-9.

202. Krissinel, E. and Henrick, K. (2004) *Acta Crystallographica D Biological Crystallography*. **60**(Pt 12 Pt 1): p. 2256-68.

203. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) *Journal of Computional Chemistry*. **25**(13): p. 1605-12.

204. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) *Nucleic Acids Research*. **34**: p. W116-8.

205. Tonthat, N.K., Milam, S.L., Chinnam, N., Whitfill, T., Margolin, W. and Schumacher, M.A. (2013) *Proceedings of the National Academy of Sciences* . **110**(26): p. 10586-91.

206. Rautengarten, C., Baidoo, E., Keasling, J.D. and Scheller, H.V. (2010) *Bioenergy Research*. **3**: p. 115-122.

207. Peter, D.M., Vogeli, B., Cortina, N.S. and Erb, T.J. (2016) *Molecules*. **21**(4).

208. Ramon-Garcia, S., Ng, C., Jensen, P.R., Dosanjh, M., Burian, J., Morris, R.P., Folcher, M., Eltis, L.D., et al. (2013) *Journal of Biological Chemistry*. **288**(48): p. 34514-28.

209. van Kessel, J.C., Marinelli, L.J. and Hatfull, G.F. (2008) *Nature Reviews Microbiology*. **6**(11): p. 851-857.

210. van der Geize, R., Hessels, G.I., van Gerwen, R., van der Meijden, P. and Dijkhuizen, L. (2001) *FEMS Microbiology Letters*. **205**(2): p. 197-202.

211. Upton, A.M. and McKinney, J.D. (2007) *Microbiology*. **153**(Pt 12): p. 3973-3982.

212. Minkler, P.E., Kerner, J., Ingalls, S.T. and Hoppel, C.L. (2008) *Analytical Biochemistry*. **376**(2): p. 275-276.

213. Maksymiuk, C., Ioerger, T., Balakrishnan, A., Bryk, R., Rhee, K., Sacchettini, J. and Nathan, C. (2015) *Tuberculosis (Edinb)*. **95**(6): p. 689-94.

214. Horinouchi, M., Hayashi, T., Kudo, T., Koshino, H., *Jpn. Kokai Tokkyo Koho*. 2006.

215. Sogabe, S., Yoshizumi, A., Fukami, T.A., Shiratori, Y., Shimizu, S., Takagi, H., Nakamori, S. and Wada, M. (2003) *Journal of Biological Chemistry*. **278**(21): p. 19387-95.

216. Saito, J., Yamada, M., Watanabe, T., Iida, M., Kitagawa, H., Takahata, S., Ozawa, T., Takeuchi, Y., et al. (2008) *Protein Science*. **17**(4): p. 691-9.

217. Hamed, R.B., Batchelar, E.T., Clifton, I.J. and Schofield, C.J. (2008) *Cell Molecular Life Science*. **65**(16): p. 2507-27.

218. Li, H.J., Li, X., Liu, N., Zhang, H., Truglio, J.J., Mishra, S., Kisker, C., Garcia-Diaz, M., et al. (2011) *Biochemistry*. **50**(44): p. 9532-44.

219. Pelletier, D.A. and Harwood, C.S. (1998) *Journal of Bacteriology*. **180**(9): p. 2330-6.

220. Corthesy-Theulaz, I.E., Bergonzelli, G.E., Henry, H., Bachmann, D., Schorderet, D.F., Blum, A.L. and Ornston, L.N. (1997) *Journal of Biological Chemistry*. **272**(41): p. 25659-67.

221. Sramek, S.J. and Frerman, F.E. (1975) *Archives of Biochemistry and Biophysics*. **171**(1): p. 14-26.

222. Winter, G. (2010) *Journal of Applied Crystallography*. **43**(1): p. 186-190.

223. Kabsch, W. (2010) *Acta Crystallographica D Biological Crystallography*. **66**(Pt 2): p. 125-32.

224. Evans, P.R. and Murshudov, G.N. (2013) *Acta Crystallographica D Biologcial Crystallography*. **69**(Pt 7): p. 1204-14.

225. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) *Journal of Applied Crystallography*. **40**(Pt 4): p. 658-674.

226. Afonine, P.V., Moriarty, N.W., Mustyakimov, M., Sobolev, O.V., Terwilliger, T.C., Turk, D., Urzhumtsev, A. and Adams, P.D. (2015) *Acta Crystallographica D Biological Crystallography*. **71**(Pt 3): p. 646-66.

227. Hersh, L.B. and Jencks, W.P. (1967) *Journal of Biological Chemistry*. **242**(2): p. 339-40.

228. Niles, R., Witkowska, H.E., Allen, S., Hall, S.C., Fisher, S.J. and Hardt, M. (2009) *Analytical chemistry*. **81**(7): p. 2804-2809.

229. Redington, R.L. (1976) *Journal of Physical Chemistry*. **80**(3): p. 229-235.

230. Byrn, M. and Calvin, M. (1966) *Journal of the American Chemical Society*. **88**(9): p. 1916-1922.

231. Denessiouk, K.A., Rantanen, V.V. and Johnson, M.S. (2001) *Proteins*. **44**(3): p. 282-91.

232. Yang, X., Dubnau, E., Smith, I. and Sampson, N.S. (2007) *Biochemistry*. **46**(31): p. 9058-9067.

233. Thomas, S.T. and Sampson, N.S. (2013) *Biochemistry*. **52**(17): p. 2895-2904.

234. Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., Buchanan, S.G., Buchanan, M.D., et al. (2005) *Proteins*. **60**(4): p. 787-96.

235. Miller, D.J., Zhang, Y.M., Subramanian, C., Rock, C.O. and White, S.W. (2010) *Nature Structural and Molecular Biology*. **17**(8): p. 971-5.

236. Anand, S., Singh, V., Singh, A.K., Mittal, M., Datt, M., Subramani, B. and Kumaran, S. (2012) *FEBS Journal*. **279**(17): p. 3214-28.

237. Reichheld, S.E., Yu, Z. and Davidson, A.R. (2009) *Proceedings of the National Academy of Sciences*. **106**(52): p. 22263-8.

238. Yang, S., Gao, Z., Li, T., Yang, M., Zhang, T., Dong, Y. and He, Z.G. (2013) *Journal of Biological Chemistry*. **288**(33): p. 23687-95.

239. Bhukya, H., Bhujbalrao, R., Bitra, A. and Anand, R. (2014) *Nucleic Acids Research*. **42**(15): p. 10122-33.

240. Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2002) *EMBO Journal*. **21**(5): p. 1210-8.

241. Harwood, C.S. and Parales, R.E. (1996) *Annual Reviews in Microbiology*. **50**: p. 553-90.

242. Middleton, B. (1973) *Biochemical Journal*. **132**(4): p. 717-730.

243. Laempe, D., Jahn, M. and Fuchs, G. (1999) *European Journal of Biochemistry*. **263**(2): p. 420-9.

244. Harwood, C.S., Burchhardt, G., Herrmann, H. and Fuchs, G. (1998) *FEMS Microbiology Reviews*. **22**(5): p. 439-458.

245. Wells, T., Jr. and Ragauskas, A.J. (2012) *Trends in Biotechnology*. **30**(12): p. 627-37.

246. Cortes, M., Singh, A.K., Reyrat, J.-M., Gaillard, J.-L., Nassif, X. and Herrmann, J.-L. (2011) *PLoS One*. **6**(12): p. e29306.

247. Roux, A.L., Viljoen, A., Bah, A., Simeone, R., Bernut, A., Laencina, L., Deramaudt, T., Rottman, M., et al. (2016) *Open Biology*. **6**(11).

248. Lee, H.J., Lang, P.T., Fortune, S.M., Sassetti, C.M. and Alber, T. (2012) *Nature Structural and Molecular Biology*. **19**(8): p. 811-818.

249. Montgomery, D.C., Sorum, A.W. and Meier, J.L. (2015) *ACS Chemical Biology*. **10**(1): p. 85-94.

250. Xu, H., Hegde, S.S. and Blanchard, J.S. (2011) *Biochemistry*. **50**(26): p. 5883-92.

251. Garcia-Fernandez, J., Papavinasasundaram, K., Galan, B., Sassetti, C.M. and Garcia, J.L. (2017) *Environmental Microbiology*.

252. Peterson, E.J.R., Reiss, D.J., Turkarslan, S., Minch, K.J., Rustad, T., Plaisier, C.L., Longabaugh, W.J.R., Sherman, D.R., et al. (2014) *Nucleic Acids Research*. **42**(18): p. 11291-11303.

253. Minch, K.J., Rustad, T.R., Peterson, E.J., Winkler, J., Reiss, D.J., Ma, S., Hickey, M., Brabant, W., et al. (2015) *Nat Communications*. **6**: p. 5829.

254. Masai, E., Yamada, A., Healy, J.M., Hatta, T., Kimbara, K., Fukuda, M. and Yano, K. (1995) *Applied Environmental Microbiology*. **61**(6): p. 2079-85.

255. Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. and Puhler, A. (1994) *Gene*. **145**(1): p. 69-73.

256. van Kessel, J.C. and Hatfull, G.F. (2008) *Methods in Molecular Biology*. **435**: p. 203-15.

257. Burian, J., Ramón-García, S., Sweet, G., Gómez-Velasco, A., Av-Gay, Y. and Thompson, C.J. (2012) *Journal of Biological Chemistry*. **287**(1): p. 299-310.

258. Nakashima, N. and Tamura, T. (2004) *Applied Environmental Microbiology*. **70**(9): p. 5557-5568.

Appendices

Appendix A Bacterial strains, plasmids and oligonucleotides in Chapters 4 and 5

Strain or construct name	Description	Reference
Strains		
RHA1	*R. jostii* RHA1	[254]
Δ*ipdAB* RHA1	RHA1 Δ*RS22695-22690*	This study
Δ*ipdC* RHA1	RHA1 Δ*RS22685*	This study
Δ*ipdABC* RHA1	RHA1 *ΔRS22685-22690*	This study
Δ*fadD3* RHA1	RHA1 Δ*RS22410*	[79]
M. smegmatis mc^2155		(NCBI 246196)
Δ*echA20 M. smegmatis*	mc^2155 Δ*msmeg_6001*, Hyg^R	This study
Δ*echA20 M. smegmatis::echA20*	mc^2155 Δ*msmeg_6001* containing pMVEchA20, Hyg^R Apr^R	This study
Δ*fadE32 M. smegmatis*	mc^2155 Δ*msmeg_6015*, Hyg^R	This study
Δ*fadE32 M. smegmatis::fadE32*	mc^2155 Δ*msmeg_6015* containing pMVFadE32, Hyg^R Apr^R	This study
Δ*ipdF M. smegmatis*	mc^2155 Δ*msmeg_6011*, Hyg^R	This study
Δ*ipdF M. smegmatis::rv3559c*	mc^2155 Δ*msmeg_6011*containing pMVrv3559c, Hyg^R Apr^R	This study
Δ*ipdAB M.smegmatis*	mc^2155 Δ*msmeg_6002-6003*, Hyg^R	This study
Mtb	*M. tuberculosis* Erdman	(NCBI 652616)
Δ*ipdAB Mtb*	Erdman Δ*erdman_3896-3897*, Hyg^R	This study
Δ*ipdAB Mtb::ipdAB*	Erdman Δ*erdman_3896-3897* containing pMVipdAB, Hyg^R Apr^R	This study
Δ*ipdC Mtb*	Erdman Δ*erdman_3898*, Hyg^R	This study
Δ*ipdC Mtb::ipdABC*	Erdman Δ*erdman_3898* containing pMVipdABC, Hyg^R Apr^R	This study
E. coli DH5α		Invitrogen -
E. coli Rosetta2 pLysS		Novagen -

Strain or construct name	Description	Reference
E. coli BL21(DE3)		New England Biolabs -
Plasmid constructs		
pK18*mobsacB*	Plasmid for allelic exchange in RHA1. Kan^R	[255]
pK18ΔipdAB	pK18*mobsacB* with upstream region of *RS22695* and downstream region of *RS22690* cloned into *Eco*RI/*Hin*dIII sites. Kan^R	This study
pK18ΔipdC	pK18*mobsacB* with upstream and downstream regions of *RS22685* cloned into *Eco*RI/*Hin*dIII sites. Kan^R	
pYUB854	Recombineering plasmid used for allelic exchange in mycobacteria. Hyg^R	[256]
pYUBud6001	pYUB854 containing upstream (*Afl*II/*Xba*I) and downstream (*Bgl*II/*Nhe*I) regions of *Msmeg_6001* flanking hyg^R. Hyg^R	This study
pYUBud6011	pYUB854 containing upstream (*Afl*II/*Xba*I) and downstream (*Bgl*II/*Nhe*I) regions of *Msmeg_6011,* flanking hyg^R. Hyg^R	This study
pYUBud6015	pYUB854 containing upstream (*Afl*II/*Xba*I) and downstream regions of *Msmeg_6015,* respectively, were introduced either side of hyg^R *Bgl*II/*Nhe*I (down). Hyg^R	This study
pYUBudipdAB	pYUB854 containing upstream (*Afl*II/*Xba*I) region of *Msmeg_6002* and downstream region of *Msmeg_6003* (*Bgl*II/*Nhe*I) flanking hyg^R . Hyg^R	This study
pYUBudipdC	pYUB854 containing upstream (*Afl*II/*Xba*I) and downstream regions of *Msmeg_6004,* respectively, were introduced either side of hyg^R into pYUB854 using AflII/XbaI (up) and *Bgl*II/*Nhe*I (down). Hyg^R	This study
pYUBudipdAB-mtb	pYUB854 containing upstream (*Afl*II/*Xba*I) region of *rv3551* and downstream region of *rv3552* (*Bgl*II/*Nhe*I) flanking hyg^R . Hyg^R	This study
pYUBudipdC.mtb	pYUB854 containing upstream (*Afl*II/*Xba*I) region and downstream region of *rv3553* (*Bgl*II/*Nhe*I) flanking hyg^R . Hyg^R	This study

Strain or construct name	Description	Reference
pMV361.apr	Integrative plasmid with the constitutive *hsp60* promoter for complementation in mycobacteria. Apr^R	[257]
pMVEchA20	pMV361.apr containing *rv3550* inserted into the *Eco*RI/*Hind*III sites as a 766 bp fragment. A stop codon and RBS were added in the 5'UTR of the insert. Apr^R	This study
pMVFadE32	pMV361.apr containing *msmeg_6015* inserted into the *Eco*RI/*Hind*III sites as a 988 bp fragment. A stop codon and RBS were added in the 5'UTR of the insert. Apr^R	This study
pMVrv3559c	pMV361.apr containing *rv3559c* inserted into the *Eco*RI/*Hind*III sites as a 811 bp fragment. A stop codon and RBS were added in the 5'UTR of the insert. Apr^R	This study
pMVipdAB	pMV361.apr containing *rv3551-3552c* inserted into the *Eco*RI/*Hind*III sites as a 1632 bp fragment. A stop codon and RBS were added in the 5'UTR of the transcript. Apr^R	This study
pMVechA20-ipdC	pMV361.apr containing *rv3550-3553c* inserted into the *Eco*RI/*Hind*III sites as a 3536 bp fragment. A stop codon and RBS were added in the 5'UTR of the insert. Apr^R	This study
pET41b+	Expression vector for *E. coli*. Kan^R	Novagen
pETRv3559	pET41b+ containing *rv3559c* (*ipdF*) with codons introducing an N-terminal His_6 tag and TEV^{Pro} site following start codonas an 835 bp fragment inserted into *Nde*I/*Hind*III sites. Kan^R	This study
pTip-QC2	Expression vector for RHA1. Thiostrepton inducible promoter. Cam^R, Amp^R	[258]
pTipR1EchA20	pTip-QC2 Expression vector for $EchA20_{RHA1.}$ 840 bp fragment containing *RS27700* (*echA20*) with codons introducing a hexahistidine tag and TEV^{Pro} site following start codon. as an 840 bp fragment inserted into *Eco*RI/*Hind*III sites. Cam^R, Amp^R	This study

Strain or construct name	Description	Reference
pTipFadA6	pTip-QC2 containing *rv3556c* with codons introducing a hexahistidine tag and TEV^{Pro} site following start codon as a 1210 bp fragment inserted using *Eco*RI/*Hind*III into pTip-QC2. Cam^R, Amp^R	This study
pTipR1IpdAB	pTip-QC2 containing *RS22695-22690* with codons introducing a hexahistidine tag and TEV^{Pro} site following start codon as a 1709 bp fragment inserted using *Eco*RI/*Hind*III into pTip-QC2. Cam^R,Amp^R	This study
pTipIpdAB	pTip-QC2 containing *rv3551-2c* with codons introducing a hexahistidine tag and TEV^{Pro} site following start codon inserted using *Eco*RI/*Hind*III into pTip-QC2. Cam^R,Amp^R	This study
pTipRv3553	pTip-QC2containing *rv3553* inserted using *Eco*RI/*Hind*III into pTip-QC2. Cam^R, Amp^R	This study
pMAL-c2x	Expression vector for *E.coli* used to make maltose binding protein (MBP)-fusion proteins. IPTG inducible promoter. Amp^R	(NEB #E8000S)
pMALDOC21	Expression vector for MBP-$IpdC_{DOC21}$. 1116 bp fragment containing *DC0014_19*. Introduced into pMAL-c2x using *Eco*RI/*Hind*III adding *DC0014_19* in frame with MBP gene. Amp^R	This study

Appendix B X-ray crystallography data collection and statistics

B.1 KstR2: HIP-CoA crystallography statistics

PDB Code	4W97
Data collection	
Space group	C2
Cell dimensions	
a, *b*, *c,* Å	
β, °	72.5, 90.5, 49.8 129.7
Resolution, Å	25.00 – 1.60
R_{merge}[a]	0.032 (0.545)[b]
$I / \sigma(I)$	48.68 (3.78)
Completeness, %	100 (99.9)
Redundancy	4.4 (4.2)
Refinement	
Resolution, Å	24.85 – 1.60
No. of reflections: working, test	31023, 1650
R-factor/free *R*-factor[c]	16.0/19.9 (23.7/29.4)
No. of refined atoms	
Protein	1626
Substrate	64
Solvent	4
Water	310
B-factors, Å^2	
Protein	28.9
Substrate	37.9
Solvent	41.2
Water	43.5
r.m.s.d.	
Bond lengths, Å	0.02
Bond angles, °	2.14

[a] $R_{merge} = \sum_{hkl} \sum_j | I_{hkl,j} - \langle I_{hkl} \rangle | / \sum_{hkl} \sum_j I_{hkl,j}$

[b] Values in parentheses refer to highest resolution shells of 1.63-1.60 Å for data collection and 1.64-1.60 Å for refinement.

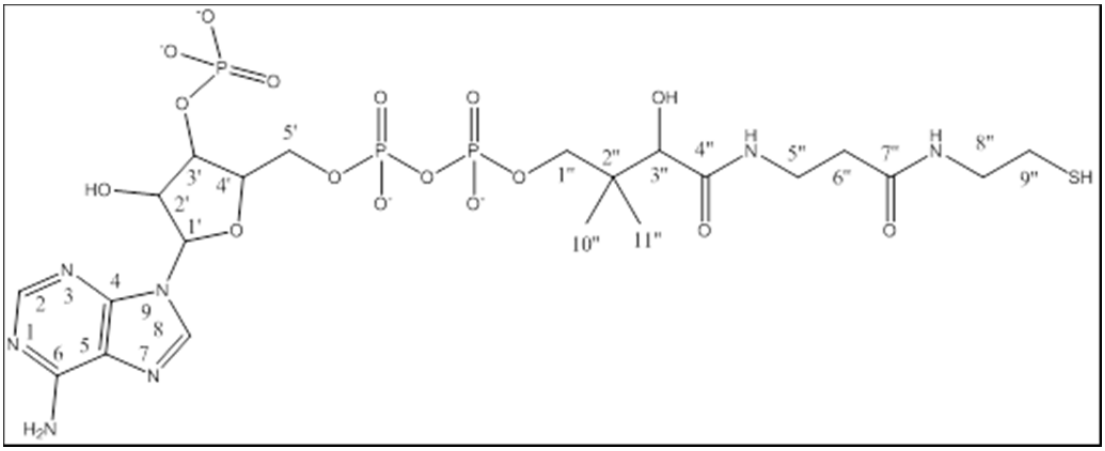
[c] R-factor = $\sum_{hkl} | F^o_{hkl} - F^c_{hkl} | / \sum_{hkl} F^o_{hkl}$. Free R-factor calculated with 5% reflections set aside.

B.2 IpdAB, IpdAB: COCHEA-CoA, and IpdAB E105^{A}A: COCHEA-CoA crystallography statistics

	$IpdAB_{RHA1}$ WT	$IpdAB_{RHA1}$ E105^{A}A·COCHEA-CoA	$IpdAB_{RHA1}$ WT·COCHEA-CoA
Wavelength	0.97948	0.97874	1.54
Resolution range	48.73 - 1.70 (1.76 - 1.70)	47.94 - 1.40 (1.45 - 1.40)	48.02 - 1.60 (1.66 - 1.60)
Space group	P 4_32_12	P 4_32_12	P 4_32_12
Unit cell (Å, Å, Å, °, °, °)	68.91 68.91 241.37 90 90 90	69.17 69.17 241.87 90 90 90	69.29 69.29 241.86 90 90 90
Total reflections	522304 (51305)	819431 (75359)	837439 (27240)
Unique reflections	65026 (6374)	116083 (11331)	77704 (6844)
Multiplicity	8.0 (8.0)	7.1 (6.7)	10.8 (4.0)
Completeness (%)	1.00 (1.00)	1.00 (0.99)	0.99 (0.89)
Mean I/sigma(I)	16.54 (2.31)	20.62 (1.53)	31.31 (2.13)
Wilson B-factor	18.06	16.09	21.51
R-merge	0.09076 (0.9778)	0.0541 (1.336)	0.0436 (0.4061)
R-meas	0.09699 (1.044)	0.05838 (1.448)	0.04563 (0.4612)
CC1/2	0.999 (0.818)	1.00 (0.624)	1.00 (0.934)
CC*	1.00 (0.949)	1.00 (0.877)	1.00 (0.983)
Reflections used in refinement	65021 (6374)	116047 (11327)	77599 (6823)
Reflections used for R-free	3311 (347)	5804 (567)	3838 (357)
R-work	0.1611 (0.2257)	0.1504 (0.2756)	0.1610 (0.4059)
R-free	0.1916 (0.2508)	0.1716 (0.2766)	0.1886 (0.3955)
CC(work)	0.971 (0.911)	0.978 (0.830)	0.973 (0.903)
CC(free)	0.962 (0.893)	0.971 (0.776)	0.964 (0.880)
Number of non-hydrogen atoms	4700	5014	4874
macromolecules	4148	4215	4183
ligands	79	78	78
Protein residues	551	543	542
RMS(bonds, Å)	0.006	0.006	0.006
RMS(angles, °)	0.83	0.92	0.93
Ramachandran favored (%)	97	97	98
Ramachandran allowed (%)	3	2.5	2.4
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	0.23	0.45	0.23
Clashscore	1.56	2.01	3.23
Average B-factor (Å^2)	23.6	25.4	27.0
macromolecules	21.3	22.0	24.3
ligands	71.7	66.3	71.9
solvent	35.7	40.4	39.6
Number of TLS groups	19	18	18

Appendix C Analyses of metabolites

CoASH



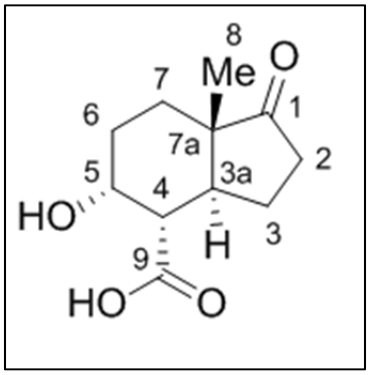
$\varepsilon_{260nm,\ pH\ 7.0} = 11.9\ mM^{-1}\ cm^{-1}$

[M+H]$^+$ (ESI-MS/MS) 768.1210 (261.1269, 428.0364); $C_{21}H_{37}N_7O_{16}P_3S^+$; λ_{max}= 258 nm

^{1}H-NMR (850 MHz, D_2O): δ= 8.69 (s,H,8), 8.29 (s,H,4), 6.19 (d,H,1'),4.85 (s,H,3''), 4.59 (s,H,4'), 4.20 (s,2H,5'), 4.03 (s,2H,2'), 3.81-3.35 (m,8H), 2.59 (t,2H,6''), 2.51 (t,2H,5''), 0.90 (s,3H,10''), 0.78 (s,3H,11'')

^{13}C-NMR (from ^{1}H-^{13}C HMBC/HSQC; 850 MHz, D_2O): δ= 178 (C4''), 177 (C7''), 158 (C6), 155 (C2), 153 (C4), 144 (C8), 122 (C5), 89 (C1'), 86 (C4'), 77, 75, 73, 68 (C5'^{c}), 45 (C8''), 41 (C2''), 38 (C5''^{c}), 38 (C6''^{c}), 31 (C9''^{c}), 23 (C11''), 21 (C10'')

3a*α*-*H*-4*α*(carboxyl)-5α-hydroxy-7a*β*-methylhexahydro-1-indanone (5α-OH HIC)



MP: 154.4-156^{o}C (colorless liquid); λ_{max}= 258 nm

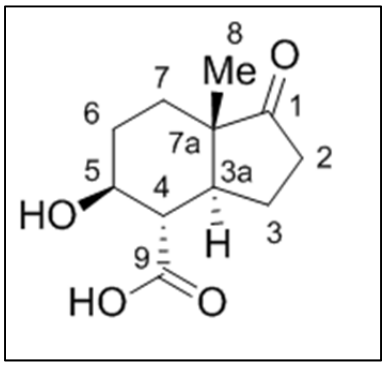
[M+H]$^+$ (ESI-MS/MS) 962.2159 (455.2209, 428.0368); $C_{32}H_{51}N_7O_{19}P_3S^+$

^{1}H-NMR (400 MHz, MeOD): δ= 4.30-4.27 (m,H,5), 2.67 (dd, *J* =12.2,2.8 Hz,H,4), 2.48-2.41 (m,H,2), 2.31-2.20 (m,2H,3a,3), 2.16-2.09 (m,H,2), 1.84-1.80 (m,2H,6), 1.68-1.58 (m,2H,3,7), 1.53-1.49 (m,H,7), 0.91 (s,3H,8)

^{13}C-NMR (100 MHz, MeOD): δ= 222 (C1), 177 (C9), 68 (C5), 49 (C7a), 48 (C4) 41 (C3a), 36 (C2), 30 (C6), 27 (C7), 24 (C3), 13 (C8)

2-TMS-5α-OH HIC GCMS R_t = 10.72 min. MS (70 eV, EI); *m/z*: 356 (4%), 341 (54%), 300 (11%), 147 (100%), 73 (78%)

3a*α*-*H*-4*α*(carboxyl)-5β-hydroxy-7a*β*-methylhexahydro-1-indanone (5β-OH HIC)



MP: 193-195°C (colorless liquid); λ_{max}= 258 nm

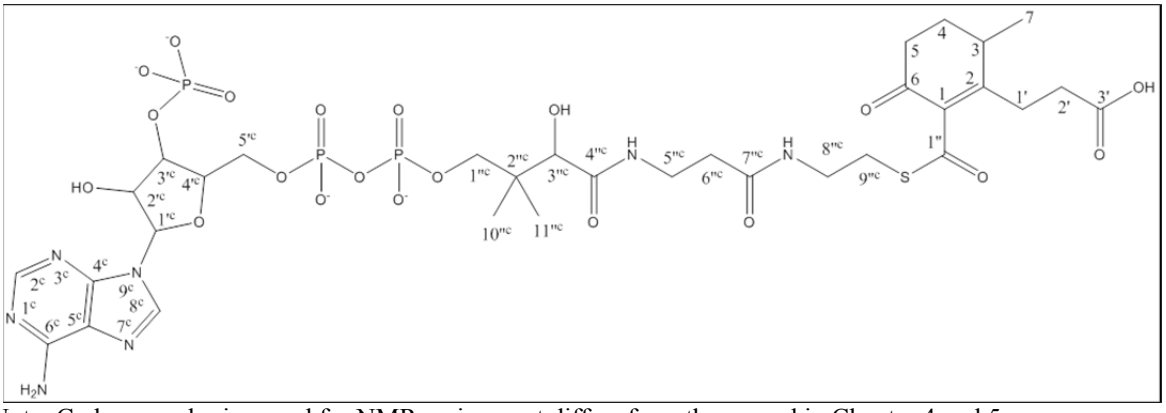
[M+H]$^+$ (ESI-MS/MS) 962.2170 (455.2209, 428.0367); $C_{32}H_{51}N_7O_{19}P_3S^+$

^{1}H-NMR (400 MHz, MeOD): δ= 3.66 (ddd, *J* = 11.2, 10.0,5.5 Hz,H,5), 2.52-2.45 (m,2H,4,2), 2.21-2.11 (m,H,2), 1.96-1.89 (m,H,6), 1.86-1.67 (m,4H,3a,3,7), 1.65-1.57 (m,H,6), 1.41-1.35 (m,H,7), 0.97 (s,3H,8)

^{13}C-NMR (100 MHz, MeOD): δ= 221 (C1), 177 (C9), 73 (C5), 52 (C4), 48 (C7a), 47 (C3a), 37 (C2), 31 (C6), 30 (C7), 23 (C3), 14 (C8)

2-TMS-5β-OH HIC GCMS R_t = 10.70 min. MS (70 eV, EI); *m/z*: 356 (9%), 341 (47%), 300 (15%), 147 (100%), 73 (96%)

2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA (COCHEA-CoA)



Note: Carbon numbering used for NMR assignment differs from those used in Chapter 4 and 5

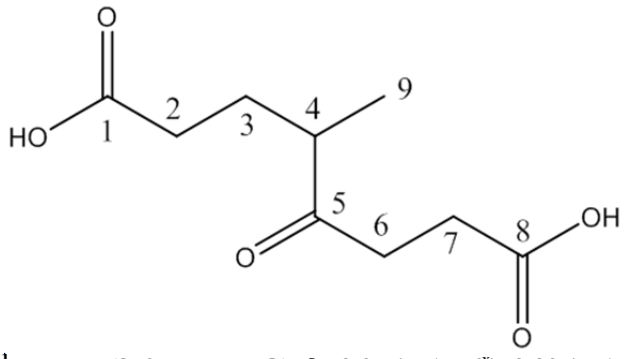
$\varepsilon_{260nm,\ pH\ 7.0}$ = 16.4 mM^{-1} cm^{-1} (calculated).

[M+H]$^+$ (ESI-MS/MS) 976.1960 (469.1994, 428.0358); $C_{32}H_{49}N_7O_{20}P_3S^+$; λ_{max}= 250 nm

^{1}H-NMR (850 MHz, D_2O): δ= 8.64 (s,H,8^{c}), 8.23 (s,H,4^{c}), 6.19 (d,H,1''^{c}),4.79 (s,H,3''^{c}), 4.63 (s,H,4''^{c}), 4.23 (s,2H,5''^{c}), 4.00 (s,2H,2''^{c}), 3.58-3.39 (m,7H), 3.16 (m,2H, 9''^{c}), 2.73 (m,H, 3), 2.65 (m,2H, 5,1'), 2.59 (t,2H, 2'), 2.44 (t,H,5''^{c}), 2.38 (m,H,5), 2.10 (m,H,4), 1.81 (m,H,4) 1.23 (d ^{3}J=9 Hz,3H,7), 0.90 (s,3H,10''^{c}), 0.78 (s,3H,11''^{c}); ^{c} denotes CoA moiety

^{13}C-NMR (from ^{1}H-^{13}C HMBC/HSQC; 850 MHz, D_2O): δ= 203 (C6), 201 (C1''), 182 (C3'), 177 (C4''c), 177 (C7''c), 175 (C2), 156 (C6c), 155 (C2c), 152 (C4c), 142 (C8c), 139 (C1), 121 (C5c), 89 (C1'c), 83 (C4'c), 77 (C3''c), 75 (C2'c), 73 (C3'c), 68, 62 (C5'c), 41 (C2''c), 38 (C5''c), 38 (C6''c), 36 (C5), 36 (C1'), 35 (C3), 32 (C2'), 31 (C9''c), 31 (C4), 24 (C11''c), 21 (C10''c), 19 (C7); c denotes CoA moiety

4-Methyl-5-oxo-octanedioc acid (MOODA)

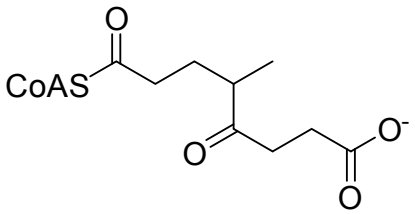


^{1}H NMR (850 MHz, D_2O): δ= 2.97 (m,1H,6^{x}), 2.89 (m,1H,6^{y}), 2.77 (m,1H,4), 2.61 (t,2H,8), 2.37 (t,2H,2), 1.94 (m,1H,3^{x}), 1.69 (m,1H,3^{y}), 1.13 (d,^{3}J= 7.1 Hz,3H, 9)

^{13}C NMR (from ^{1}H-^{13}C HMBC/ HSQC; 850 MHz, D_2O): δ= 221 (C5), 180 (C1), 179 (C8), 48 (C4), 38 (C6), 32 (C2), 30 (C7), 29 (C3), 18 (C9)

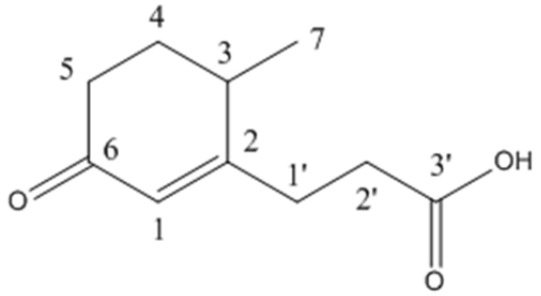
2-TMS-MOODA GCMS R_t = 10.02 min. MS (70 eV, EI); *m/z*: 346 (1%), 331 (12%), 241 (14%), 173 (100%), 125 (32%), 73 (82%)

4-Methyl-5-oxo-octanedioyl-CoA (MOODA-CoA)



[M+H] 952.1960 (445.1999, 428.0371) m/z; $C_{30}H_{49}N_7O_{20}P_3S^{+}$; λ_{max} = 258 nm
^{1}H-NMR (600 MHz, D_2O): δ(ppm) = 8.53 (s, 1H), 8.24 (s, 1H), 6.2 (d, 1H), 4.65 (s, 1H), 4.2 (s, 2H), 4.0 (s, 2H), 3.8 (m, 1H), 3.6-3.5 (m, 2H), 3.50 (q, 2H), 3.49 (m, 3H), 3.30-3.45 (m, 4H), 2.95 (q, 2H), 2.83 (m, 1H), 2.73 (m, 1H), 2.68 (m, 1H), 2.58 (q, 2H), 2.3-2.4 (m, 4H), 1.64 (m, 1H), 1.05 (d, 3H), 0.86 (s, 3H), 0.72 (s, 3H)

2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1

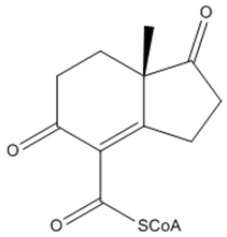


1-TMS-COCHE GCMS R_t = 9.32 min. MS (70 eV, EI); *m/z*: 254 (83%), 239 (25%), 221 (59%), 197 (40%), 137 (92%), 122 (100%)

^{1}H-NMR (600 MHz, D_2O): δ= 5.8 (s,1H,1), 2.6 (m,6H,5,3',1'), 2.4 (m,1H,3), 2.1 (m,1H,4^x), 1.8 (m,1H,4^y), 1.2 (d,^{3}J= 8 Hz,3H,7)

^{13}C-NMR (from HSQC/HMBC; 600 MHz, D_2O): δ= 205 (C6), 180 (C3'), 174 (C2), 128 (C1), 36 (C5), 34 (C3), 34 (C2'), 32 (C4), 32 (C1'), 17 (C7)

(7aS)-7a-Methyl-1,5-dioxo-2,3,5,6,7,7a-hexahydro-1H-indene-4-carboxyl-CoA (HIEC-CoA)



[M+H] 958.1899 (451.1914, 428.0378) m/z; $C_{32}H_{47}N_7O_{19}P_3S^+$; λ_{max} = 253 nm