

Stony Brook University



OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

A Dissertation Presented
by
Suzanne T. Thomas
to
The Graduate School
in Partial Fulfillment of the
Requirements
for the Degree of
Doctor of Philosophy
in
Chemistry

Stony Brook University

August 2012

Stony Brook University

The Graduate School

Suzanne T. Thomas

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

**Nicole S. Sampson, Ph. D., Dissertation Advisor
Department of Chemistry**

**Orlando D. Schärer, Ph. D., Chairperson
Department of Pharmacological Sciences**

**Jessica C. Seeliger, Ph. D., Third Member
Department of Pharmacological Sciences**

**Antonius Koller, Ph. D., Outside Member
Department of Pathology
Stony Brook University**

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

**Inhibitors of 3 β -Hydroxysteroid Dehydrogenase and
Metabolic and Enzymatic Characterization of the Cholesterol Metabolizing
Intracellular Growth Operon of *Mycobacterium tuberculosis***

By

Suzanne T. Thomas

Doctor of Philosophy

In Chemistry

Stony Brook University

2012

New drugs with novel mechanisms of action are required to meet the severe threat to human health posed by the emergence of multidrug and extensively drug resistant strains of *Mycobacterium tuberculosis* (*M. tuberculosis*). The cholesterol metabolism pathway in *M. tuberculosis* is a potential source of energy as well as secondary metabolite production that is important for survival of *M. tuberculosis* in the host macrophage. Oxidation and isomerization of 3 β -hydroxysterols to 4-en-3-ones by *M. tuberculosis* 3 β -hydroxysteroid dehydrogenase (3 β -HSD) is required for sterol metabolism, and inhibitors of 3 β -HSD are important for targeting the cholesterol metabolic pathway. In this work, we evaluated a series of azasteroids for inhibition of 3 β -HSD. Our structure–activity studies indicate that the 6-aza version of cholesterol is the best and tightest binding competitive inhibitor ($K_i = 100$ nM) of the steroid substrate and are consistent with cholesterol being the preferred substrate of *M. tuberculosis* 3 β -HSD.

The *intracellular growth (igr)* operon is required for *in vitro* growth using cholesterol as a sole carbon source. The function of *igr* operon and its role in cholesterol metabolism is yet to be established. Here we describe the biosynthetic preparation of isotopically labeled ¹³C- [1,7,15,22,26]-cholesterol and employ it as a tool to investigate the

cholesterol-derived metabolite profile of the *M. tuberculosis* H37Rv Δ *igr* mutant strain by high resolution LC/MS. Culture supernatants from the *igr* mutant accumulate a cholesterol-derived metabolite not observed in H37Rv wild type or complemented strains. Multidimensional NMR and mass spectral analysis revealed the structure of this cholesterol-derived catabolite to be a late stage metabolic product: methyl 1 β -(2'-propanoate)-3 α -H-4 α -(3'-propanoic acid)-7 α β -methylhexahydro-5-indanone.

The computationally annotated functions of the six genes of the *igr* operon are a lipid transfer protein (*ltp2/Rv3540c*), two (*R*)-specific enoyl-CoA hydratases (*Rv3541c* and *Rv3542c*), two acyl-CoA dehydrogenases (*fadE29/Rv3543c* and *fadE28/Rv3544c*), and a cytochrome P450 (*cyp125/Rv3545c*). Heterologous expression in *E. coli* demonstrated that FadE28 forms a heteromeric complex with FadE29, and that likewise, Rv3542c forms a heteromeric complex with Rv3541c. Biophysical characterization of each complex established they form novel $\alpha_2\beta_2$ heterotetramers. Using synthetic substrates analogous to the metabolite identified in *M. tuberculosis* H37Rv Δ *igr* mutant strain, we verified the catalytic activity of the purified, recombinant FadE28-FadE29 and Rv3541c-Rv3542c protein complexes to be dehydrogenation and hydration of the 2'-propanoate-CoA side chain. We conclude the *igr* operon is required for degradation of the 2'-propanoate side chain fragment during metabolism of cholesterol by *M. tuberculosis*.

Table of Contents

List of Figures	viii
List of Tables	x
List of Schemes	xi
List of Abbreviations	xii
Chapter 1 - Introduction	1
I. <i>Mycobacterium tuberculosis</i> infection and treatment	2
II. Intracellular metabolism of <i>M. tuberculosis</i>	5
III. Cholesterol metabolism is important for <i>M. tuberculosis</i> pathogenesis	7
IV. The cholesterol metabolism pathway	9
V. <i>M. tuberculosis</i> 3 β -hydroxysteroid dehydrogenase	20
VI. The <i>intracellular growth</i> operon	22
VII. Cytochrome P450 Cyp125	24
VIII. Acyl-CoA dehydrogenases	25
IX. 2-Enoyl-CoA hydratases	29
X. Specific aims	32
Chapter 2 - Results and Discussion	34
I. Inhibition of the <i>M. tuberculosis</i> 3 β -hydroxysteroid dehydrogenase by azasteroids	35
II. Pathway profiling in <i>M. tuberculosis</i> : elucidation of a cholesterol-derived catabolite and the enzymes that catalyze its metabolism	43
III. Biochemistry of acyl-CoA dehydrogenase FadE28-FadE29 from <i>M. tuberculosis</i>	62
IV. Biochemistry of 2-enoil-CoA hydratase Rv3541c-Rv3542c from <i>M. tuberculosis</i>	76
Chapter 3 - Experimental Methods	85
I. Materials	87

II.	General Methods	87
III.	Instrumentation	88
IV.	Buffers	88
V.	Expression plasmid construction and plasmid table	89
VI.	Expression and protein purification	90
VII.	In-gel tryptic digestion and MALDI-TOF mass fingerprinting	94
VIII.	Azasteroids	94
IX.	3 β -HSD assay	94
X.	Isotopic labeling of LDL-cholesterol	95
XI.	Metabolic labeling of <i>M. tuberculosis</i>	96
XII.	Growth of <i>M. tuberculosis</i> in the presence of free cholesterol	96
XIII.	Biochemical analysis of LDL-derived [¹⁴ C]-lipids	97
XIV.	Biochemical extraction and analysis of cholesterol-derived bacterial lipid metabolite	97
XV.	Purification and characterization of cholesterol-derived metabolite 1	98
XVI.	FadE28-FadE29 substrate synthesis	98
XVII.	Dehydrogenase assay	100
XVIII.	Hydratase assay	100
XIX.	Identification and quantification of flavin cofactor	100
XX.	Analytical gel-filtration	100
XXI.	Analytical ultracentrifugation sedimentation equilibrium	101
XXII.	LC/UV/MS of intact protein complexes	101
Chapter 4 - Conclusions and Future Perspectives		103
I.	6-azasteroid inhibitors of <i>M. tuberculosis</i> cholesterol metabolism	104
II.	Cholesterol-derived metabolite profiling with [1,7,15,22,26- ¹³ C]-cholesterol	105

III.	Investigating the FadE proteome of <i>M. tuberculosis</i>	107
IV.	Studies of heteromeric (<i>R</i>)-hydratases	108
V.	Final conclusions	110
Chapter 5 - References		111
Appendix		132
	Expression study of <i>vaccinia virus</i> 3 β -HSD	133

List of Figures

1-1	Structure of cholesterol	8
1-2	FadEs in cholesterol metabolism	13
1-3	The <i>igr</i> operon	23
1-4	Acyl-CoA dehydrogenase crystal structures	27
1-5	The catalytic base of acyl-CoA dehydrogenase	28
1-6	Enoyl-CoA hydratase crystal structures	31
2-1	Azasteroid frameworks	36
2-2	Kinetics of <i>M. tuberculosis</i> 3 β -HSD inhibition by 7 at varied concentrations of NAD ⁺	41
2-3	Kinetics of <i>M. tuberculosis</i> 3 β -HSD inhibition by 7 at varied concentrations of DHEA	42
2-4	¹³ C- or ¹⁴ C- metabolic labeling of LDL-derived cholesterol	44
2-5	Analysis of metabolically labeled LDL-[¹³ C]- or -[¹⁴ C]-cholesterol	46
2-6	TLC of <i>M. tuberculosis</i> extracts cultured with LDL-[1,7,15,22,26- ¹⁴ C]-cholesterol.	47
2-7	LC/MS analysis of <i>M. tuberculosis</i> H37Rv, Δ <i>igr</i> , and <i>igr</i> complement extracts cultured with LDL-[¹³ C]-cholesterol and natural abundance LDL-cholesterol	49
2-8	Assigned structures of Δ <i>igr</i> metabolite ions identified by LC/MS/MS	51
2-9	Structural characterization of Δ <i>igr</i> metabolite 1	55
2-10	MS/MS/MS fragmentation analysis of Δ <i>igr</i> metabolite 1	56
2-11	Expression and purification of the <i>igr</i> operon	58
2-12	UV-visible spectra of purified FadE29 and FadE28-FadE29	63
2-13	Tryptic peptides from FadE28 and FadE29	64
2-14	Analytical size exclusion chromatography of FadE29 and FadE28-FadE29	65

2-15	Analytical ultracentrifugation sedimentation equilibrium of FadE29 and FadE28-FadE29	66
2-16	Reverse phase LC/UV chromatogram of FadE28-FadE29	66
2-17	UV/LC/MS of isolated flavin from FadE28-FadE29	67
2-18	FAD binding in acyl-CoA dehydrogenases	69
2-19	MALDI-TOF MS spectra of FadE28-FadE29 assay	71
2-20	Bioinformatic analysis of the FadE proteome of <i>M. tuberculosis</i>	73
2-21	Regulation of <i>M. tuberculosis fadE</i> genes	74
2-22	Tryptic peptides from Rv3541c and Rv3542c	79
2-23	Analytical size exclusion chromatography of Rv3541c-Rv3542c	80
2-24	Analytical ultracentrifugation sedimentation equilibrium of Rv3541c-Rv3542c	81
2-25	Reverse phase LC/UV chromatogram of Rv3541c-Rv3542c	81
2-26	Substrates of Rv3541c-Rv3542c	82
2-27	Detection of Rv3541c-Rv3542c assay products by MALDI-TOF MS	83
A-1	SDS-PAGE of purified <i>vaccinia virus</i> 3 β -HSD- MBP	134
A-2	Codon optimized gene sequence of <i>vaccinia virus A44L</i>	135

List of Tables

1-1	Current Tb treatments and drug candidates currently in clinical trials	4
1-2	Cholesterol side-chain metabolism	15
1-3	Cholesterol ring system metabolism	20
1-4	Structural information for (<i>R</i>)-hydratases	32
2-1	Inhibition of <i>M. tuberculosis</i> 3 β -HSD by azasteroids	40
2-2	Inhibition characteristics of azasteroid 7	41
2-3	Cholesterol-derived metabolite parent ions (MH ⁺ or MNa ⁺) identified using LC/MS/MS	50
2-4	Specific activity data for FadE28-FadE29	60
2-5	The (<i>R</i>)-hydratase motif	77
3-1	Plasmids used in this work	90
3-2	Calculated extinction coefficients for proteins studied in this work	102
A-1	Expression constructs for <i>vaccinia virus</i> 3 β -HSD	134

List of Schemes

1-1	Flux of metabolites from cholesterol catabolism	7
1-2	Metabolic tracing to study cholesterol metabolism in <i>M. tuberculosis</i>	9
1-3	β -oxidation enzymes are encoded by multiple copies in the <i>M. tuberculosis</i> genome	11
1-4	Cholesterol side-chain metabolism	14
1-5	Cholesterol ring system metabolism	19
1-6	(<i>R</i>)-hydratase is important for the metabolism of fatty acids with <i>cis</i> double bonds	30
2-1	Partial cholesterol metabolism pathway in <i>M. tuberculosis</i> .	36
2-2	Reaction catalyzed by 3 β -HSD	37
2-3	Proposed function of the <i>igr</i> operon	59

Abbreviations

(<i>R</i>)-hydratase	(<i>R</i>)-specific enoyl-CoA hydratase
(<i>S</i>)-hydratase	(<i>S</i>)-specific enoyl-CoA hydratase
2 × YT	2 × yeast extract tryptone broth
3-HSA	3-hydroxy-9,10-seconandrost-1,3,5(10)triene-9,17-dione
3,4-DHSA	3-hydroxy-9,10-seconandrost-1,3,5(10)triene-9,17-dione
3β-HSD	3β-hydroxysteroid dehydrogenase
4,9-DHSA	4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10), 2-diene-4-ioc acid
9-OH-ADD	9α-hydroxy-androsta-1,4-diene-3,17-dione
9-OH-AD	9α-hydroxy-4-androstene-3,17-dione
<i>A. caviae</i>	<i>Aeromonas caviae</i>
ACAD	acyl-CoA dehydrogenase
AD	androst-4-ene-3,17-dione
ADD	androst-1,4-diene-3,17-dione
AIDS	acquired immune deficiency syndrome
amp	ampicillin
ATP	adenosine triphosphate
BSA	bovine serum albumin
BVMO	Baeyer-Villiger monooxygenase
<i>C. testosteroni</i>	<i>Comamonas testosteroni</i>
cam	chloramphenicol
cfu	colony forming units
CH ₂ Cl ₂	methylene chloride
CH ₃ CN	acetonitrile
CHCl ₃	chloroform

ChoX	cholesterol oxidase
CoA	coenzyme A
COSY	correlation spectroscopy
cpm	counts per minute
CV	column volumes
DEPT	distortionless enhancement by polarization transfer
DHEA	dehydroepiandrosterone
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DOHNAA	9,17-dioxo-1,2,3,4,10,19-hexanoandrostan-5-ic acid
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
ECH	enoyl-CoA hydratase
eq	equivalents
ESI	electrospray ionization
Et ₂ O	diethyl ether
Et ₃ N	triethylamine
EtOAc	ethyl acetate
EtOH	ethanol
Fad	fatty acid degrading
FAD	flavin adenine dinucleotide
GD	glutaryl-CoA dehydrogenase
GMC	glucose-methanol-choline
HDD	2-hydroxy-hexa-2,4-dienoic acid
HEPES	hydroxyethyl piperazineethanesulfonic acid
His ₆	hexahistidine
HIV	human immunodeficiency virus

HMBC	heteronuclear multiple bond correlation
HMG-CoA	3-hydroxy-3-methylglutarate coenzyme A
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
hygro	hygromycin
i2VD	short/branched chain acyl-CoA dehydrogenase
i3VD	isovaleryl-CoA dehydrogenase
iBD	isobutyryl-CoA dehydrogenase
IC ₅₀	half maximal inhibitory concentration
icl	isocitrate lyase
<i>igr</i>	intracellular growth
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-thiogalactoside
kan	kanamycin
kDa	kilodalton
K _{ic}	competitive inhibition constant
K _{iu}	uncompetitive inhibition constant
K _M	Michaelis-Menten constant
LB	Luria-Bertani
LC	liquid chromatography
LCAD	long chain acyl-CoA dehydrogenase
LDL	low density lipoprotein
Ltp	lipid transfer protein
<i>M. smegmatis</i>	<i>Mycobacterium smegmatis</i>
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
MALDI	matrix assisted laser desorption ionization
MaoC	monoamine oxidase C

MCAD	medium chain acyl-CoA dehydrogenase
MDR-Tb	multi-drug resistant Tb
MeOH	methanol
MFE	multifunctional enzyme
MS	mass spectrometry
MWCO	molecular weight cut off
NAD+	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
OADC	oleate-albumin-dextrose-NaCl-catalase
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDIM	phthiocerol dimycocerosate
PHA	polyhydroxyalkanoate
<i>R. equi</i>	<i>Rhodococcus equi</i>
<i>R. jostii</i>	<i>Rhodococcus jostii</i>
rpm	revolutions per minute
rt	room temperature
SAR	structure-activity relationship
SCAD	short chain acyl-CoA dehydrogenase
SDR	short chain dehydrogenase/reductase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SQ	single quadrupole
TAPS	N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
Tb	tuberculosis
TCEP	tris(2-carboxyethyl)phosphine

TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIC	total ion chromatogram
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
TOF	time of flight
UV	ultraviolet
VLCAD	very long chain acyl-CoA dehydrogenase
VLCAD2	very long chain 2 acyl-CoA dehydrogenase
WT	H37Rv wild type
XDR-Tb	extensively drug resistant Tb
<i>Δcyp125</i>	<i>cyp125</i> knockout
<i>Δigr</i>	<i>igr</i> operon knockout

Chapter 1

Introduction

I.	<i>Mycobacterium tuberculosis</i> infection and treatment	2
II.	Intracellular metabolism of <i>M. tuberculosis</i>	5
III.	Cholesterol metabolism is important for <i>M. tuberculosis</i> pathogenesis	7
IV.	The cholesterol metabolism pathway	9
V.	<i>M. tuberculosis</i> 3 β -hydroxysteroid dehydrogenase	20
VI.	The <i>intracellular growth</i> operon	22
VII.	Cytochrome P450 Cyp125	24
VIII.	Acyl-CoA dehydrogenases	25
IX.	2-Enoyl-CoA hydratases	29
X.	Specific aims	32

I. *Mycobacterium tuberculosis* infection and treatment

Tuberculosis (Tb), caused by *Mycobacterium tuberculosis* (*M. tuberculosis*), is a serious worldwide public health threat. One third of the world's population is infected with the disease, with approximately 9 million new cases occurring each year. Over the past two centuries one billion people have died from Tb infection (1) and approximately 1.5 million deaths each year are attributed to the disease, making it the second leading cause of death by an infectious disease (2, 3). The HIV epidemic led to an increase in active Tb infections and mortality rates. 10% of individuals infected with *M. tuberculosis* are also HIV positive. In recent years, drug-resistant strains of *M. tuberculosis* have emerged that make treating the infection more challenging. Novel treatments to combat resistant bacteria are urgently needed (4, 5).

Tuberculosis infection is spread through inhalation of aerosols containing the bacilli. The infection is highly contagious and a single digit number of bacilli can establish infection within a host. It is predicted that every person with an active infection will infect approximately 15 additional people.

The primary site of infection is the respiratory system. In the lung, bacilli are phagocytosed by alveolar macrophages. *M. tuberculosis* is able to inhibit maturation of the phagosome, preventing formation of the phagolysosome and acidification of the vacuole (5). During the early stages of infection, the bacilli rapidly divide and increase in number a thousand times (6). The host immune response results in accumulation of immune cells and the formation of a granuloma or tubercle, a major feature of Tb infection (7). The macrophage differentiates to form foamy macrophages that are loaded with lipid droplets within the core of the granuloma (8). A fibrous cuff surrounds the granuloma with lymphocytes at the perimeter. At this stage the bacteria can persist within the host granuloma for decades and the individual is considered to have latent Tb (9). During latent infection, the bacilli exist in a dormant, low-growth state and infected individuals are not contagious (10).

Approximately 10% of latent Tb infections will progress to active infection at some point during the host's lifetime. This conversion is triggered when the immune system is compromised and can be caused by AIDS, poor nutrition, and old age. Death of macrophages causes caseation at the core of the granuloma (7). Necrosis then occurs

leading to the release of bacteria into the airway. The individual is now contagious and symptomatic.

Antibiotics, such as streptomycin, and improved sanitation had a great impact on reducing the incidence of Tb in the 1950's. *M. tuberculosis* remains difficult to treat, however, because of its intrinsic drug resistance mechanisms, including drug efflux and low permeability due to the mycolic acids in the cell wall (11, 12). Current Tb therapies require six months of treatment with a combination of several antibiotics. The front-line drugs used to treat the infection are isoniazid, which targets acyl-carrier protein reductase, InhA, preventing cell wall mycolic acid synthesis, and rifampin, which targets RNA polymerase, inhibiting RNA synthesis (Table 1-1). Others included in the 6-month regimen are pyrazinamide and ethambutol.

The length of treatment causes problems with patient compliance, which can lead to resistant bacteria. Bacteria acquire resistance due to spontaneous mutations, which are selected for during inadequate treatment (13). The emergence of multi-drug resistant Tb (MDR-Tb) (3), caused by bacteria resistant to isoniazid and rifampin, and extensively drug resistant Tb (XDR-Tb) (14, 15), caused by bacteria resistant to these front line drugs as well as second line treatments, make new drugs to combat the infection a necessity. The current drug regimen is able to cure 95% of individuals with drug-susceptible Tb while cure rates of MDR-Tb drop to 50 to 70% and require 18 to 24 months of treatment with 8 to 10 drugs (16). Most worrisome is infection with XDR-Tb, which is very difficult to treat and for which mortality rates are very high. According to the World Health Organization, XDR-Tb has been identified in over 77 countries worldwide (17). Several drug candidates are currently in the pipeline for Tb treatment, but no new treatments have been implemented in over 40 years (Table 1-1) (18).

Table 1-1. Current Tb treatments and drug candidates currently in clinical trials.

Drug	Target	Clinical Phase	Ref
Bedaquiline (TMC207)	ATP Synthesis	II	(19)
Isoniazid	Cell Wall Biosynthesis	Current Tb drug	(20)
Ethambutol	Cell Wall Biosynthesis	Current Tb drug	(21)
BTZ043	Cell Wall Biosynthesis	I	(22)
PA824	Cell Wall Biosynthesis	II	(23)
OPC67683	Cell Wall Biosynthesis	III	(24)
Gatifloxacin	DNA Gyrase Inhibitor	III	(25)
Mozifloxacin	DNA Gyrase Inhibitor	III	(26)
Streptomycin	Protein Synthesis	Current Tb drug	(27)
Pyrazinamide	Protein Synthesis	Current Tb drug	(28)
PNU100480	Protein Synthesis	I	(29)
AZD5847	Protein Synthesis	I	
Rifampin	RNA Synthesis	Current Tb drug	(30)
SQ109	Cell Wall Biosynthesis	II	(31)
Sudoterb	Unknown	II	

A majority of Tb infections exist in a latent phase, in which Tb bacilli survive within the host. Current antibiotics, as well as the majority of drug candidates in clinical trials, target processes of active infection (Table 1-1). Therefore, they are ineffective in treating slow growing or dormant bacteria. One exception is bedaquiline, a novel diarylquinolone drug targeting ATP synthase. Bedaquiline has recently been shown to be as effective against dormant bacteria as replicating *M. tuberculosis* (32, 33). ATP synthase activity, though downregulated during dormancy, was shown to be essential for bacilli survival. This is the first *M. tuberculosis* drug or drug candidate demonstrating activity against non-replicating bacilli, making it a promising antimycobacterial agent. Additional drugs that target latent *M. tuberculosis* are necessary to reduce the number of infected individuals worldwide.

Sequencing of the *M. tuberculosis* genome reported in 1998, has provided a wealth of information that can be used for understanding Tb pathogenesis (4, 34). The genome

contains 4,000 genes and the function of 45% were predicted based on similarity to known genes. A large portion of the genome is not annotated and 16% have no similarity to known proteins (34, 35). Gaining functional information for the large number of uncharacterized genes in the *M. tuberculosis* genome is a daunting task. However, discovering the biological role of these genes will lead to a better understanding of the pathogenesis of the bacteria and may result in the identification of novel virulence factors and drug targets for Tb treatment.

II. Intracellular metabolism of *M. tuberculosis*

The adaptation of *M. tuberculosis* to the intracellular phagosome environment is important for pathogenesis and enables the bacilli to reside within the host for decades. In order to survive, bacteria require at least one carbon source for the production of energy and biosynthetic intermediates. *In vitro*, *M. tuberculosis* readily utilizes glycerol and glucose as carbon and energy sources. During infection, the human body is the source of nutrition for the bacteria and the availability of such nutrients in the phagosome may be limited. Several lines of evidence suggest that during this time *M. tuberculosis* shifts its metabolism to utilize preferentially host-derived lipids for nutrition.

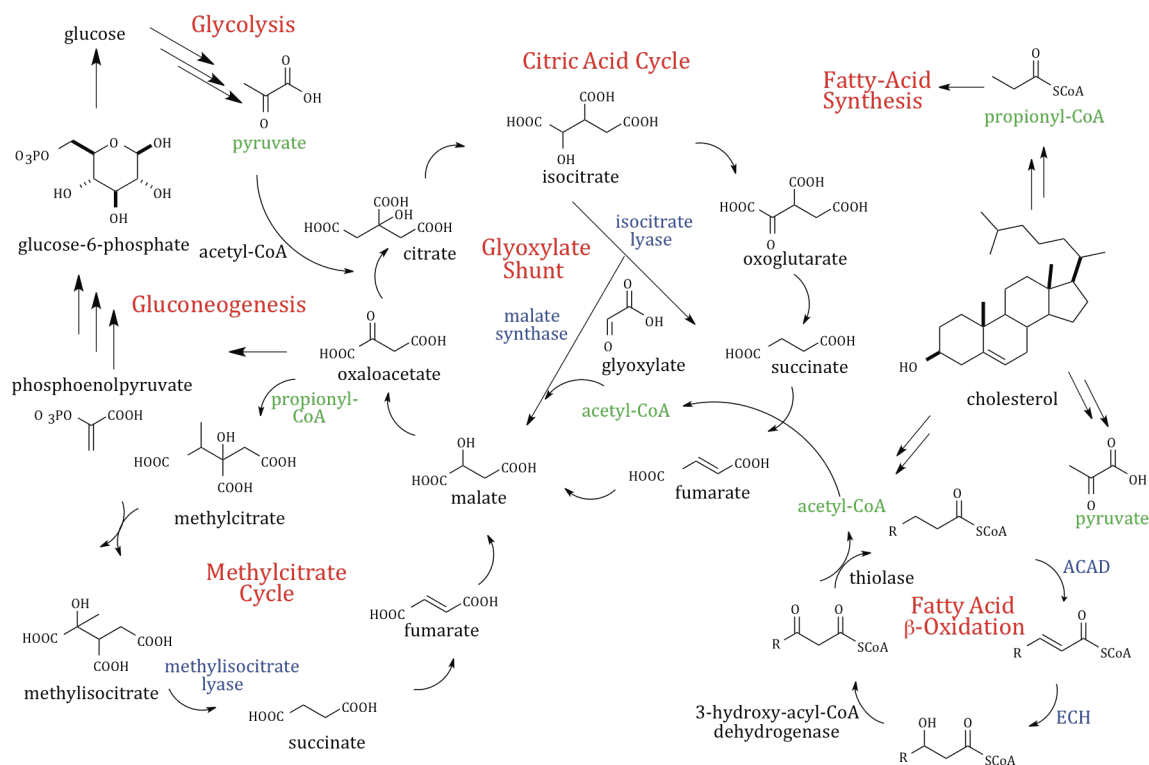
Seminal studies by Segal and Bloch demonstrated that bacilli freshly isolated from mouse lungs preferentially metabolized fatty acids over carbohydrates (36). This was the first evidence that fatty acids may play an important metabolic role in *M. tuberculosis* during infection. The importance of fatty acid metabolism is supported by the observation that the genome of *M. tuberculosis* contains a large number of genes, approximately 250, that are annotated as being involved in lipid and fatty acid metabolism (34). Global expression profiling experiments have shown that many of the hypothesized lipid-metabolizing genes are upregulated during infection of macrophages and mice (37, 38). In addition lipid-metabolizing genes were shown to be upregulated in Tb granulomas excised from Tb patients (39).

The *M. tuberculosis* genome contains two isocitrate lyase (*icl*) genes and one malate synthase gene (40). Isocitrate lyase and malate synthase are essential enzymes in the glyoxylate shunt, a pathway required for growth on fatty acids as a carbon source (Scheme 1-1). Acetyl-CoA, generated by fatty acid β -oxidation is converted to oxaloacetate through

the glyoxylate shunt and citric acid cycle. Oxaloacetate can be further transformed into carbohydrates by gluconeogenesis. Double knockout of *icl1* and *icl2* blocks growth on fatty acids as a carbon source *in vitro* while growth on glycerol and glucose are not impaired, indicating *icl1/icl2* are strictly utilized for fatty acid metabolism (41). In addition, an infection is not established in mice dosed with the double mutant, indicating *icl1* and *icl2* are essential genes (42). Evidence suggests that Icl1 may also function as a 2-methyl isocitrate lyase as part of the methyl citrate cycle (43). This pathway is necessary for propionate metabolism produced from the β -oxidation of odd chain fatty acids (Scheme 1-1) (44).

The granulomas in which *M. tuberculosis* resides during infection contains foamy macrophages, which are known to be lipid rich (9). *M. tuberculosis* is able to migrate towards the lipid bodies when engulfed within foamy macrophages. It is proposed that host lipids in foamy macrophages may supply the bacteria with the nutrients necessary for persistence (9).

Carbon metabolism by bacteria *in vivo* is a fundamental process that is key to survival and virulence (45). Metabolism *in vivo*, surprisingly, is poorly understood and *M. tuberculosis* metabolism is no exception (46). Many genes annotated as being involved in *M. tuberculosis* lipid and fatty acid metabolism have not been studied and their role in virulence yet to be substantiated. Biochemical knowledge of carbon metabolism pathways is necessary in order to target specific enzymatic conversions with therapeutics.



Scheme 1-1. Flux of metabolites from cholesterol catabolism. The metabolism of cholesterol results in the production of acetyl-CoA, propionyl-CoA, and pyruvate. Acetyl-CoA can enter into the glyoxylate shunt, allowing the bacteria to produce oxaloacetate, which can enter into gluconeogenesis. Propionyl-CoA can be utilized to synthesize fatty acids or enter into the methylcitrate cycle followed by the citric acid cycle to generate ATP and biosynthetic intermediates. Pyruvate can be decarboxylated to form acetyl-CoA. Shown is the proposed flux of metabolites from cholesterol degradation. Pathways are shown in red, key enzymes discussed in the text are shown in blue, and cholesterol metabolites, acetyl-CoA, propionyl-CoA, and pyruvate are shown in green. ECH, enoyl-CoA hydratase; ACAD, acyl-CoA dehydrogenase.

III. Cholesterol metabolism is important for *M. tuberculosis* pathogenesis

A growing body of evidence indicates that *M. tuberculosis* utilizes cholesterol (Figure 1-1) during persistent infection *in vivo* and this ability may be important for pathogenesis. It is well established that *M. tuberculosis* can grow on cholesterol as a carbon source *in vitro*. *In vivo* *M. tuberculosis* has access to adequate amounts of cholesterol. Lipids comprising the foamy macrophages, where *M. tuberculosis* resides during infection, include cholesterol, cholesterol esters, triacylglycerol, and lactosylceramide, all components of low density lipoprotein (LDL), the major cholesterol transport molecule in the blood. LDL

particles contain 8% cholesterol and 42% cholesterol esters by weight (47). Confocal microscopy of *M. tuberculosis* infected macrophages, stained with the cholesterol specific stain, filipin, showed that the bacilli reside in close proximity to host cholesterol (37).

Moreover, *M. tuberculosis* is able to import cholesterol via the ABC transporter encoded by the *mce4* operon (48). Mutations in the *mce4* operon were detrimental to the growth of the bacterium on cholesterol, indicating its importance for the ability to utilize cholesterol. Furthermore, cholesterol import is critical for persistence of the bacteria. The *mce4* knockout strain is impaired in growth during the chronic phase of infection in the mouse model and macrophages.

The use of isotopically labeled carbon sources is a powerful technique for studying metabolic pathways, allowing one to trace the fate of carbon through a given system. Metabolic studies using commercially available isotopically labeled [4-¹⁴C] and [26-¹⁴C] cholesterol have been conducted to trace the fate of these carbons during metabolism *in vitro* (Scheme 1-2) (49). The C4 of ring A is converted to CO₂ while C26 of the side chain is incorporated into mycobacterial lipids, such as phthiocerol dimycocerosate (PDIM). The CO₂ is presumably generated during the formation of ATP in the citric acid cycle. C25-C26-C27 are likely metabolized to propionyl-CoA, which can enter into biosynthetic pathways, including lipid biosynthesis as well as the methyl citrate pathway to feed the citric acid cycle (Scheme 1-1 and Scheme 1-2). Therefore, *M. tuberculosis* may use cholesterol as a carbon and energy source during infection.

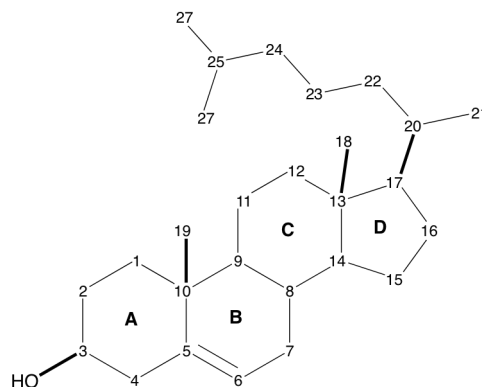
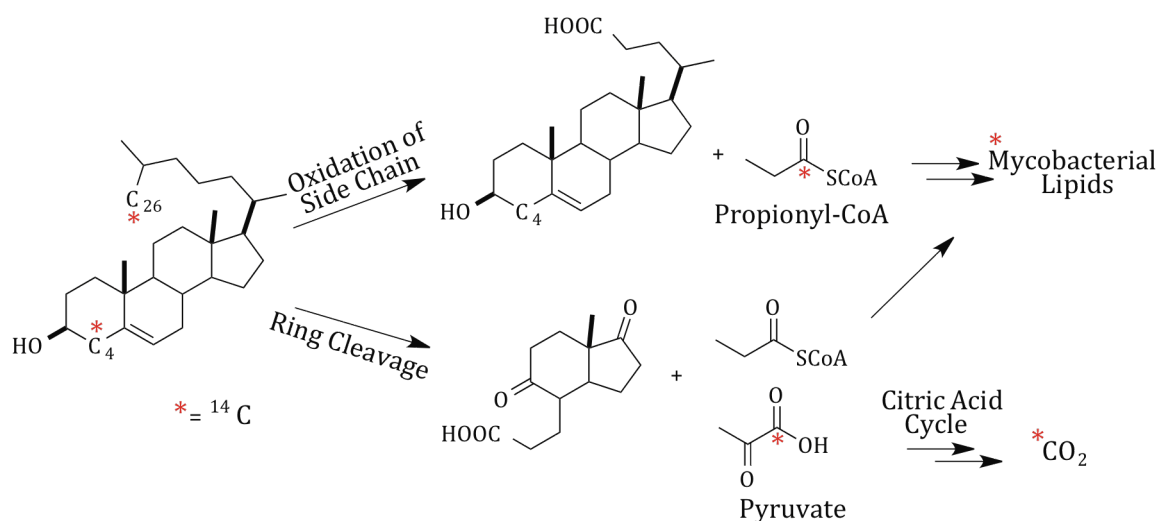


Figure 1-1. Structure of cholesterol. Rings of the steroid nucleus are labeled A-D and carbon atoms are numbered 1-27.



Scheme 1-2. Metabolic tracing to study cholesterol metabolism in *M. tuberculosis*. Metabolic tracing studies with [4- ^{14}C] and [26- ^{14}C] cholesterol revealed that carbon from cholesterol was incorporated into mycobacterial lipids and CO_2 . * represents ^{14}C label.

IV. The cholesterol metabolism pathway

Actinomycetes are Gram-positive bacteria that most commonly inhabit the soil. They play an important role in the carbon cycle because of their ability to decompose a variety of organic compounds. It is well established that actinobacteria can metabolize sterols, including cholesterol. Species of *Arthrobacter*, *Mycobacteria*, *Nocardia*, *Rhodococcus*, and *Streptomyces* can all utilize cholesterol as a sole carbon source (50). Sterol metabolism by actinomycetes has been extensively studied in order to identify pathway intermediates that may be of pharmaceutical use (51, 52). For example, 17-keto steroids, like androstenedione (AD), are key starting materials for the preparation of clinically useful steroids, such as testosterone, estradiol, progesterone, cortisone, and cortisol.

Side chain metabolism

Cholesterol metabolism includes opening the steroid nucleus and cleaving the hydrocarbon side chain. Early studies focused on side chain metabolism to obtain ring intact species of commercial value from inexpensive starting materials like cholesterol. In the 1960s Sih et al. characterized partially metabolized side chain intermediates from *Nocardia* cultured with cholesterol (53-55). These studies established that complete

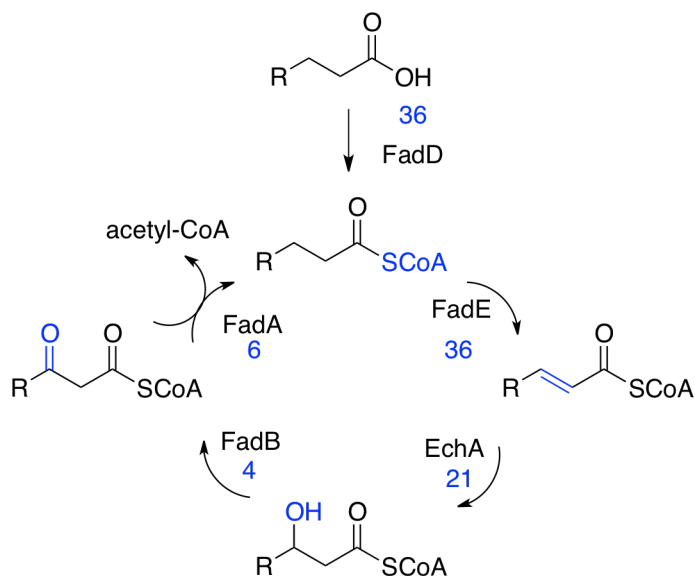
metabolism of the cholesterol side chain proceeds through C24 and C22 intermediates (Scheme 1-4). Propionate is lost during the formation of the C24 intermediate and can enter into fatty acid biosynthesis pathways or the methylcitrate cycle (Scheme 1-1). Acetate is lost during the formation of the C22 intermediate and can enter into the glyoxylate shunt. These metabolites suggest that the side chain is metabolized via conventional fatty acid β -oxidation. A final loss of propionate resulted in full side chain metabolism to the C17 keto steroid. A third round of β -oxidation is not possible because enoyl-CoA hydratase (ECH) would form a tertiary alcohol and is not capable of forming the subsequent β -keto-acyl-CoA. An atypical oxidation sequence is likely.

Little is known of the genes involved in the oxidation of the side chain. Early studies lacked sequence information for sterol-metabolizing microorganisms. Even with sequence data now available, gene assignment is complicated due to the large number of annotated fatty acid oxidation genes (56). It is challenging to identify which are involved in the metabolism of cholesterol and impossible to assign enzymes to specific steps in side chain metabolism from sequence data alone.

In order for the side chain to undergo β -oxidation it must first be activated and converted to a CoA thioester. It has long been proposed that a cytochrome P450 is responsible for oxidation of C26(27) methyl to a carboxylic acid. Actinobacteria have a large number of cytochrome P450s, making it challenging to assign gene function. Only recently has the enzyme responsible for catalysis been identified in any actinomycete. *Rhodococcus jostii* (*R. jostii*) *cyp125* is required for growth on cholesterol but not for growth on side chain-activated 5-cholestene-26-oic acid-3 β -ol (57). Recombinant activity has not been demonstrated. In *M. tuberculosis*, *cyp125*, the homologue of *cyp125* in *R. jostii*, has been assigned this function. Recombinant Cyp125 catalyzes the oxidation of the side chain of cholesterol and cholest-4-ene-3-one to the C26 carboxylic acid. An unknown CoA ligase then forms the thioester, the degradation of which proceeds via β -oxidation.

Bacterial β -oxidation includes four enzymatic steps including an acyl-CoA dehydrogenase (FadE), an enoyl-CoA-hydratase (EchA), a 3-hydroxy-acyl-CoA-dehydrogenase (FadB), and a 3-keto-acyl-CoA thiolase (FadA) (Scheme 1-3). Similar to other actinomycetes, it has been impossible to assign specific genes to β -oxidation of the

cholesterol side chain from sequence data alone. The *M. tuberculosis* genome contains numerous fatty acid degradation genes including a broad family of β -oxidation genes consisting of 36 *fadE*, 21 *echA*, 4 *fadB*, and 6 *fadA* genes (34). In addition, there are 36 acyl-CoA ligases (*fadD*) that generate the acyl-CoA thioester that enters into the β -oxidation cycle.



Scheme 1-3. β -oxidation enzymes are encoded by multiple copies in the *M. tuberculosis* genome. Shown are the four enzymatic steps of β -oxidation and the number of annotated genes assigned each function in the *M. tuberculosis* genome. The acyl-CoA ligase step is also shown.

Transcriptional profiling (58) and global phenotypic profiling experiments (59) of *M. tuberculosis* have aided in reducing the total number of lipid metabolizing genes to those likely involved in cholesterol metabolism. *M. tuberculosis* cultured with or without cholesterol revealed 52 genes within an 83-gene region that were upregulated by cholesterol. Phenotypic profiling identified 96 genes important for growth on cholesterol. Table 1-2 and Table 1-3 show a subset of the annotated fatty acid metabolizing genes that have been implicated in cholesterol metabolism based on these results. They include the genes predicted to degrade the C17 side chain and catabolize the sterol rings.

Surprisingly, 13 of 36 annotated *fadEs* are upregulated by cholesterol and 9 are required for growth on cholesterol. The predicted cholesterol metabolism pathway

requires only four acyl-CoA dehydrogenation steps, three for the oxidation of the side chain and one for the proposed metabolism of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid (DOHNAA) (*vide infra*) (Figure 1-2). *fadE* genes have not been assigned experimentally to specific steps in the pathway.

Three enoyl-CoA hydratases have been proposed to be involved in cholesterol metabolism, *echA9*, *echA19*, and *echA20*. *echA9*, *echA19*, and *echA20* are in the cholesterol regulon, though *echA19* and *echA20* are not required for growth of the bacteria on cholesterol. No β -hydroxy-acyl-CoA dehydrogenase encoding genes are located in the cholesterol regulon and it is not clear which genes are required for this transformation in the pathway.

Two acyl-CoA thiolases are found in the cholesterol regulon, *fadA5* and *fadA6*, but only *fadA5* is required for growth of *M. tuberculosis* on cholesterol, suggesting it is involved in cholesterol metabolism. Nesbitt *et al.* showed that *in vitro* this enzyme can catalyze thiolysis of acetoacetyl-CoA (58). *M. tuberculosis* cultured with cholesterol accumulates AD and androst-1,4-diene-3,17-dione (ADD), while androsterones are not detected in the *fadA5* knockout strain. These experiments demonstrate that FadA5 is one of the thiolase enzymes required for full metabolism of the side chain of cholesterol. In addition, the *fadA5* knockout strain shows attenuation in the mouse model of infection. Colony forming units (cfu) in the lungs showed a marked decrease after 8 weeks of infection compared to wild type and complement strains.

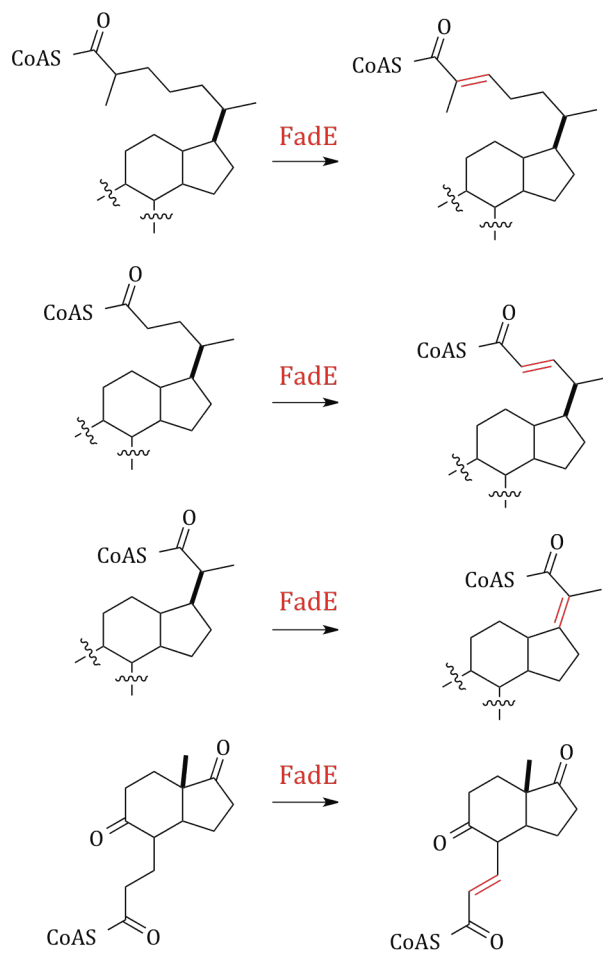
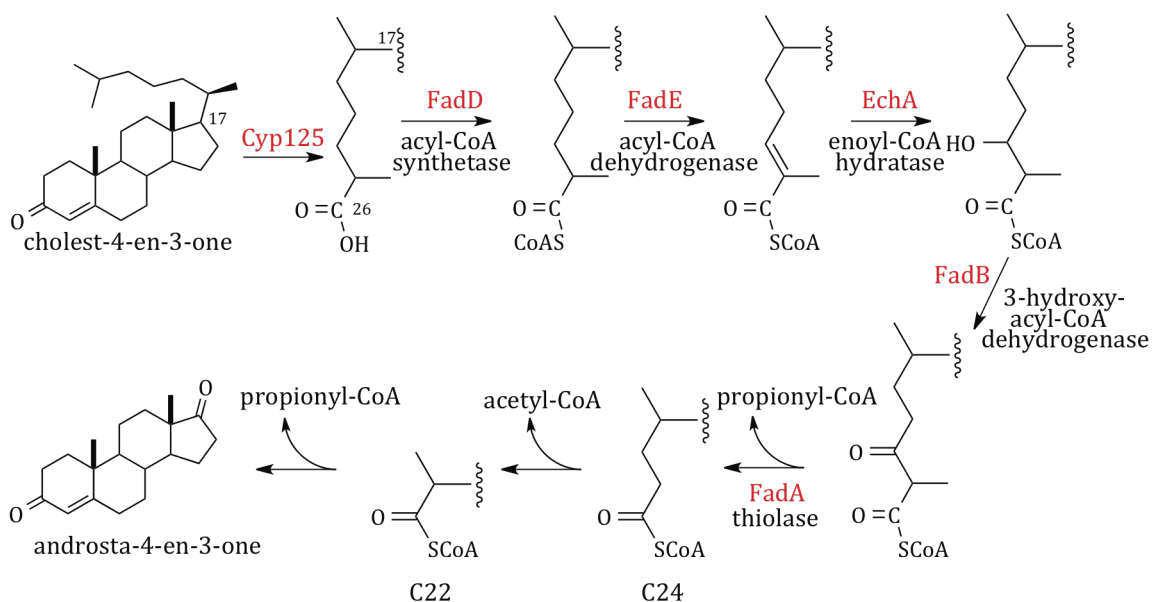


Figure 1-2. FadEs in cholesterol metabolism. Four steps in the proposed cholesterol metabolism pathway require dehydrogenation by a FadE.



Scheme 1-4. Cholesterol side-chain metabolism. The cholesterol side chain is proposed to be metabolized by β -oxidation. The enzymes that catalyze the activation of the side chain and the 4 steps of β -oxidation are shown in red. The cholesterol side chain loses 3 carbons as propionyl-CoA during the first round of β -oxidation. A second round of β -oxidation further shortens the side chain by two carbons, lost as acetyl-CoA. The mechanism for the formation of the C17 keto steroid is not known.

Table 1-2. Cholesterol side-chain metabolism. The genes proposed to be involved in the metabolism of the side chain of cholesterol by *M. tuberculosis* are presented. The experiments conducted for each gene are also included.

Gene name	<i>M. tb</i> gene number	Putative function	Studies performed	Reference
<i>cyp125</i>	<i>Rv3545c</i>	cytochrome P450	B, D, F	(57, 60-64)
<i>cyp142</i>	<i>Rv3518c</i>	cytochrome P450	B	(65, 66)
<i>fadD3</i>	<i>Rv3561</i>	acyl-CoA ligase	D, E	
<i>fadD19</i>	<i>Rv3515c</i>	acyl-CoA ligase	D, E	
<i>fadD36</i>	<i>Rv1193</i>	acyl-CoA ligase	D	
<i>fadE5</i>	<i>Rv0244c</i>	ACAD	D, E	
<i>fadE14</i>	<i>Rv1346</i>	ACAD	E	
<i>fadE17</i>	<i>Rv1934c</i>	ACAD	E	
<i>fadE18</i>	<i>Rv1933c</i>	ACAD	E	
<i>fadE25</i>	<i>Rv3274c</i>	ACAD	D	
<i>fadE26</i>	<i>Rv3504</i>	ACAD	E	
<i>fadE27</i>	<i>Rv3505</i>	ACAD	E	
<i>fadE28</i>	<i>Rv3544c</i>	ACAD	B, C, D, E, F	(63, 64) This work
<i>fadE29</i>	<i>Rv3543c</i>	ACAD	B, C, D, E, F	(63, 64) This work
<i>fadE30</i>	<i>Rv3560c</i>	ACAD	D, E	
<i>fadE31</i>	<i>Rv3562</i>	ACAD	D, E	
<i>fadE32</i>	<i>Rv3563</i>	ACAD	D, E	
<i>fadE33</i>	<i>Rv3564</i>	ACAD	D, E	
<i>fadE34</i>	<i>Rv3573c</i>	ACAD	D, E	
<i>echA9</i>	<i>Rv1071c</i>	(S)- hydratase	D	
<i>echA19</i>	<i>Rv3516</i>	(S)- hydratase	E, G	
<i>echA20</i>	<i>Rv3550</i>	(S)- hydratase	E	
	<i>Rv3541c</i>	(R)- hydratase	B, C, D, E, F	(63, 64) This work
	<i>Rv3542c</i>	(R)- hydratase	B, C, D, E, F	(63, 64) This work
<i>fadA5</i>	<i>Rv3546</i>	thiolase	B, C, D, E, F	(58)
<i>fadA6</i>	<i>Rv3556c</i>	thiolase	E	

- A. Bioinformatic annotation
- B. Recombinant expression and enzymatic function confirmed
- C. Function examined with mutant strain *in vivo*
- D. Required for growth on cholesterol (59)
- E. Upregulated by cholesterol (58)
- F. Growth phenotype *in vivo*
- G. No growth phenotype *in vivo*

A and B ring metabolism

Much of what is known about sterol ring metabolism was discovered in part based on similarities to known transforming enzymes from *Comamonas testosteroni* (*C. testosteroni*), *Rhodococcus sp.*, and other actinomycetes. *M. tuberculosis* enzymes involved in A and B ring metabolism have been characterized to varying extents. The genes required for metabolism of the C and D rings are not known.

Ring metabolism initiates at the A and B rings and it is established to occur via aromatization of the A ring followed by meta-cleavage of the B ring. The first step is the oxidation and isomerization of cholesterol to cholest-4-ene-3-one. In actinobacteria this reaction is catalyzed by either a 3 β -hydroxysteroid dehydrogenase (3 β -HSD) or cholesterol oxidase (ChoX). *Nocardia sp.* (67), *C. testosteroni* (68), and *R. jostii* (69) utilize a 3 β -HSD while *Streptomyces spp.* (70), *Rhodococcus equi* (*R. equi*) (71), and *Gordonia cholesterolivorans* (72) utilize a cholesterol oxidase to form cholest-4-ene-3-one.

In *M. tuberculosis* this function is assigned to *Rv1106c*, encoding a 3 β -HSD. The recombinant enzyme has been enzymatically characterized as a 3 β -HSD and is able to convert cholesterol to cholest-4-ene-3-one (73). Knockout experiments indicate 3 β -HSD is required for growth of *M. tuberculosis* on cholesterol (73). In addition, *Rv1106c* knockout culture supernatant no longer catalyzes the oxidation of cholesterol to cholest-4-ene-3-one (73, 74).

Ketosteroid Δ 1-dehydrogenase

The A ring then undergoes 1,2-desaturation catalyzed by a 3-ketosteroid Δ 1-dehydrogenase (KstD). KstD flavoproteins involved in sterol metabolism have been identified in *Nocardia* (75), *Rhodococcus* (76-78), and *M. smegmatis* (79). Interestingly some bacteria have been shown to encode more than one KstD enzyme and deletion of both were required to block 1,2-desaturation (80).

In *M. tuberculosis*, KstD is encoded by *Rv3537* (78). Activity has been demonstrated *in vitro* with substrate 5 α -androstane-3,17-dione and 17 β -hydroxy-5 α -androstane-3-one. Genome analysis indicates *Rv3537* is the sole KstD encoding gene in *M. tuberculosis*. A *Rv3537* disrupted strain of *M. tuberculosis* was unable to grow on cholesterol as a carbon source and accumulated 9-hydroxy-4-androstene-3,17-dione (9-OHAD) (81).

3-Ketosteroid 9 α -hydroxylase

Next a 3-ketosteroid-9 α -hydroxylase catalyzes the hydroxylation at C9, which leads to aromatization of the A ring and opening of ring B. KshA, the oxygenase, and KshB, the reductase, make up the two-component Rieske oxygenase. While Ksh activity has been observed in several actinomycetes (82-84), it was not until recently that this activity was demonstrated with recombinant enzyme from *Rhodococcus* for substrates AD and ADD (85).

The 3-ketosteroid-9 α -hydroxylase homologs in *M. tuberculosis*, KshA/Rv3526 and KshB/Rv3571, have been recombinantly expressed in *E. coli* and purified (86). Activity was reconstituted *in vitro* with several substrates including AD, ADD, the CoA thioester of 3-oxo-4-pregnene-20-carboxylic acid, and the CoA thioester of 3-oxo-1,4-pregnadiene-20-carboxylic acid (87).

3-HSA-hydroxylase

C. testosteronei is able to utilize testosterone as a sole carbon source and studies elucidating this pathway have provided insight into cholesterol ring system metabolism. For instance, gene disruption of *tesA1* and *tesA2* of *C. testosteronei* grown with testosterone were shown to accumulate 3-hydroxy-9,10-secoandrost-1,3,5(10)-triene-9,17-dione (3-HSA) and *tesA1* and *tesA2* are required for the formation of 3,4-dihydroxy-9,10-secoandrost-1,3,5(10)-triene-9,17-dione (3,4-DSHA) (88). TesA1 and TesA2 are assigned as the flavin-dependent oxygenase, reductase pair responsible for converting 3-HSA to 3,4-DSHA.

In *M. tuberculosis* this reaction is assigned to *hsaA/Rv3570c* and *hsaB/Rv3567c*. Conversion of 3-HSA to 3,4-DSHA has been demonstrated *in vitro* with recombinant HsaAB (89). Mutagenesis analysis showed that *Rv3570c* is required for growth of *M. tuberculosis* in macrophages (90).

2,3-Dehydroxyphenyl dioxygenase

Studies in *C. testosteronei* identified *tesB* as the iron-dependent extradiol dioxygenase, responsible for the meta-cleavage of the A ring of 3,4-DHSA to give 4,5-9,10-diseco- α 3-hydroxy-5,9,17-trioxoandrost-1(10),2-diene-4-oic acid (4,9-DSHA) (91, 92). In *Rhodococcus sp.* HsaC (*Rv3568c*) is homologous to TesB and a *hsaC-disrupted* strain grown on cholesterol accumulates 3,4-DHSA (56).

In *M. tuberculosis* *hsaC/Rv3568c* encodes the dioxygenase responsible for the formation of 4,9-DSHA. Activity has been demonstrated *in vitro* with recombinant enzyme and substrate 3,4-DHSA (56, 93). *HsaC* is required for growth of *M. tuberculosis* on cholesterol but not on glycerol. Knockout studies reveal that *hsaC* is important for pathogenesis. Immuno-compromised mice and guinea pigs infected with Δ *hsaC* lived significantly longer than those infected with wild type (93). In addition harvested lungs showed lower bacterial loads after 8 weeks of infection. These results establish that *hsaC* is important for pathogenesis of *M. tuberculosis*.

4,9-DSHA hydrolase

In *C. testosteroni*, TesD was been confirmed as the hydrolase that cleaves 4,9-DSHA to form DOHNAA and 2-hydroxy-hexa-2,4-dienoic acid (HHD). In *M. tuberculosis* HsaD (*Rv3569c*) catalyzes the same reaction (56, 94). The hydrolase activity has been demonstrated with recombinant enzyme and substrate 4,9-DSHA. In addition, HsaD is important for *M. tuberculosis* survival within macrophages (90).

Metabolism of HDD and DOHNAA

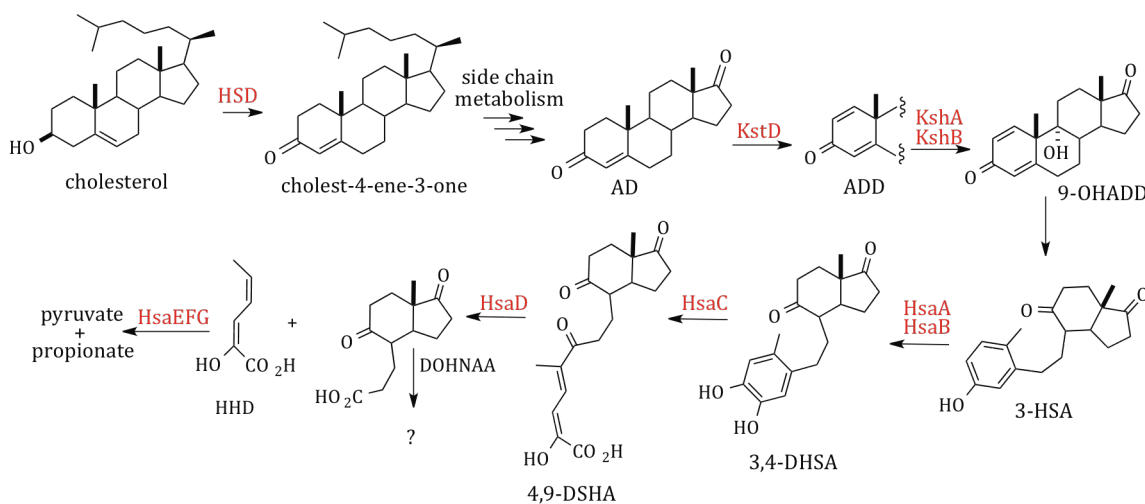
The genes for the metabolism of HDD by *M. tuberculosis* have been proposed based on homology to enzymes in *C. testosteroni*. HDD is metabolized to 4-hydroxy-2-oxohexanoic acid in *C. testosteroni* by 2-hydroxypentadienoate hydratase, TesE (95). Next *tesG*, encoding a 4-hydroxy-2-oxovalerate aldolase, forms pyruvate and propionaldehyde. Pyruvate can be converted to carbohydrates or fatty acids. TesF, an acetaldehyde dehydrogenase, converts propionaldehyde to propionyl-CoA.

In *M. tuberculosis* HsaEFG (*Rv3534c/Rv3535c/Rv3536c*) are hypothesized to metabolize HDD (56). HsaE, HsaF, and HsaG share 41%, 47%, and 56% amino acid identity with TesE, TesG, and TesF from *C. testosteroni*, respectively. The activities of HsaEFG have not been verified and gene knockout studies reveal that *hsaEFG* are not required for growth on cholesterol (59).

The metabolic fate of DOHNAA in *M. tuberculosis* is unknown. Several C and D ring intact hexahydroindanone intermediates have been characterized in actinomycetes, including *Nocardia*, *Streptomyces*, and *Rhodococcus* (96-100). Hayakawa et al. demonstrated that some actinomycetes were not able to metabolize DOHNAA like hexahydroindanones further and nitrogenous and amino acid derivatives have been

observed (96, 97, 101, 102). It has been suggested that metabolism of the C and D rings initiates by lactonization of the cyclic carbonyl by a Baeyer-Villiger monooxygenase (BVMO) (56). C and D ring lactones have been identified in bacteria and fungi cultured with sterols, supporting this hypothesis (103-108). The BVMO in *M. tuberculosis* has not been identified and the cholesterol regulon lacks an obvious gene candidate (56).

Studies in *Rhodococcus equi* suggest the propionate moiety of DOHNAA is metabolized by β -oxidation (100, 109). Nakamatsu et al. proposed that propionate was metabolized to an acetyl group via α -oxidation (99). The genes required for α or β -oxidation of the propionate group of DOHNAA have not been identified, though it is proposed that *fadE30* in *R. equi* is responsible for the first dehydrogenation step (109). *R. equi* FadE30 shares 68% amino acid identity with FadE30 (*Rv3560c*) from *M. tuberculosis*. Further study is necessary to elucidate the later steps in the cholesterol metabolism pathway in *M. tuberculosis*.



Scheme 1-5. Cholesterol ring system metabolism. The metabolic steps of ring degradation by *M. tuberculosis* have been partially elucidated. Enzymes responsible for the presented transformations are shown in red. See Table 1-3 for studies performed for each enzyme.

Table 1-3. Cholesterol ring system metabolism. The genes involved in the metabolism of the cholesterol ring system by *M. tuberculosis* are presented. The experiments conducted for each gene are also included.

Gene name	<i>M. tb</i> gene number	Putative function	Studies performed	Reference
<i>hsd</i>	<i>Rv1106c</i>	3 β -hydroxysteroid dehydrogenase	B, C, D, E, G, H	(73, 74)
<i>kstD</i>	<i>Rv3537</i>	ketosteroid Δ 1-dehydrogenase	B, C, D, E, H	(78, 81)
<i>kshA</i>	<i>Rv3526</i>	3-Ketosteroid 9 α -hydroxylase	B, D, E, F, H	(86, 87, 110)
<i>kshB</i>	<i>Rv3571</i>	3-Ketosteroid 9 α -hydroxylase	B, D, E, F, H	(86, 87, 110)
<i>hsaA</i>	<i>Rv3570c</i>	3-HSA-hydroxylase	B, D, E, H	(89)
<i>hsaB</i>	<i>Rv3567c</i>	3-HSA-hydroxylase	B, D, E	(89)
<i>hsaC</i>	<i>Rv3568c</i>	2,3-dehydroxyphenyl dioxygenase	B, E, F, H	(93)
<i>hsaD</i>	<i>Rv3569c</i>	4,9-DSHA hydrolase	B, E, H	(111)
<i>hsaE</i>	<i>Rv3536c</i>	hydratase	A, E	(56)
<i>hsaF</i>	<i>Rv3534c</i>	aldolase	A, E	(56)
<i>hsaG</i>	<i>Rv3535c</i>	acetaldehyde dehydrogenase	A, E	(56)

- A. Bioinformatic annotation
- B. Recombinant expression and enzymatic function confirmed
- C. Function examined with mutant strain *in vivo*
- D. Required for growth on cholesterol
- E. Upregulated by cholesterol (58)
- F. Growth phenotype *in vivo*
- G. No growth phenotype *in vivo*
- H. Growth phenotype in macrophages (90)

V. *M. tuberculosis* 3 β -hydroxysteroid dehydrogenase

The first step in cholesterol metabolism is predicted to be the conversion of cholesterol to cholest-4-ene-3-one. The genome of *M. tuberculosis* contains an annotated 3 β -*hsd/Rv1106c*. The gene product of *M. tuberculosis Rv1106c* has been enzymatically characterized and catalyzes the oxidation and isomerization of 3 β -hydroxy-5-ene-steroids, cholesterol, dehydroepiandrosterone (DHEA), and pregnenolone to 3-oxo-4-ene steroids, cholest-4-ene-3-one, AD, and progesterone, respectively (73).

The short-chain dehydrogenase/reductase superfamily

3 β -HSD is a member of the short chain dehydrogenase/reductase (SDR) super family and is required for the biosynthesis of steroid hormones, including glucocorticoids, mineralocorticoids, androgens, and estrogens (112). Members of this family require nicotinamide adenine dinucleotide for catalysis. They consist of two structural/functional domains. The N terminal domain is the nucleotide, cofactor-binding domain and has a

characteristic Rossmann fold (113). The C terminal domain varies within the family and is responsible for substrate binding.

Mutagenesis and crystallographic data show that members of the SDR family, including 3 β -HSD, require an Asn-Ser-Tyr-Lys tetrad for activity (114, 115). In *M. tuberculosis* 3 β -HSD the tetrad is conserved as residues Asn107, Ser131, Tyr158, and Lys162. Tyr is key in the dehydrogenase reaction and functions as the catalytic base (114). Ser is important for substrate binding and Lys interacts with the catalytic Tyr, lowering the pKa of the hydroxyl group (116, 117). Crystallographic studies indicate that Asn plays a structural role, linking the substrate loop and active site by interactions with a bound water molecule in the active site (118).

Cholesterol oxidation and isomerization to cholest-4-ene-3-one in M. tuberculosis

The genome of *M. tuberculosis* also contains an annotated cholesterol oxidase (*ChoD/Rv3409c*). Cholesterol oxidase is a member of the glucose-methanol-choline (GMC) oxidoreductase family and catalyzes the same reaction as 3 β -HSD, though by a different mechanism. ChoD has been suggested to have cholesterol oxidation activity due to the temporary accumulation of cholest-4-ene-3-one in *M. smegmatis* over expressing *choD* (119). Studies by Yang et al., demonstrate that *Rv3409c* is not required for growth of *M. tuberculosis* on cholesterol and the recombinant enzyme cannot convert cholesterol to cholest-4-ene-3-one (74).

Knockout experiments indicate 3 β -HSD is required for growth of *M. tuberculosis* on cholesterol while growth of *Rv3409c* knockout strain was not affected (74). In addition, culture supernatant lost the ability to oxidize cholesterol to cholest-4-ene-3-one in *Rv1106c* knockout strains indicating 3 β -HSD is the sole cholesterol-oxidizing enzyme in *M. tuberculosis* (73). *Rv1106c* is not upregulated by cholesterol and is not part of the cholesterol regulon (58).

M. smegmatis culture supernatant from a *ChoD* (*MSMEG1604*) mutant strain was still able to convert cholesterol to cholest-4-ene-3-one (74, 120). Likewise in *Mycobacterium sp.* the *choD* knockout strain was still able to form 3-keto-4-ene steroids (121). These results suggest that ChoD is not a cholesterol oxidase and is not responsible for 3 β -hydroxy-5-ene

steroid oxidation in *M. tuberculosis* or in other mycobacterial species. A role for ChoD in cholesterol metabolism has not been substantiated and its function is still unknown.

Rv1106c is important for cholesterol metabolism, but it is not important for *M. tuberculosis* growth in the macrophage (74). Growth rates of *3 β -hsd* knockout strain replicated at a similar rate to wild type in macrophages. In addition, infection studies in the guinea pig showed identical cfu in the lungs of wild type, *3 β -hsd* knockout, and *3 β -hsd* complement infection models. It is likely multiple carbon sources are available to the bacteria during infection and that *Rv1106c* is not necessary for nutrition acquisition (74). In most bacteria, carbon sources are catabolized in order of their ability to support growth. Recently *M. tuberculosis* has been shown to catabolize multiple carbon sources simultaneously (122). The required role of cholesterol metabolism in pathogenesis of *M. tuberculosis* is unlikely to be the supply of energy to the bacteria.

VI. The *intracellular growth operon*

The *intracellular growth (igr)* operon (*Rv3545c-Rv3540c*) is implicated in lipid metabolism by *M. tuberculosis*. The computationally annotated functions of the *igr* operon consist of a cytochrome P450 (*cyp125/Rv3545c*), two acyl-CoA dehydrogenases (*fadE28/Rv3544c* and *fadE29/Rv3543c*), two (*R*)-specific enoyl-CoA hydratases (*Rv3541c* and *Rv2542c*) and a lipid transfer protein *ltp2* (*Rv3540c*) (Figure 1-3). The *igr* operon is conserved in pathogenic and nonpathogenic mycobacteria, including *M. smegmatis*, *M. leprae*, *M. avium*, and *M. bovis*. In other actinomycetes the operon is only partially conserved. In *Rhodococcus*, *Streptomyces*, and *Gordonia* *cyp125* is not found within the operon.

The annotated functions of the *igr* operon suggest it may be involved in lipid metabolism. However, the *igr* operon knockout (Δ *igr*) strain grew just as well as wild type *M. tuberculosis* on pyruvate, valerate, isovalerate, propionate, palmitate, dodecanoate, glycerol, Tween, and dextrose suggesting it is not involved in fatty acid metabolism (64). Further studies have revealed that the operon is important for cholesterol metabolism. The Δ *igr* strain does not grow on cholesterol as a sole carbon source *in vitro* (64).

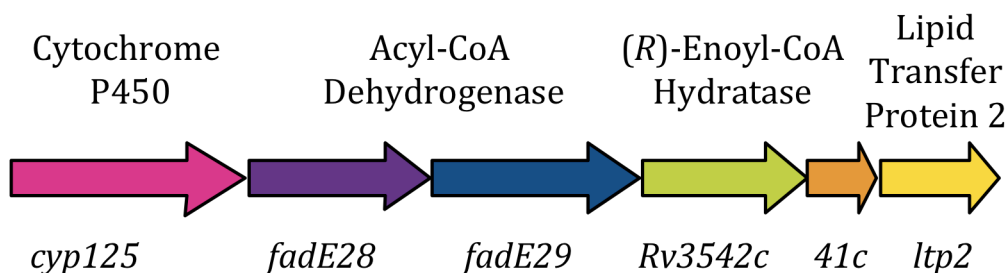


Figure 1-3. The *igr* operon. Organization of the *igr* operon (*Rv3545c-Rv3540c*), including the annotated functions of each gene.

The *igr* operon is upregulated by cholesterol and is located within the cholesterol regulon in the *M. tuberculosis* genome (58). The Δ *igr* mutant is still able to import and partially degrade cholesterol *in vitro*. Metabolic tracing studies with isotopically labeled [4-¹⁴C] and [26-¹⁴C] cholesterol indicated that C4 was incorporated into CO₂ and C26 was found in mycobacterial lipids. This is the same result as wild-type metabolic tracing studies, therefore the function of the operon could not be ascertained (Scheme 1-2) (37).

Growth of a Δ *igr* knockout strain of *M. tuberculosis* is attenuated in resting macrophages and early in the infection process in immunocompetent mice. Cfu levels in the lungs and spleen were greatly reduced compared to wild type. The growth attenuation observed in Δ *igr* demonstrates the operon is important for intracellular growth, giving the *igr* its name. However, this phenotype is minimized in INF- γ activated macrophages, which correlates with the onset of the adaptive immune response in mouse infections (123). Interestingly, the Δ *igr* knockout strain displays a cholesterol-sensitive phenotype *in vitro* when the bacteria enter a static growth phase in glycerol and glucose-containing media supplemented with low concentrations of cholesterol (0.1 mM) (64). This cholesterol-sensitive phenotype is characterized by the bacterium maintaining an intact respiratory electron transport chain in the absence of cell division (124). Moreover, the *in vitro* and *in vivo* cholesterol-sensitive phenotype of the Δ *igr* knockout can be suppressed by genetically disrupting *yrbE4a/Rv3501c*, which is the first gene in the operon that encodes the multi-subunit cholesterol importer, Mce4 of *M. tuberculosis* (124, 125). These data suggest that disruption of the *igr* operon results in accumulation of a toxic cholesterol-derived

metabolite and that this toxicity can be relieved by blocking cholesterol uptake. The biochemical function and molecular cause for the growth phenotype are not known.

VII. Cytochrome P450 Cyp125

The *M. tuberculosis* genome encodes for 20 cytochrome P450 enzymes and studying their function and structure is an active area of research (126-130). Cytochrome P450s are heme proteins that catalyze monooxygenation reactions in which one atom of oxygen is incorporated into the substrate from molecular oxygen. They are capable of catalyzing the oxidation of a variety of non-activated hydrocarbon substrates with high regio- and stereo-specificity. They are found in all domains of life and are involved in anabolic and catabolic pathways, including the biosynthesis and metabolism of steroids.

Cyp125/Rv3545c is annotated as a cytochrome P450 and is the only gene that has been functionally characterized from the *igr* operon prior to this work. The gene was recombinantly expressed, in separate accounts, in *R. jostii* *RHA1* or in *E. coli* with coexpression of folding chaperones, GroEL and GroES (60, 62, 64). Purified protein was shown to bind cholesterol and cholest-4-ene-3-one with sub micromolar affinity (60). Cyp125, with reductase KshB (*Rv3571*) or spinach ferredoxin reductase, and NADH, was able to transform cholesterol and cholest-4-ene-3-one *in vitro* to the C26 hydroxy and the C26 carboxylic acid products. *In vivo* formation of the carboxylic acid would be necessary for activation of the side chain and further conversion to the CoA thioester, in order to undergo β -oxidation.

The crystal structure of Cyp125 has also been reported (61). Structures were obtained without ligand as well as with AD and the antitubercular drug econazole. These ligands were found to bind within the active site cavity of the enzyme. Docking was used to incorporate cholesterol into the structure and it was found for the lowest energy structure that the C26 and C27 of cholesterol were 5.3 and 6.3 Å from the heme iron center, respectively. The position of these carbons relative to the active site of the enzyme substantiates that this is where the transformation occurs.

Knockout *cyp125* (Δ *cyp125*) *M. tuberculosis* CDC1551 cultures grown with cholesterol accumulate cholest-4-ene-3-one, suggesting that this is the physiological substrate for the enzyme *in vivo* (62). In culture, cholest-4-ene-3-one is toxic to H37Rv and

CDC1551 *M. tuberculosis* strains (62). Interestingly *cyp125* is required for growth on cholesterol in CDC1551 but is not required in the H37Rv strain. *In vitro* recombinant Cyp124 and Cyp142 oxidized the side chain of cholesterol, but only Cyp142 could support the growth of *M. tuberculosis* on cholesterol in $\Delta cyp125$, whereas Cyp124 could not (65, 66). It was demonstrated that *cyp124* is not expressed in *M. tuberculosis* H37Rv and CDC1551 wild-type strains (66). Therefore, Cyp142 and not Cyp124 provide compensatory activity in H37Rv.

The metabolite profile of H37Rv $\Delta cyp125$ strain has not been elucidated. Due to the compensatory activity of Cyp142 in H37Rv the metabolic phenotype is likely to be different in this strain than CDC1551 where cholest-4-ene-3-one was found to accumulate. In addition, the *in vivo* phenotype of CDC1551 $\Delta cyp125$ has not been investigated and it is unclear if it is consistent with the growth attenuation of H37Rv Δigr .

VIII. Acyl-CoA Dehydrogenases

The *M. tuberculosis* genome encodes several homologs of each type of β -oxidation enzyme. In the case of acyl-CoA dehydrogenases (ACADs) there are 36 genes computationally annotated as encoding for this activity (34). Prokaryotic ACADs are known as Fatty acid degrading E (FadE) enzymes. *FadE28/Rv3544c* and *FadE29/Rv3543c* of the *igr* operon have been computationally annotated as ACADs.

ACADs are a family of flavoproteins which catalyze the α,β -unsaturation of acyl esters (131). They require flavin adenine dinucleotide (FAD) for catalysis. A catalytic glutamate abstracts the alpha proton from an acyl-CoA substrate and the β -hydride is transferred to N5 of the isoalloxazine ring of FAD. Reduced FAD is re-oxidized by the transfer of reducing equivalents to electron transfer protein. The reaction proceeds by a concerted mechanism (131, 132).

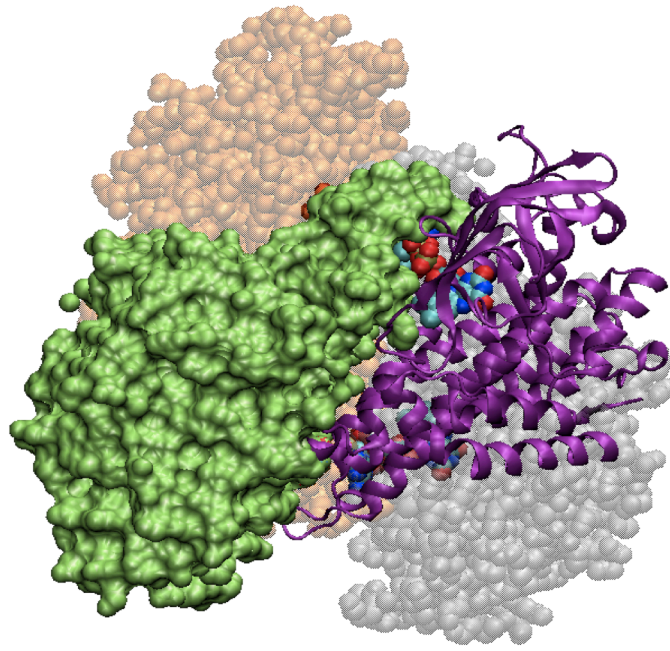
ACADs have been identified in plants (133), animals, bacteria, nematodes (134), and fungi (135). The family of enzymes has been extensively studied in eukaryotes (136). There are nine major eukaryotic classes of ACADs that vary in their substrate specificities. Additional classes are still being identified (137). Five classes are involved in fatty acid metabolism as part of β -oxidation, including short chain (SCAD), medium chain (MCAD), long chain (LCAD), very long chain (VLCAD), and very long chain 2 (VLCAD2). The

remaining members are involved in amino acid oxidation pathways and include isovaleryl-CoA dehydrogenase (i3VD) for leucine, short/branched chain acyl-CoA dehydrogenase (S/BCAD or i2VD) for isoleucine, glutaryl-CoA dehydrogenase (GD) for lysine and tryptophan, and isobutyryl-CoA dehydrogenase (iBD) for valine.

Though the substrate specificity varies between each class the sequences are highly homologous. They all have a similar fold, consisting of three structural domains: N- and C-terminal α -helical domains, and a middle β -sheet domain. All known classes of ACADs exist as homotetramers, except for VLCAD and VLCAD2, which are homodimers (Figure 1-4) (138). ACADs contain one non-covalently bound FAD that binds at the interface of two monomers and is essential for catalysis.

ACAD utilize a catalytic base, most often a glutamate, for abstraction of the α -proton of the acyl-CoA substrate. Crystallography and mutagenesis studies have elucidated that the catalytic base is located at one of two positions in helix G or helix J/K (139-142). The catalytic glutamate in helix J/K orients into the same active site as the catalytic glutamate in helix G in the three-dimensional protein structures of ACADs. For SCAD, MCAD, i2BD, iBD, VLCAD, VLCAD2, and GD the catalytic base is located in helix J/K while for i3VD and LCAD the catalytic glutamate is located in helix G. These two catalytic residues are separated by 100 amino acids and lie in different helices, but mechanistically catalyze the reaction identically (Figure 1-5).

A



B

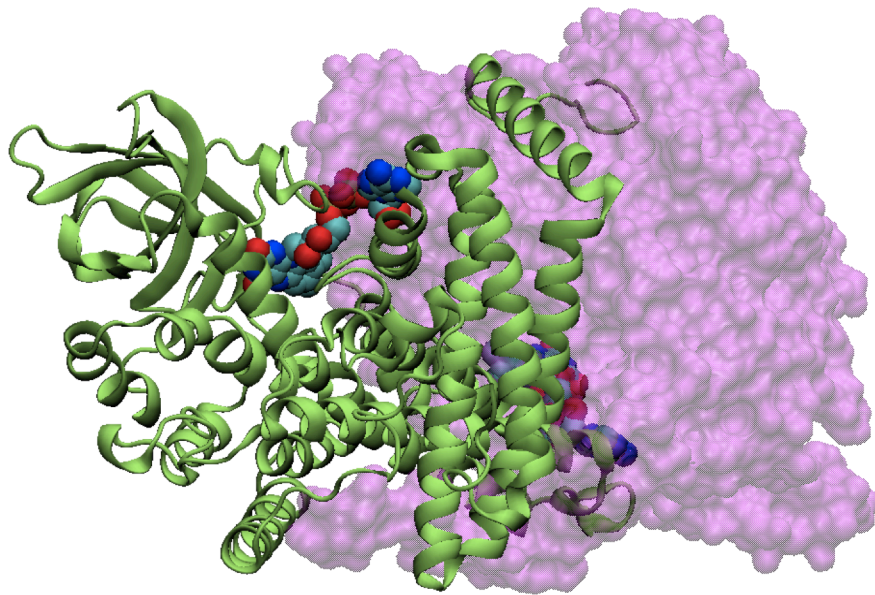


Figure 1-4. Acyl-CoA dehydrogenase crystal structures. Three-dimensional structure of (A) homotetrameric human MCAD and (B) homodimeric human VLCAD. FAD is shown as a Van der waals surface. PDB 1EGC and 3B96, respectively.

A

	Helix G		Helix J/K	
SCAD	MGRIGIASQ	277	EIYEGTSEIQRLV	401
MCAD	KTRPVVAAG	286	QIYEGTSQIQRLI	410
VLCAD	NGRFGMAAA	324	RIFEGTNDILRLF	448
VLCAD2	SGRFSMGSV	309	LIFEGTNEILRMY	435
i2VD	EGRIGIAAQ	299	TIYEGASNIQLNT	423
GD	NARYGIAWG	299	NTYEGTHDIHALI	423
iBD	GGRINIASC	282	QILEGSNEVMRIL	407
i3VD	LERLVLAGG	289	EIGAGTSEVRRLV	413
LCAD	QERLLIADV	297	PIYGGTNEIMKEL	421



Figure 1-5. The catalytic base of acyl-CoA dehydrogenase. (A) The confirmed catalytic glutamate in representative human acyl-CoA dehydrogenase from each of the 9 characterized classes is shown in red. (B) The catalytic glutamate in helix J/K in teal orients into the same active site as the catalytic glutamate in helix G in purple.

The proposed cholesterol metabolism pathway by *M. tuberculosis* includes four acyl-CoA dehydrogenation steps, three involved in side chain metabolism and one for the degradation of DOHNAA (Figure 1-2). Of the 36 annotated ACADs, 13 were found to be upregulated by cholesterol (58). This suggests the possibility of functionally-redundant enzymes. Or possibly ACAD activity is required for the metabolism of the C and/or D rings of cholesterol or additional sterol transformations yet to be proposed. Thus far, no *M. tuberculosis* ACADs have been characterized or been shown to be directly involved in sterol dehydrogenation. The biochemical functions of *fadE28* and *fadE29* are investigated in this work.

IX. 2-enoyl-CoA hydratases

The second step of β -oxidation is the hydration of enoyl-CoA esters to β -hydroxyacyl-CoA products by enoyl-CoA hydratase. In *M. tuberculosis* 21 genes are computationally annotated as enoyl-CoA hydratases (34). Sequence analysis of *Rv3541c* and *Rv3542c*, of the *igr* operon, suggest that these genes encode for (*R*)-specific-enoyl-CoA hydratases ((*R*)-hydratase).

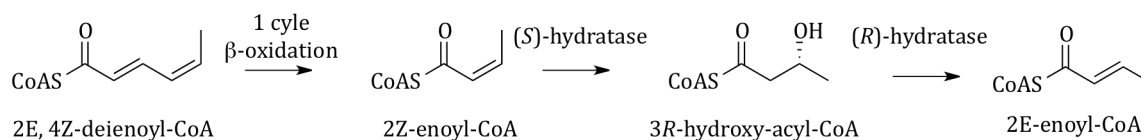
(*R*)-hydratases yield 3*R*-hydroxyacyl-CoA products, which differ from the classical enoyl-CoA hydratase, (*S*)-specific-enoyl-CoA hydratase ((*S*)-hydratase) which forms the 3*S*-hydroxyacyl-CoA stereoisomer. (*S*)-hydratases have been well studied, such as the classic example crotonase, and are involved in fatty acid metabolism by β -oxidation (143).

Several (*R*)-hydratases have been identified and studied in bacteria, including those in *Aeromonas caviae* (*A. caviae*) (144), *Pseudomonas aeruginosa* (145), *Rhodospirillum rubrum* (146), and *E. coli* (147). Evidence suggests they are involved in polyhydroxyalkanoate (PHA) biosynthesis (144, 146, 147). PHAs are polyesters used to store carbon and energy. 3*R*-hydroxyacyl-CoA monomers, produced by (*R*)-hydratases, are shunted towards PHA synthesis and polymerized by PHA synthase (144, 148).

In mammals, (*R*)-hydratases are found as part of multifunctional enzyme 2 (MFE-2). MFE-2 contains three functional domains with (3*R*)-hydroxyacyl-CoA dehydrogenase, (*R*)-hydratases, and sterol carrier protein activities. MFE-2 has been shown to be important for β -oxidation of very-long-chain and α -methyl-branched fatty acids, and bile acid synthesis (149). For example, MFE-2 in rat liver peroxisomes has been shown to be involved in cholic acid synthesis (150). The biosynthesis of cholic acid requires shortening of the side chain of C27 steroids by β -oxidation. Recombinant (*R*)-hydratase is able to catalyze the hydration of (24*E*)-3 α ,7 α ,12 α -trihydroxy-5 β -cholest-24-enoyl-CoA, an intermediate in bile acid synthesis. In rats, MFE-2 knockout mice accumulate very long chain fatty acids, branched chain fatty acids, and bile acid intermediates (151).

In yeast, (*R*)-hydratase is found as part of MFE-2 and has been studied in *Saccharomyces cerevisiae* (152). Unlike mammalian MFE-2, which includes three functional domains, in yeast, MFE-2 includes only two activities, (*R*)-hydratase and (3*R*)-hydroxyacyl-CoA dehydrogenase. They are implicated in peroxisomal β -oxidation.

(*R*)-hydratases have been identified in plants including *Arabidopsis thaliana* (153). Like in bacteria, (*R*)-hydratases in plants are monofunctional enzymes. They play an important role in the metabolism of fatty acids with cis double bonds on an even-numbered carbon (Scheme 1-6). Similar monofunctional (*R*)-hydratases have been predicted in plants based on sequence homologue, including in classes of Monocotyledons and Dicotyledons (153).



Scheme 1-6. (*R*)-hydratase is important for the metabolism of fatty acids with cis double bonds. Following one round of β -oxidation, 2E,4Z-dienoyl-CoA is converted to 2Z-enoyl-CoA. (*S*)-hydratase involved in fatty acid metabolism converts 2Z-enoyl-CoAs to the 3*R*-hydroxy-acyl-CoA isomer. (*R*)-hydratase then converts 3*R*-hydroxy-acyl-CoA to 2E-enoyl-CoA, which can under go β -oxidation.

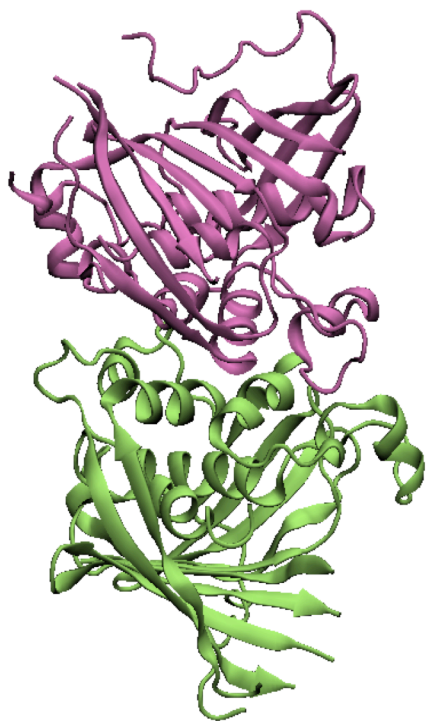
Multiple sequence alignment analysis of (*R*)-hydratase family members revealed a highly conserved, consensus motif, [YF]-X(1,2)-[LVIG]-[STGC]-G-D-X-N-P-[LIV]-H-X(5)-[AS], where X is any amino acid (150). This motif is conserved in the prokaryote and eukaryote (*R*)-hydratase subfamilies. The motif includes the catalytic dyad consisting of aspartic acid and histidine residues (154). The reaction mechanism has not been highly studied, but is it predicted to occur by an acid-base mechanism with an enol-like transition state. The catalytic residues are responsible for activating water and the reaction proceeds via a concerted mechanism (155).

The (*R*)-hydratase family has a characteristic hot dog fold consisting of a central helix wrapped in several β -sheets (149, 154). The hot dog fold is found in several functional families, including dehydratases, thioesterases, and acetyltransferases, all of which utilize CoA substrates. Interestingly, the oligomerization state and quaternary associations amongst hot dog fold proteins is not conserved. Dimeric, tetrameric, and hexameric oligomers have been crystallized. For example, thioesterases involved in the phenylacetic acid catabolic pathway form homotetramers while acyl-CoA thioesterases form homohexamers (156). Hot dog fold crystal structures with hexameric oligomers have interactions at the interface between the N-terminal helices, in a head-to-tail arrangement,

and with back to back stacking of β -sheets, highlighting how the associations of the monomeric subunits differ for a given oligomeric state (157-163).

In mammals and fungi, (*R*)-hydratase is part of the peroxisomal multifunctional protein (MFE) and forms a homodimer of double hot dogs (Figure 1-6A and Table 1-4). Each double hot dog monomer, with a mass of 32 kDa, includes one active site found in the C terminal domain. The N terminal domain lacks the central alpha helix and includes the cavity for binding of the hydrocarbon acyl part of the substrate molecule (155). It is hypothesized that the double hot dog fold arose from gene duplication and the loss of one active site allows the enzyme to accommodate bulky substrates, consistent with its preference for long chain substrates (149). In prokaryotes, (*R*)-hydratases form a homodimer of a single hot dog domain of 14 kDa (Figure 1-6B). Each monomer includes one active site (154).

A



B

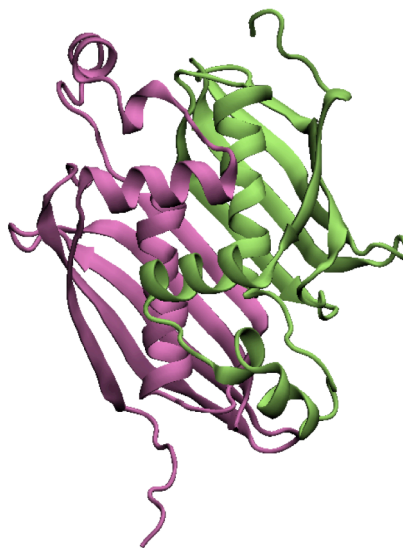


Figure 1-6. Enoyl-CoA hydratase crystal structures. Three-dimensional structure of (A) dimer of double hot dog fold human (*R*)-hydratase and (B) dimer of single hot dog fold *A. caviae* (*R*)-hydratase. PDB 1S9C and 1IQ6 respectively.

Table 1-4. Structural information for (*R*)-hydratases. Shown is a summary of structural information for (*R*)-hydratases identified in mammals, fungi, plants, and bacteria.

Kingdom	Activities	Oligomeric structure	No. active sites per oligomer	Ref
Mammals	Tri-functional	Dimer of double hot dogs	Two	(149)
Fungi	Di-functional	Dimer of double hot dogs	Two	(155)
Plants	Mono-functional	NA	-	
Bacteria	Mono-functional	Dimer of single hot dogs	Two	(154)

NA: No crystal structure available

Sequence analysis of Rv3541c indicates it may form a single hot dog (*R*)-hydratase, with a characteristic size of 14 kDa. Rv3542c is about twice the size of typical hot dog domains in prokaryotes at 34 kDa, indicating it may possibly fold as a double hot dog. The biochemical roles of Rv3541c and Rv3542c are unknown. Due to their importance in *M. tuberculosis* intracellular growth and involvement in cholesterol metabolism, the function of these genes is investigated in this work.

X. Specific Aims

The goal of this thesis is to gain mechanistic understanding of cholesterol metabolism by *M. tuberculosis*. Experiments conducted were designed to target the cholesterol metabolism pathway by small molecule inhibitors and elucidate the biochemical role of the *igr* operon. Fully elucidating the chemical steps in the cholesterol metabolic pathway is necessary to understand the role of this pathway in mycobacterial survival and infection as well as to define novel targets for drug development.

The specific aims include:

1. Examine the structure-activity relationship of azasteroid inhibitors against 3 β -hydroxysteroid dehydrogenase. We employed a series of azasteroids as a tool set to determine the structure-activity relationship (SAR) for *M. tuberculosis* 3 β -HSD. The SAR would provide insight into the true substrate for the enzyme *in vivo*. In addition, inhibitors of *M. tuberculosis* 3 β -HSD would target the cholesterol metabolic pathway, which has been shown to be important for pathogenesis.

2. Elucidate the cholesterol-derived metabolite profile of Δ *igr* H37Rv *M. tuberculosis*. We applied ¹³C-metabolite profiling using isotopically labeled cholesterol to study the cholesterol metabolism pathway in *M. tuberculosis*. The Δ *igr* and wild-type strains were

investigated to identify metabolites that accumulate or disappear upon disruption of the *igr* operon, towards understanding the functional role of the *igr* operon and the molecular cause of the growth attenuation in the Δ *igr* strain.

3. Investigate the biochemical function of the gene products of the *igr* operon. Using recombinant enzymes we investigated the enzymatic activities encoded by the *igr* operon. We aim to understand the functional role of these proteins in cholesterol metabolism.

Chapter 2

Results and Discussion

I.	Inhibition of the <i>M. tuberculosis</i> 3 β -hydroxysteroid dehydrogenase by azasteroids ^a	35
II.	Pathway profiling in <i>M. tuberculosis</i> : elucidation of a cholesterol-derived catabolite and the enzymes that catalyze its metabolism ^b	43
III.	Biochemistry of acyl-CoA dehydrogenase FadE28-FadE29 from <i>M. tuberculosis</i>	62
IV.	Biochemistry of 2-enoyl-CoA hydratase Rv3541c-Rv3542c from <i>M. tuberculosis</i>	76

^a This section is adapted from published work in *Bioorganic & Medicinal Chemistry Letters* (164).

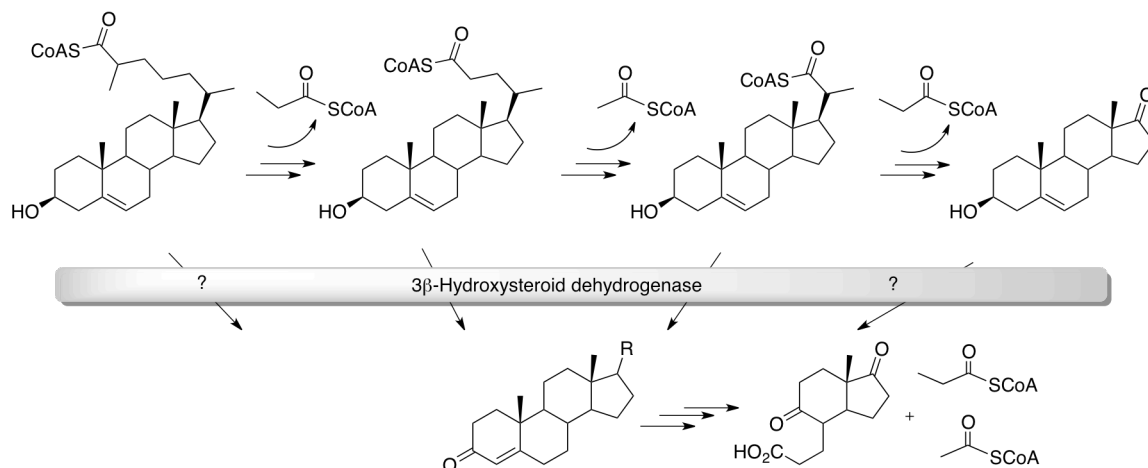
^b This section is adapted from published work in *The Journal of Biological Chemistry* (165).

I. Inhibition of the *M. tuberculosis* 3 β -hydroxysteroid dehydrogenase by azasteroids

3 β -hydroxysteroid dehydrogenase (3 β -HSD) catalyzes the oxidation and isomerization of Δ^5 -3 β -hydroxysteroids to Δ^4 -ketosteroids. In mammals, this enzyme is required for the biosynthesis of steroid hormones. Pregnenolone, 17-hydroxypregnenolone, DHEA, and androst-5-ene-3,7-diol are all substrates of mammalian 3 β -HSDs (112). 3 β -HSD orthologs have been identified in plants, fish, amphibians, viruses, and actinomycete bacteria (67, 166-170). Their respective substrate specificities and metabolic functions are not as well established.

3 β -HSD (Rv1106c) is proposed to catalyze the first step of the cholesterol metabolism pathway of *M. tuberculosis* (Scheme 1-5). However, the physiological substrate specificity has not been fully investigated. *M. tuberculosis* 3 β -HSD also can catalyze the oxidation and isomerization of DHEA and pregnenolone to their respective α,β unsaturated ketones with equal efficiency (73). Direct comparison of substrate specificities is difficult in this system because the conditions employed to solubilize the steroids differ, and substrate inhibition by the NAD⁺ cofactor is observed at millimolar concentrations. Therefore, the relative binding affinities for the enzyme are not readily derived from the kinetic experiments.

As a consequence, the precise sequence of catalytic events in the *M. tuberculosis* cholesterol metabolism pathway is not known. In the predicted pathway for *M. tuberculosis* cholesterol metabolism, the steroid skeleton can undergo oxidative degradation simultaneous with side-chain truncation (Scheme 2-1). The substrate preference of each enzyme in the pathway has been explored to a limited extent with the exception of 3-ketosteroid dehydrogenase (KstD), which is suggested to prefer the 5 α -androstane-3,17-dione and 5 α -testosterone as substrates (78), and the *meta*-cleavage product hydrolase HsaD for which X-ray crystal structures suggest a preference for 4,9-DHSA as the substrate (94). In order to develop antibacterial therapies that target the pathway and to understand the metabolic consequence of blocking enzymes within the pathway, the preferred sequence of reactions must be elucidated.



Scheme 2-1. Partial cholesterol metabolism pathway in *M. tuberculosis*. The physiological substrate of 3 β -hydroxysteroid dehydrogenase is not clear.

Azasteroids are validated, therapeutically-useful compounds that inhibit enzymes in steroid biosynthetic pathways (Figure 2-1). For example, finasteride and dutasteride inhibit 5 α -reductase-catalyzed production of dihydrotestosterone from testosterone. Inhibition of 5 α -reductase is indicated for treatment of benign prostatic hyperplasia and some prostate cancers. Investigation of the 6-azasteroid series was undertaken by Frye and coworkers at GlaxoWellcome during SAR work to develop enzyme selectivity (171). However, cross-reactivity with human 3 β -HSD could not be eliminated and development efforts were refocused on the 4-azasteroid series (172-174).

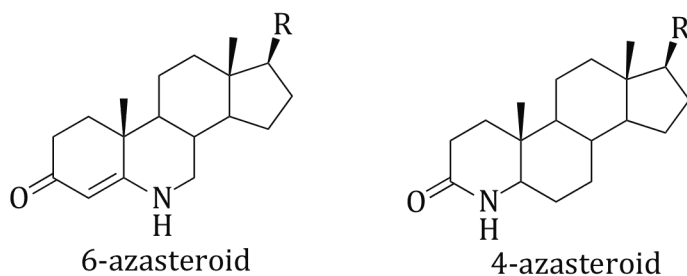
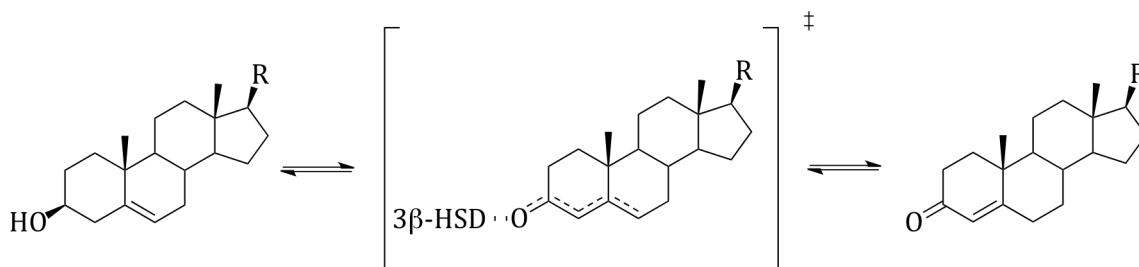


Figure 2-1. Azasteroid frameworks.

M. tuberculosis 3 β -HSD shares 29% amino acid sequence identity with type I and type II human 3 β -HSD (UniProtKB ID P14060 and P26439) and these enzymes catalyze the

same reaction. Both the active site catalytic tetrad and Rossmann fold motif for NAD⁺ cofactor binding are conserved. As the 6-azasteroid moiety is proposed to act as a transition state mimic of the 3 β -HSD-catalyzed reaction (172), we expected that the transition state analogy would apply to the *M. tuberculosis* enzyme (Scheme 2-2). Given the low amino acid identity of non-catalytic residues between orthologs, we did not expect inhibitor specificity of the *M. tuberculosis* enzyme to necessarily parallel that of the human enzyme.



Scheme 2-2. Reaction catalyzed by 3 β -HSD.

We reasoned that a comprehensive study of 6-azasteroids would provide rapid entry into the SAR of *M. tuberculosis* 3 β -HSD and insight into the true substrate for the enzyme *in vivo*. Moreover, 6-azasteroids have excellent biodistribution and pharmacokinetic properties in humans (173). Inhibitors of *M. tuberculosis* 3 β -HSD are important for targeting the cholesterol metabolic pathway and would require little development before *in vivo* analysis of enzyme inhibition could be undertaken. Here we report the *in vitro* inhibition SAR for *M. tuberculosis* 3 β -HSD using a family of azasteroids to explore the enzyme specificity.

Three series of azasteroids were tested to survey the importance and tolerance of substituents at a) the 17-position of the D-ring, b) the 4-7-positions of the A- and B-rings, and c) the 1,2 positions of the A-ring (Table 2-1). In order to identify the most potent compounds, the IC₅₀'s for 21 different azasteroids were measured at the K_M of DHEA (120 μ M) and at 2x K_M of NAD⁺ (400 μ M). Previously, we had demonstrated that *M. tuberculosis* 3 β -HSD follows a compulsory order mechanism in which NAD⁺ binds first (73). Therefore, we expected that competitive inhibitors of steroid binding would bind to the E-NAD⁺ complex. However, inhibitors were tested with a less than fully saturating concentration of

NAD⁺ because substrate inhibition occurs at millimolar levels of the cofactor (73). The IC₅₀'s were determined by fitting the initial velocities at 8 different inhibitor concentrations ranging from 6 nM to 400 μM to Equation 3-1. The maximum concentration that was used ranged from 50-400 μM due to the limited solubility of some of the azasteroids.

The mechanism of inhibition was determined for azasteroids **3**, **7**, and **17**, which had IC₅₀'s that varied over 2 orders of magnitude. We measured steady-state rates as a function of both DHEA and inhibitor concentrations and globally fit the data to Equation 3-2. All three inhibitors were found to be competitive inhibitors of DHEA. We concluded that modifications of the steroid ring framework that reduced the efficacy of inhibition did not alter the mechanism of inhibition. Kinetic competition with DHEA is consistent with the proposal that the 6-azasteroid binds in the steroid binding site and that the IC₅₀'s serve as a valid indicator of relative binding to the steroid binding site.

Functional groups of varying polarity were employed to interrogate the importance of the 6-azasteroid side chain at C17. Carboxylic acid **1**, did not inhibit *M. tuberculosis* 3β-HSD at concentrations below 200 μM. In contrast, incorporation of an amide at the 17-carbon was tolerated. Aromatic amide substituents, e.g., **3-5**, were effective inhibitors, whereas polar heterocycle **2** obviated efficacy. Further constraint of the amide as an oxazole **6** improved inhibition a further 3-fold. Elimination of heteroatoms at C17 and substitution of the 6-azasteroid with the native cholesterol 8-carbon side chain to provide azasteroid **7** increased potency an order of magnitude compared to the oxazole. The binding preference for large hydrophobic substituents indicates that C27 sterols are most likely the preferred substrate for *M. tuberculosis* 3β-HSD.

Addition of a methyl group at C7 to the sterol framework of **7** to provide **8** reduced potency 10-fold. Next, we explored structure-activity space around the A,B ring system using a simple amide substituent at the C17 position. When the C17 was modified as the t-butyl amide, **9**, the inhibition potency was poorer than the aromatic substituents. The diethyl amide substituent was selected for further analysis due to its facile synthesis and to compare inhibition constants in a readily measured concentration range. Substitution of the 6-azasteroid framework with linear alkyl moieties at either C4 or N6, **10-15**, reduced potency, and in most cases also reduced solubility. Alkylation at both C4 and N6 in a

bridged fashion, **16**, restored potency. Interestingly, swapping N6 and C4 to provide the 4-azasteroids **17** (finasteride) and **18** (dutasteride) had a deleterious effect on potency, which is consistent with the 6-azasteroid framework mimicking the intermediate in the isomerization half-reaction (Scheme 2-2).

Introduction of unsaturation at C1-C2 in azasteroid **9** was tolerated. However, cyclopropanation across C1-C2 in azasteroids **19** and **20** reduced efficacy at least 5-fold. Therefore, the steric requirements about the A,B ring system are quite stringent for the *M. tuberculosis* 3 β -HSD.

This 6-azasteroid series shows potent inhibition against human adrenal 3 β -HSD. Increased bulk at C17, e.g., trifluoromethyl aryl substituents like in azasteroid **3**, decreased potency (174). Moreover, the human adrenal 3 β -HSD tolerated methylation at N6 and cyclopropanation at C1-C2 (172). However, larger N6 substituents, e.g, butyl and hexyl reduced potency against the human enzyme 1000-fold (172). Therefore, future work to develop *M. tuberculosis* specific inhibitors should focus on elaboration of the C17 structure-activity space and introduction of constrained substituents between N6 and C4.

In our screen of 20 compounds, azasteroid **7**, which most closely mimics cholestenone, was the most effective inhibitor. Further inhibition analyses were carried out with azasteroid **7** in which the concentrations of both DHEA and NAD⁺ were varied. The azasteroid competitively inhibits DHEA binding ten times more effectively than NAD⁺ binding (Table 2-2, Figure 2-2, and Figure 2-3). Moreover, the azasteroid shows strong uncompetitive inhibition with respect to NAD⁺. Previously, we demonstrated that catalysis by *M. tuberculosis* 3 β -HSD follows a compulsory order binding mechanism in which NAD⁺ binds first (73). The competitive behavior versus DHEA and uncompetitive behavior versus NAD⁺ suggests that the azasteroid binds most effectively to the E-NAD⁺ complex. Therefore, increasing levels of NAD⁺ cofactor cannot overcome inhibition of 3 β -HSD by azasteroids.

Table 2-1. Inhibition of *M. tuberculosis* 3 β -HSD by azasteroids^a

Structure	No	R	IC ₅₀ M. tb 3 β -HSD (μ M)	Structure	No	IC ₅₀ M. tb 3 β -HSD (μ M)
	1	OH	> 200		6	9 \pm 1
	2	NH-phenyl-4-morpholino	> 100		7	0.5 \pm 0.1
	3	NH[(2-t-butyl, 5-CF ₃) phenyl]	27 \pm 3			
	4	NH-2-benzophenone	27 \pm 2			
	5	NH[(2,5-bis-t-butyl) phenyl]	23 \pm 2			
R				R ₁ , R ₂		
	8		8 \pm 2		11	> 400
	9	NHC(Me) ₃	140 \pm 15		12	> 260
					13	156 \pm 15
					14	> 50
					15	> 50
					16	43 \pm 3
R				R		
	17	NHC(Me) ₃	312 \pm 18		19	> 200
	18	NH[(2,5-bis-CF ₃) phenyl]	> 150		20	540 \pm 390

^aExperimental Conditions: 100 mM TAPS pH 8.5, 150 mM NaCl₂, 30 mM MgCl₂, 2% DMSO, at 30 °C. Fixed concentrations of DHEA and NAD⁺ were 120 μ M and 400 μ M, respectively. Azasteroids were tested up to 400 μ M or their maximum solubility, whichever was less.

Table 2-2. Inhibition characteristics of azasteroid **7**^a

Substrate	Pattern type	K_{ic}	K_{iu}
DHEA	C	124 ± 6 nM	n.o. ^b
NAD ⁺	Mixed	870 ± 150 nM	98 ± 12 nM

^aExperimental Conditions: 100 mM TAPS pH 8.5, 150 mM NaCl₂, 30 mM MgCl₂, 2% DMSO at 30 °C. Fixed concentrations of DHEA and NAD⁺ were 120 μM and 400 μM, respectively. C: competitive inhibition; Mixed: mixed competitive and uncompetitive inhibition. K_{ic} and K_{iu} : competitive and uncompetitive inhibition constants, respectively. ^bNot observed.

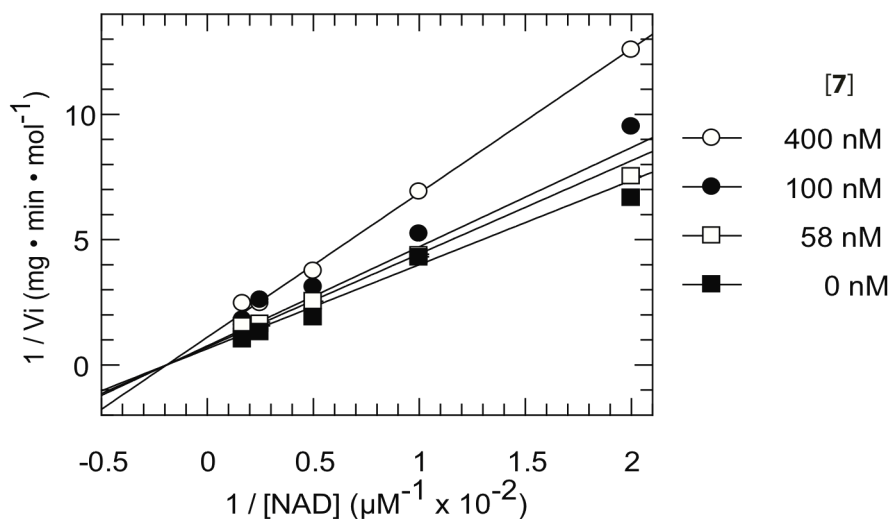


Figure 2-2. Kinetics of *M. tuberculosis* 3β-HSD inhibition by **7** at varied concentrations of NAD⁺. DHEA concentration was fixed at 120 μM.

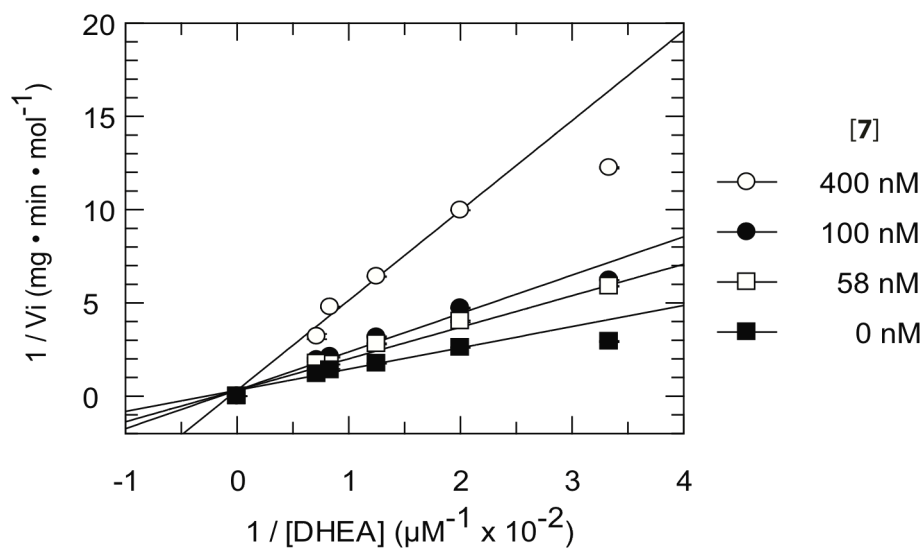


Figure 2-3. Kinetics of *M. tuberculosis* 3 β -HSD inhibition by **7** at varied concentrations of DHEA. NAD⁺ concentration was fixed at 400 μ M.

In summary, three series of azasteroids were evaluated for binding and inhibition of *M. tuberculosis* 3 β -HSD dehydrogenase. 6-Azasteroids with large, hydrophobic side chains at the C17 position are the most effective inhibitors. Substitutions at C1, C2, C4 and N6 were poorly tolerated. Our structure-activity studies indicate that the 6-aza version of cholesterol is the best and tightest binding inhibitor and are consistent with cholesterol being the preferred substrate of *M. tuberculosis* 3 β -HSD dehydrogenase.

II. Pathway profiling in *M. tuberculosis*: elucidation of a cholesterol-derived catabolite and the enzymes that catalyze its metabolism^a

In order to pursue a full analysis of cholesterol metabolism in the Δigr mutant strain of *M. tuberculosis* H37Rv, we required tools that trace the fate of the B-D rings and additional carbons of the sterol side chain. Here we describe the biosynthetic preparation of isotopically labeled [1,7,15,22,26-¹⁴C]-cholesterol or [1,7,15,22,26-¹³C]-cholesterol. The distribution of labels throughout the sterol ring system and side chain makes these reagents useful metabolic tracers to study cholesterol metabolism in *M. tuberculosis*.

*Biosynthetic preparation of isotopically labeled LDL-cholesterol.**

Mevalonate is the first dedicated precursor required for cholesterol biosynthesis (175-177). To metabolically label cholesterol in low-density lipoprotein (LDL) particles, HepG2 (human liver) cells were cultured with either 2-[¹³C]- or [¹⁴C]-mevalonolactone, to produce LDL particles containing [1,7,15,22,26-¹³C]-cholesterol or [1,7,15,22,26-¹⁴C]-cholesterol (Figure 2-4) (178). Mevalonolactone is the soluble and membrane permeable form of mevalonate that is assimilated by cellular metabolism. Mevastatin, the HMG-CoA reductase inhibitor, was added to suppress the cellular conversion of unlabeled acetyl-CoA pools to mevalonate, which otherwise would be incorporated into cholesterol and reduce the heavy isotope incorporation.

The ¹³C- or ¹⁴C-labeled cholesterol was isolated as soluble LDL particles from the HepG2 culture supernatant and unincorporated mevalonolactone removed by ultrafiltration. TLC analysis of the LDL lipids demonstrated that cholesterol and cholesterol-esters contained ¹⁴C-labeled carbons (Figure 2-5A). The LDL-derived phospholipid and triacylglycerol components contained no ¹⁴C label, suggesting that the label observed in cholesterol ester is likely not located in the esterified fatty acid. MALDI-TOF mass spectra of LDL-derived [¹³C]-cholesterol contained the [M-H₂O+H]⁺ dehydration product ion (m/z= 369), as well as an [M-H₂O+H+5]⁺ ion (m/z = 374) that corresponds to

^a These experiments were performed in collaboration with Brian C. VanderVen and David G. Russell. Specific experiments performed by B. C. V. and D. G. R. are indicated by an asterisk.

the expected cholesterol labeled with five ^{13}C atoms (Figure 2-5B). The $m/z=369$ ion had a natural isotope envelope for the +1 to +4 ions indicating that partially labeled cholesterol was not formed. The isotopic incorporation ranged from 10 to 30% depending on the preparation. DEPT135 NMR analysis was used to confirm the positions of the five ^{13}C labeled atoms (Figure 2-5C). ^{13}C resonances were compared to known cholesterol resonances, confirming labeling at the C1, C7, C15, C22, and C26 positions (179).

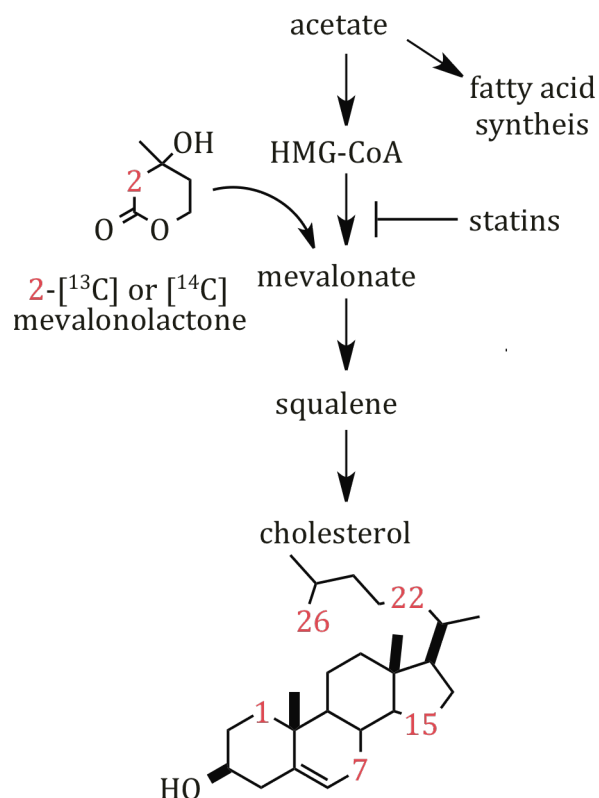


Figure 2-4. ^{13}C - or ^{14}C - metabolic labeling of LDL-derived cholesterol. Isotopically labeled 2- ^{13}C - or ^{14}C -mevalonolactone is converted into mevalonate and is incorporated into the cholesterol biosynthesis pathway resulting in labeled [1,7,5,22,26- ^{13}C]-cholesterol or [1,7,5,22,26- ^{14}C]-cholesterol. HMG-CoA: 3-hydroxy-3-methylglutarate coenzyme A.

*A ^{14}C -labeled cholesterol-derived metabolite accumulates in the Δ igr mutant when supplied LDL-[1,7,15,22,26- ^{14}C]-cholesterol. **

M. tuberculosis wild type H37Rv (WT), *igr* knockout strain (Δ igr) and Δ igr complement strain were grown in 7H9 OADC media, supplemented with LDL-[^{14}C]-cholesterol or nonradioactive LDL- cholesterol. After two weeks, extracts were isolated

from the pellet and culture media of each culture, separately. LC/MS analysis of the bacterial lipid extracts from nonradioactive LDL-cholesterol cultures confirmed that no free cholesterol remained in WT, *Δigr*, or complemented *Δigr* extracts. After two weeks of culture with LDL-[¹⁴C]-cholesterol, ¹⁴C counts were two-fold higher for *Δigr* mutant extracts compared to WT or complemented *Δigr* strains consistent with complete metabolism to CO₂ or low molecular weight carbon in wild type. In the *Δigr* strain, the metabolism of cholesterol is incomplete. The majority of the remaining ¹⁴C counts (85%) in the lipid extracts from the *Δigr* strain was secreted into the culture supernatant. TLC analysis revealed that a new polar species was formed in the *Δigr* strain that was absent in extracts from the WT and complemented *Δigr* strains (Figure 2-6).

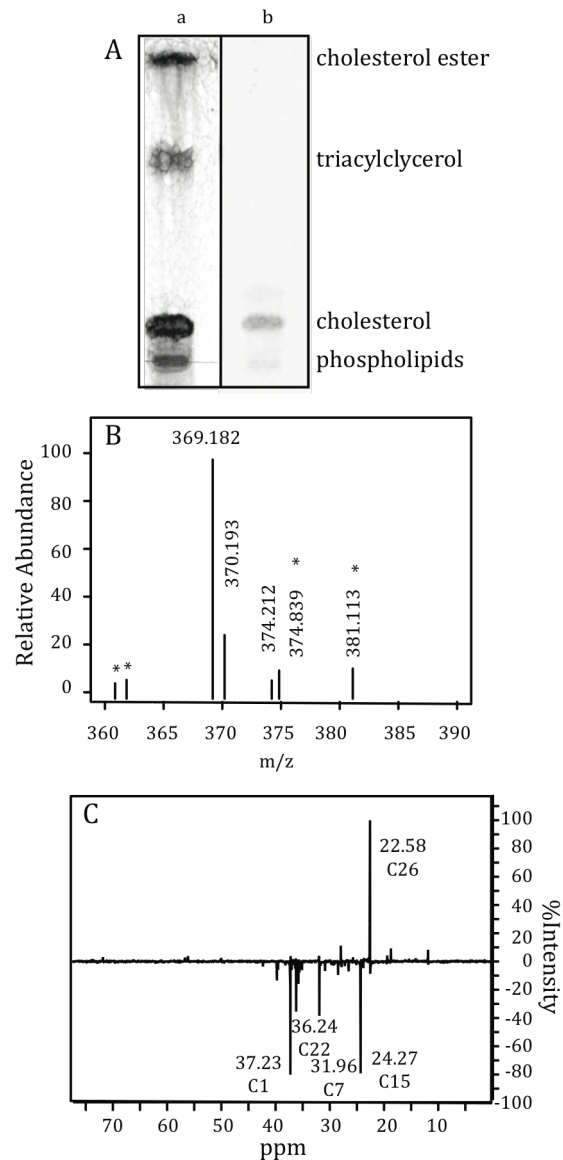


Figure 2-5. Analysis of metabolically labeled LDL- ^{13}C - or - ^{14}C -cholesterol. (A) TLC analysis of LDL- ^{14}C -cholesterol developed by (a) charring and (b) autoradiography. (B) MALDI-TOF MS spectrum of LDL- ^{13}C -cholesterol. Peaks labeled with * are matrix ions. (C) DEPT135 spectrum of isolated LDL- ^{13}C -cholesterol, confirming the position of isotopically labeled carbons.

Application of LDL-[1,7,15,22,26- ^{13}C]-cholesterol as an isotopic tracer of cholesterol metabolism intermediates by LC/MS.

To further characterize partially degraded cholesterol metabolites in *M. tuberculosis*, the bacteria were cultured with LDL-[1,7,15,22,26- ^{13}C]-cholesterol and the

extracts were analyzed by liquid chromatography and high-resolution mass spectrometry using XCMS software for feature matching and detection. ^{13}C labeled, cholesterol-derived metabolites were identified by their unnatural ^{13}C isotopomers. The presence of isotopomers was further confirmed by comparison to natural abundance LDL-cholesterol metabolic profiles.

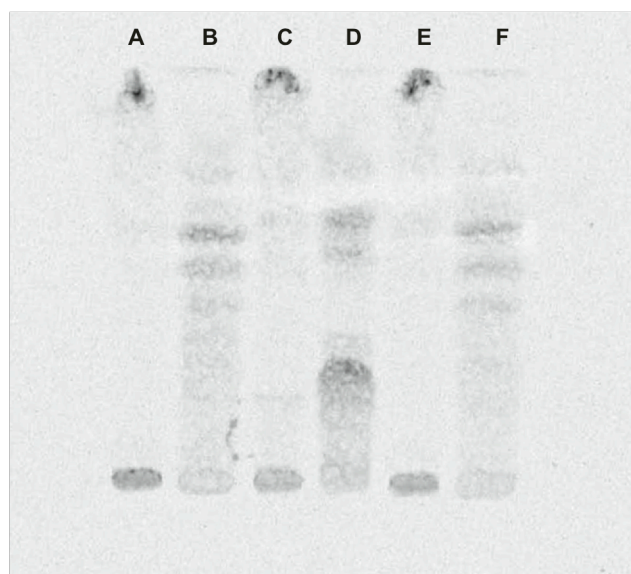


Figure 2-6.* TLC of *M. tuberculosis* extracts cultured with LDL-[1,7,15,22,26- ^{14}C]-cholesterol. 10000 cpm were loaded per lane. TLC was visualized by autoradiography. (A) *M. tuberculosis* H37Rv cell pellet. (B) *M. tuberculosis* H37Rv culture medium. (C) *M. tuberculosis* Δ *igr* cell pellet. (D) *M. tuberculosis* Δ *igr* culture medium. (E) *igr* complement cell pellet. (F) *igr* complement culture medium.

*Wild-type M. tuberculosis accumulates androst-4-ene-3,17-dione and androsta-1,4-diene-3,17-dione, while the Δ *igr* mutant does not.*

Four species with ^{13}C labels were observed in the WT and *igr* complemented strain that were not detected in Δ *igr* extracts (Figure 2-7 and Table 2-3). The two most prominent species, based on the peak area from the extracted ion chromatogram were androsta-1,4-diene-3,17-dione (ADD) and androst-4-ene-3,17-dione (AD) (Figure 2-7A). AD and ADD are known intermediates of the cholesterol metabolism pathway and had been previously identified in culture supernatants of *M. tuberculosis* H37Rv cultured with cholesterol (58). The molecular formula, obtained from high-resolution mass data and number of ^{13}C labels match their structures (Table 2-3A). The remaining two metabolites

observed in WT and *igr* complement extracts are keto and hydroxy oxidation products of AD and ADD (Table 2-3A). Interestingly, AD and ADD were not observed in the extracts isolated from the Δ *igr* mutant indicating that cholesterol metabolism is most likely blocked prior to AD and ADD formation in the Δ *igr* mutant strain.

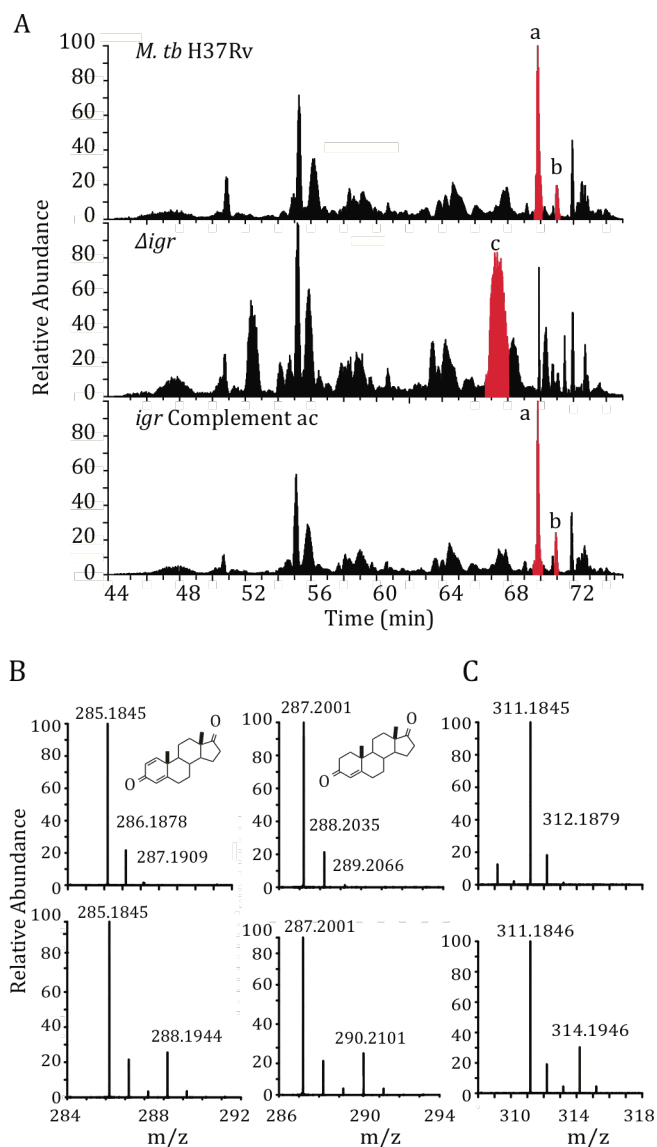


Figure 2-7. LC/MS analysis of *M. tuberculosis* H37Rv, Δ *igr*, and *igr* complement extracts cultured with LDL-[^{13}C]-cholesterol and natural abundance LDL-cholesterol for 2 weeks. (A) Total ion chromatograms (TIC) of *M. tuberculosis* H37Rv, Δ *igr*, and *igr* complement. Only the TIC from samples cultured with LDL-[^{13}C]-cholesterol are shown. Major metabolites are labeled, with a full list in Table 2-3. Metabolite peaks A and B were only observed in *M. tuberculosis* H37Rv and *igr* complement and peak C was only observed in Δ *igr*. (B) MS of *M. tuberculosis* H37Rv and *igr* complement cholesterol derived metabolite peaks A and B, which are androsta-1,4-diene-3,17-dione (ADD) and androst-4-ene-3,17-dione (AD), respectively. (C) MS of cholesterol-derived metabolite C that accumulates in the Δ *igr* mutant and is absent in the H37Rv and *igr* complemented strains. Top panels of B and C are from growth on natural abundance LDL-cholesterol and bottom panels from growth on isotopically labeled LDL-[^{13}C]-cholesterol.

Table 2-3. Cholesterol-derived metabolite parent ions (MH⁺ or MNa⁺) identified using LC/MS/MS.

A. WT H37Rv and igr complement ions						
Exact Mass	Molecular Formula	Calculated Mass	Retention Time (min)	Number of ¹³C labels	Compound	Integrated Peak Intensity
299.1638	C ₁₉ H ₂₃ O ₃	299.1647	61.9	3	4 + 0 - 2H	1.3E+07
303.1950	C ₁₉ H ₂₇ O ₃	303.1960	66.5	3	5 + 0	7.2E+06
285.1845	C ₁₉ H ₂₅ O ₂	285.1855	69.9	3	4	3.0E+08
287.2001	C ₁₉ H ₂₇ O ₂	287.2011	71.0	3	5	1.8E+08
B. Δ igr ions						
Exact Mass	Molecular Formula	Calculated Mass	Retention Time (min)	Number of ¹³C labels	Compound	Integrated Peak Intensity
355.1746	C ₁₈ H ₂₇ O ₇	355.1757	57.0	3		9.7E+06
297.1691	C ₁₆ H ₂₅ O ₅	297.1702	57.8	3	6	1.9E+08
295.1534	C ₁₆ H ₂₃ O ₅	295.1545	58.0	3	7	1.2E+08
283.1897	C ₁₆ H ₂₇ O ₄	283.1909	59.1	3		1.3E+08
281.1742	C ₁₆ H ₂₅ O ₄	281.1753	62.8	3		6.7E+07
355.2109	C ₁₉ H ₃₁ O ₆	355.2121	62.8	3		5.6E+07
398.1989	C ₁₃ H ₃₀ O ₇	298.1992	62.9	3		6.1E+07
709.4133	-		62.9	3		3.0E+06
295.1899	C ₁₇ H ₂₇ O ₄	295.1909	63.4	3		4.1E+08
277.1431	C ₁₆ H ₂₁ O ₄	277.1440	65.8	3		2.8E+07
323.1850	C ₁₈ H ₂₇ O ₅	323.1859	66.2	3		6.9E+07
311.1846	C ₁₇ H ₂₇ O ₅	311.1858	67.4	3	1	2.6E+09
309.1692	C ₁₇ H ₂₅ O ₅	309.1702	68.4	3	8	8.7E+08
279.1950	C ₁₇ H ₂₇ O ₃	279.1960	70.4	3		2.9E+08
337.2006	C ₁₉ H ₂₉ O ₅	337.2015	70.8	3		8.9E+07
744.5028	-		70.9	3		3.7E+06
293.1745	C ₁₇ H ₂₅ O ₄	293.1753	71.2	3		2.3E+07
325.2004	C ₁₈ H ₂₉ O ₅	325.2015	71.5	3	9	6.6E+07
305.2107	C ₁₉ H ₂₉ O ₃	305.2117	71.9	3		1.5E+07
343.2265	C ₂₂ H ₃₁ O ₃	343.2273	72.3	4	10	2.7E+07
345.2420	C ₂₂ H ₃₃ O ₃	345.2430	72.3	4	11	1.6E+07
351.2161	C ₂₀ H ₃₁ O ₅	351.2172	72.7	3		8.6E+06
371.2574	C ₂₄ H ₃₅ O ₃	371.2586	73.6	4	12	5.4E+07
429.2991	C ₂₇ H ₄₁ O ₄	429.3005	73.7	5	13	4.9E+06

Data were analyzed with XCMS. The exact masses were utilized to determine molecular formulae. Unnatural ¹³C isotopomers were used to identify cholesterol-derived compounds. Isotopomers and fragmentation were used to determine likely structures. Compound numbers correspond to the structures shown in Figure 2-8.

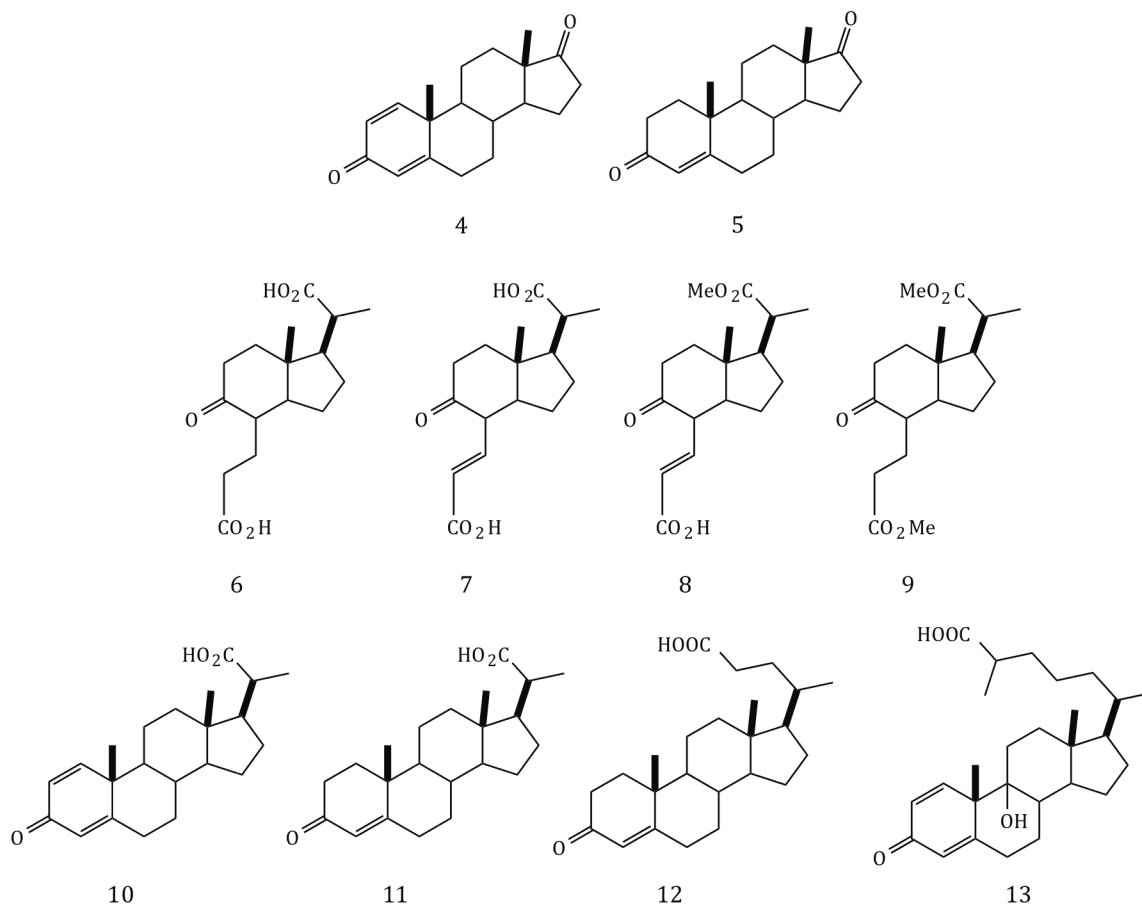


Figure 2-8. Assigned structures of Δigr metabolite ions identified by LC/MS/MS.

Identification of cholesterol derived metabolites present in Δigr extracts and absent in WT and igr complemented strains.

Several cholesterol-derived metabolites unique to Δigr extracts were identified by the unnatural isotopomer distribution of their parent ions in high-resolution mass spectra (Table 2-3B). Peak C was the major component based on peak area of the extracted ion chromatogram (Figure 2-7A and Table 2-3B) and the major ion had a $m/z = 311.1846$ (75%) and a minor ion with $m/z = 309.1692$ (25%). The remaining cholesterol-derived metabolites were present in 10- to 100-fold lower abundance.

Analysis of Δigr extracts cultured with unlabeled, free cholesterol dissolved in Tyloxapol micelles indicated that the metabolic profile is unchanged and is not dependent on culturing with LDL. Free cholesterol was then used as the carbon source to grow

sufficient quantities of *Δigr* cultures to isolate the main metabolite for spectroscopic and structural characterization.

Isolation and structural characterization of the major cholesterol metabolite, 1, formed by the Δigr strain.

Ethyl acetate extraction of *M. tuberculosis* H37Rv *Δigr* cultures grown with free cholesterol, followed by fractionation by reverse phase HPLC yielded the major metabolite **1** in ~85% purity. The metabolite structure was established by tandem mass spectrometry and ^1H , ^{13}C , COSY, ^1H - ^{13}C -HSQC, and HMBC NMR spectroscopy (Figure 2-8 and Figure 2-9).

The NMR spectra of **1** clearly mapped an indane carbon skeleton, presumably derived from the C and D rings of cholesterol. In the five-membered ring, the methine proton (δ 1.70) correlated to C1 (δ 51.8) in the ^1H - ^{13}C HSQC. The C1 methine correlated to the C2 methylene (δ 1.83; δ 1.52) by COSY. These methylene protons correlated to C2 (δ 27.7) in the ^1H - ^{13}C -HSQC spectrum. The C2 methylene coupled to the C3 (δ 24.6) methylene protons (δ 1.42; δ 1.74). The quaternary carbon, C7a (δ 42.6) in the ^{13}C spectrum showed an HMBC correlation to C3a (δ 55.1). C3a had an attached methine proton at δ 1.68.

The six-membered ring was established from C7a (δ 42.58) that correlated to the C8 (δ 10.5) methyl protons (δ 1.06) and C7 (δ 37.9) methylene protons (δ 1.65; 2.15) in the HMBC spectrum. The C7 methylene protons (δ 1.65;2.15) coupled to the C6 (δ 37.3) methylene protons (δ 2.24;2.60). A carbonyl was observed in the ^{13}C spectrum at δ 213, consistent with a ketone at C5.

This C5 ketone correlated to protons at δ 2.60, δ 2.24, δ 2.15, δ 1.68, and δ 1.76 in the HMBC. By ^1H - ^{13}C HSQC δ 1.68 and δ 1.76 were attached to C3'' (δ 22.6). Although no correlation to a proton at C4 was observed, this carbon was assigned to a signal of δ 49.5 with an attached methine proton at δ 2.51 by ^1H - ^{13}C HSQC correlation. By HMBC, C4 correlated to C3'' methylene protons (δ 1.68; δ 1.76). In the ^1H - ^{13}C HSQC, C2'' (δ 33.4) showed a correlation to methylene protons (δ 2.36; δ 2.33). In the HMBC, C3'' correlated to the C2'' methylene protons. Reciprocally, correlations were observed for C2'' and the C3'' methylene protons.

The C2'' methylene protons correlated to C1'' (δ 174.7). Based on the molecular formula of **1**, C1'' is predicted to be a carboxylic acid. However, by HMBC, C1'' also has long-range correlations with a singlet at δ 3.70, indicating that the carboxylic acid may be

partially esterified. The C1"-C3" propanoic acid group was assigned using long-range correlations making their assignment likely but not definite. COSY correlations were difficult to interpret due to co-purified metabolites of similar structure, primarily the dehydrated 3'-propenoate side chain ($m/z = 309.1692$), which we identified by mass spectral analysis. The NMR assignments made are consistent with published ^1H and ^{13}C data for $3\alpha\text{-H-}4\alpha\text{-(3'-propanoic acid)-}7\alpha\beta\text{-methylhexahydro-1,5-indadione}$ (DOHNAA) (95, 100).

The hexahydroindanone skeleton has been isolated from several steroid-metabolizing bacterial species including species of *Rhodococcus*, *Nocardia*, *Arthrobacter*, and *Streptomyces* (99-101, 180, 181). To our knowledge, this is the first report of the formation of hexahydroindanone species in *M. tuberculosis*.

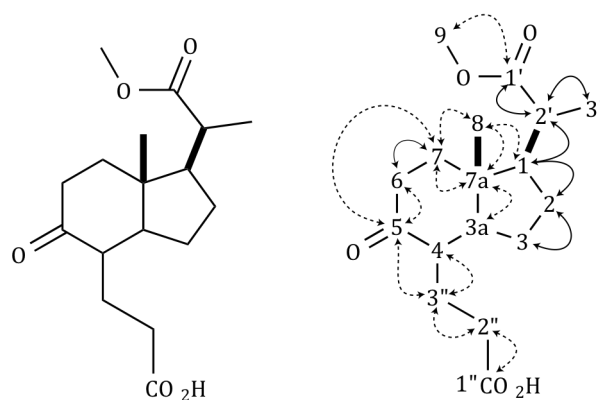
The presence of a three-carbon 2'-propanoate side chain on the five-membered ring at C1 was readily established. The carboxylate side chain was isolated as the methyl ester. The ^1H spectra showed two methyls at $\delta 1.21$ (s; 3H) and $\delta 3.67$ (d; 3H), which correlated to C3' ($\delta 16.1$) and C9 ($\delta 50.64$), respectively in the $^1\text{H-}^{13}\text{C}$ HSQC. The C1' carbonyl of the methyl ester was observed at $\delta 177$ in the ^{13}C NMR spectrum and correlated to the methyl at $\delta 3.67$ and C2' ($\delta 42.2$) methine at $\delta 2.52$ (dd, 1H) in the HMBC. The C2' methine is coupled to the C3' methyl ($\delta 1.21$) and C1 ($\delta 51.8$) methine ($\delta 1.70$) in the COSY spectrum. This assignment revealed that the side chain of cholesterol had been partially degraded, with a 2'-propanoate group remaining on what was formerly the D-ring.

In addition, the structure of the 2'-propanoate side chain was confirmed by tandem MS. Dehydrated parent ions ($\text{MH}^+ - 18$) of ^{12}C and ^{13}C -labeled metabolite **1** ($m/z = 293$ and 296 , respectively) were further fragmented (Figure 2-10). A loss of 32 Th and 50 Th was observed and attributed to the loss of MeOH and MeOH/CO respectively. A loss of 51 Th was observed for ^{13}C labeled **1** upon loss of MeOH/CO. This 1 Th increase indicates a loss of ^{13}C label and is consistent with the initial ^{13}C -labeling of cholesterol at C22 of cholesterol. A loss of 88 and 89 Th from $m/z 293$ and $m/z 296$ indicates a loss of $\text{C}_4\text{H}_8\text{O}_2$, which was assigned to the loss of the methyl 2'-propanoate side chain from C1 of **1**. The minor metabolite ($m/z = 309.1692$) corresponds to the same structure with an additional unsaturation. This metabolite has an identical fragmentation and labeling pattern to

metabolite **1**. The NMR and mass spectral data are consistent with an additional unsaturation in the moiety substituted at C4, which is most likely a 3'-propenoic acid.

We observed ions corresponding to the diacid ($m/z = 297.1691$) and the dimethyl ester ($m/z = 325.2004$) forms of **1** in the mass spectral profiles (Table 2-3). In addition, we observed the 3-oxo-4-pregnene-20-carboxylic acid precursor to **1**, which has all four steroid rings intact ($m/z = 345.2420$). Lastly, we observed the β -oxidation precursor to 3-oxo-4-pregnene-20-carboxylic acid ($m/z = 371.2574$) that has a five-carbon side chain.

Interestingly, metabolite **1** contains a 2'-propanoate side chain at C1, indicating the *igr* operon is not essential for the first two cycles of β -oxidation required to produce this shortened side chain from cholesterol. The *igr* operon encodes an incomplete cadre of β -oxidation enzymes, as it lacks genes for two key enzymes: 3-hydroxyacyl-CoA-dehydrogenase and 3-ketoacyl-CoA-thiolase.



1

Carbon No.	¹³ C (δ , ppm)	¹ H (δ , ppm)
9	50.6	3.67
1'	177.1	Q
3'	16.1	1.21
2'	42.2	2.52
8	10.5	1.06
1	51.9	1.70
2	27.7	1.83, 1.52
3	24.6	1.42, 1.75
3a	55.1	1.68
7a	42.6	Q
7	38.0	1.65, 2.15
6	37.4	2.60, 2.24
5	213.5	Q
4 ^a	49.5	2.51
3 ^{na}	22.6	1.68, 1.76
2 ^{na}	33.4	2.33; 2.36
1 ^{na}	174	Q

Figure 2-9. Structural characterization of *Δigr* metabolite **1**. Structure of **1** established from MS, tandem MS, and multidimensional NMR data. ¹H and ¹³C NMR signals are listed. Dotted arrows represent HMBC correlations. Solid arrows represent COSY correlations. Only key correlations are displayed. ^a Could not be assigned in COSY.

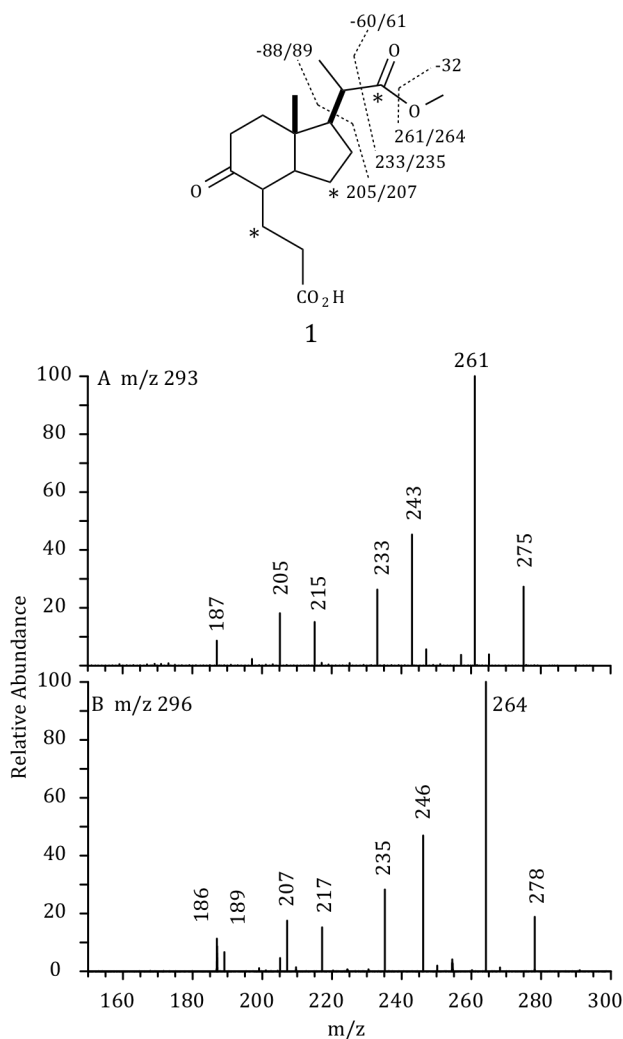


Figure 2-10. MS/MS/MS fragmentation analysis of Δ *igr* metabolite **1**. MS fragmentation data for the $[\text{MH}-18]^+$ ion of metabolite **1**. The loss of 18 Th is assumed to result from elimination of water from the enol tautomer. This fragmentation is not shown in the figure. (A) $[\text{MH}-18]^+ = m/z$ 293 (natural abundance) and (B) $[\text{MH}-18]^+ = m/z$ 296 (three ^{13}C labels). Sites of ^{13}C labeling are indicated by *.

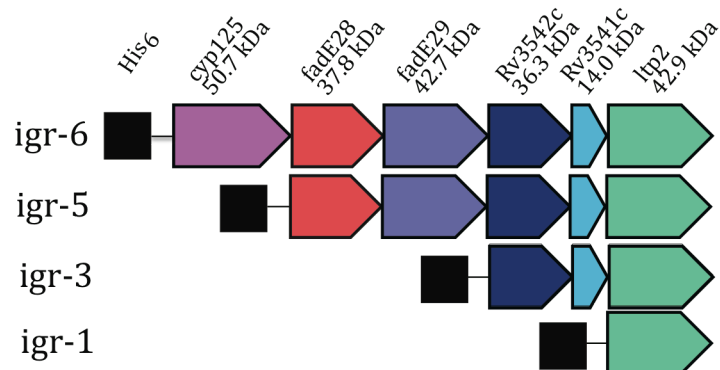
Proteins encoded in the igr operon form heteromeric complexes.

The native *igr* operon structure was used to heterologously express all six genes in *E. coli* using a single construct. The entire operon was cloned into expression vector pET28b and the first open reading frame of the operon was expressed as an N-terminal His₆-tagged protein for purification (Figure 2-11A, **igr-6**). Expression of construct **igr-6** resulted in isolation of soluble Cyp125 by immobilized metal affinity chromatography

(IMAC) purification. Next, *cyp125* was deleted to generate construct **igr-5**. Expression and purification by IMAC resulted in co-isolation of FadE28 and FadE29, even though FadE29 did not contain a His₆-tag. Therefore, these proteins are isolated as a complex. Then *fadE28* and *fadE29* were deleted to generate construct **igr-3**. Upon expression of **igr-3** and purification by IMAC, N-His₆ tagged Rv3542c and tagless Rv3541c were isolated, again indicating a protein complex was formed. Heterologous expression of His₆-tagged *ltp2* (**igr-1**) resulted in insoluble protein. Expression of FadE28, Rv3542c, or Rv3541c individually resulted in insoluble or unstable protein. Expression of FadE29 resulted in soluble, apoprotein (Chapter 2.3 and 2.4). Therefore, the protein expression data suggest that FadE28 forms a heteromeric complex with FadE29, and that likewise, Rv3542c forms a heteromeric complex with Rv3541c.

We proposed that these protein complexes catalyze the acyl-CoA dehydrogenation and enoyl-CoA hydration, respectively of the 2'-propanoate side chain to provide a quaternary alcohol. This alcohol would then readily undergo a retroaldol C1-C2' cleavage reaction catalyzed by Ltp2 to form the ketone at C1 and propionyl-CoA (Scheme 2-3). This cleavage is favorable because the thermodynamically stable ketone is formed. In contrast, conventional β -oxidation of a fatty acid requires oxidation of the 3-hydroxyacyl-CoA and thiolase cleavage because direct formation of the less stable aldehyde through retroaldol cleavage is thermodynamically uphill.

A



B

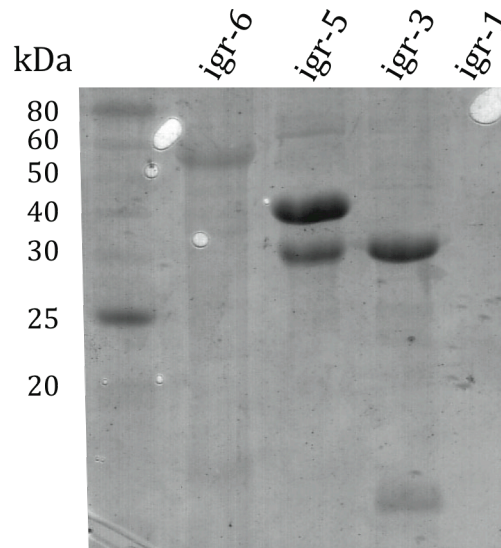
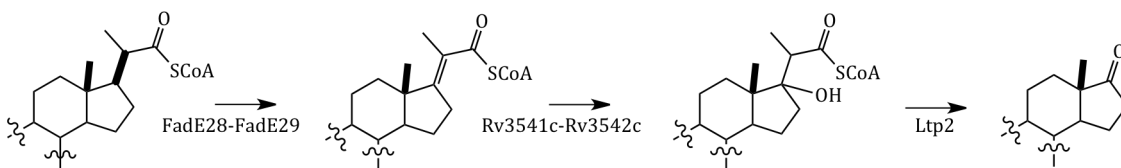


Figure 2-11. Expression and purification of the *igr* operon. (A) Constructs for expression of the *igr* operon in *E. coli* were prepared in vector pET28b. Each construct introduced an N-terminal His₆ tag on the first gene product for purification. (B) SDS-PAGE analysis of proteins purified by immobilized affinity chromatography from expression of constructs in A. Protein identities were confirmed by trypsin digest and MALDI-TOF MS fingerprinting.

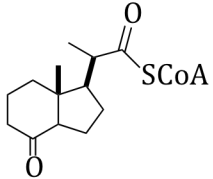
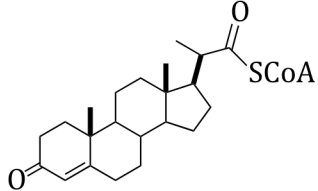


Scheme 2-3. Proposed function of the *igr* operon. The function of the *igr* operon was assigned to be degradation of the 2'-propanoate side chain. The proposed catalytic function of each enzyme or enzyme complex is shown.

FadE28-FadE29 catalyzes the dehydrogenation of 2'-propanoate-CoA esters of hexahydroindanone and pregnenone.

In order to test our hypothesis that the *igr* operon enzymes catalyze the metabolism of the 2'-propanoate substituent derived from the side chain of cholesterol, two potential polycyclic hydrocarbon substrates bearing a 2'-propanoate-CoA side chain were synthesized in addition to four short straight or branched fatty acyl-CoA esters (Table 2-4). Purified FadE28-FadE29 complex was assayed for ACAD activity. Several CoA thioester substrates, including hexahydroindanone **2** and pregnenone **3**, propionyl-CoA, butyryl-CoA, isobutyryl-CoA, and isovaleryl-CoA were assayed. Each of these substrates was assayed at 100 μM with up to 80 $\mu\text{g/mL}$ of FadE28-FadE29. Oxidation was detected spectroscopically and the formation of product was confirmed by MALDI-TOF mass spectrometry. FadE28-FadE29 catalyzed the dehydrogenation of **2** and **3**, but not of propionyl-CoA, butyryl-CoA, isobutyryl-CoA, or isovaleryl-CoA. Negative controls without enzyme or without substrate were conducted and no activity was detected. Thus, a short, straight or branched fatty acid is insufficient as a substrate. The specific activities of FadE28-FadE29 using **2** and **3** were 0.53 ± 0.07 and $2.38 \pm 0.11 \mu\text{mol min}^{-1} \text{mg}^{-1}$, respectively. FadE28-FadE29 shows a 5-fold preference for the pregnenone carbon skeleton over the hexahydroindanone skeleton under these assay conditions. This preference suggests that *in vivo*, the side chain can be metabolized to AD before further ring degradation occurs. This order of metabolism is further supported by the isolation of AD and ADD from wild-type extracts (58) and the reported substrate specificity of the 3-keto-5 α -steroid- Δ 1-dehydrogenase, KstD (78). However, a recent report suggests that the KshA/KshB hydroxylase is more specific for partial side-chain degradation intermediates (87). Most likely, the ring system and side chains are degraded in tandem.

Table 2-4. Specific activity data for FadE28-FadE29. Enzyme assays were conducted with ferrocenium hexafluorophosphate as the electron acceptor with 3.2 μg of enzyme and 100 μM of substrate at 25 $^{\circ}\text{C}$, pH 7.4.

	Substrate	Specific Activity ($\mu\text{mol min}^{-1}\text{mg}^{-1}$)
2		0.53 ± 0.07
3		2.38 ± 0.11
	Propionyl-CoA	na ¹
	Butyryl-CoA	na
	Isobutyryl-CoA	na
	Isovaleryl-CoA	na

¹na: no activity observed at FadE28-FadE29 concentrations up to 80 $\mu\text{g}/\text{mL}$

The 3'-propanoate substituent at C4 is also hypothesized to be degraded by β -oxidation to yield acetyl-CoA and the formate substituent. The absence of the above mentioned 3-hydroxyacyl-CoA-dehydrogenase and 3-ketoacyl-CoA-thiolase in the *igr* operon suggests that the *igr* enzymes do not catalyze the oxidation of this substituent. Consistent with this proposal, we observe minor metabolites corresponding to the C4-substituent β -oxidation intermediates in our NMR spectra. Moreover, van der Geize and coworkers recently reported that FadE30 of *Rhodococcus equi* (*R. equi*) is responsible for β -oxidation of the 3'-propanoate substituent at C4 of the 7ab-methyl-hexahydro-5-indanone skeleton (109). *R. equi* FadE30 shares 68% amino acid identity with FadE30 (Rv3560c) from *M. tuberculosis*. Gene knockout of *R. equi fad30* blocks growth of the bacterium on DOHNAA whereas *fadE30* growth on AD accumulates DOHNAA. These results strongly suggest that *fadE30* encodes the ACAD responsible for oxidation of the hexahydroindanone C4 substituent. *R. equi* FadE30 shares only 14% and 33% amino acid identity with FadE28 and FadE29 respectively. These low identities further support that the *FadEs* in the *igr* operon are not required for metabolism of C4 propanoate moiety.

In conclusion, we assign the primary function of the *igr* operon to be oxidation, hydration and retro-aldol cleavage of the 2'-propanoate side chain to provide AD and its ring-degraded analogs during cholesterol metabolism. Future studies will address understanding the mechanistic details of these enzymes.

III. Biochemistry of acyl-CoA dehydrogenase FadE28-FadE29 from *M. tuberculosis*

Fatty acids are metabolized by β -oxidation, the first step of which is catalyzed by an acyl-CoA dehydrogenase (ACAD). Based on early metabolite identification studies in multiple actinomycetes, it has been proposed that *M. tuberculosis* degrades the side chain of cholesterol by β -oxidation. In earlier work, we demonstrated that FadE28 and FadE29 form a heteromeric complex (165). This complex catalyzes the oxidation of 2-propanoyl-CoA side chains of polycyclic substrates. Moreover, the complex does not catalyze conversion of short straight or branched chain acyl-CoAs into enoyl-CoAs. Herein, we present full biophysical characterization of FadE28-FadE29.

FadE29 is isolated as a monomeric apoenzyme when expressed in E. coli and M. smegmatis

FadE28 and FadE29 were expressed separately in *E. coli* as N-terminal His₆ tagged proteins. FadE28 was found to be insoluble and localized to inclusion bodies. FadE29 was isolated and purified by IMAC, anion exchange, and size exclusion chromatography. UV-visible spectroscopy revealed the protein was purified as the holoprotein, without the flavin cofactor necessary for catalysis (Figure 2-12). Attempts to reconstitute with FAD were not successful.

FadE29 was analyzed by analytical size exclusion chromatography and analytical ultracentrifugation sedimentation equilibrium experiments (Figure 2-14 and Figure 2-15). The estimated molecular weight by analytical size exclusion chromatography was 52 kDa, which is close to the predicted monomeric molecular weight with a His₆ tag of 44.9 kDa. The analytical ultracentrifugation data fit an ideal model with a molecular weight of 41.9 kDa. Thus, under the conditions tested, FadE29, was monomeric. All characterized ACADs to date form homotetrameric, dimers of dimers, except for VLCAD, which assembles as a homodimer (Figure 1-4).

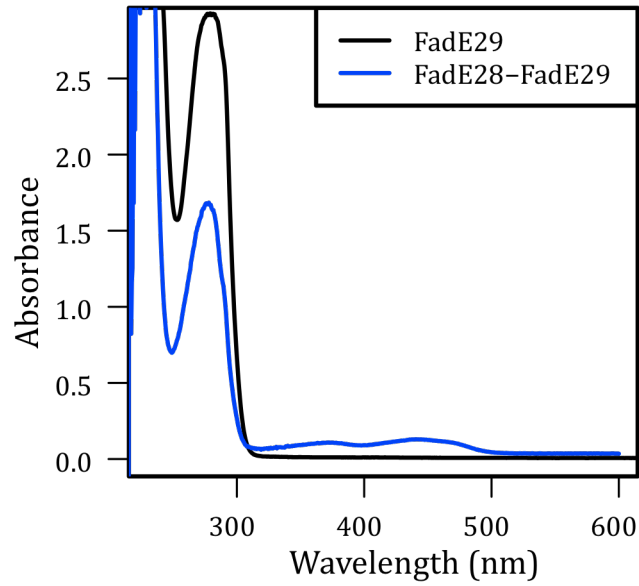


Figure 2-12. UV-visible spectra of purified FadE29 and FadE28-FadE29. FadE29 is shown in black and FadE28-FadE29 is shown in blue. FadE29 was obtained as the apoprotein while FadE28-FadE29 shows characteristic flavin absorbances.

FadE29 was then expressed in *Mycobacterium smegmatis* (*M. smegmatis*), a nonpathogenic, faster growing relative of *M. tuberculosis*. The protein was isolated by IMAC and further purified by anion exchange and size exclusion chromatography. Analysis by analytical size exclusion chromatography indicated that the protein elutes with the same retention time as monomeric FadE29 expressed in *E. coli* (Figure 2-14). Analysis of FadE29 expressed in *E. coli* and *M. smegmatis* demonstrates that the protein is not in the predicted quaternary structure of ACADs, lacks FAD, and therefore is inactive.

FadE28-FadE29 forms an obligate $\alpha_2\beta_2$ tetramer

Isolation of monomeric FadE29 apoprotein and insoluble FadE28, indicated an inability to fold to their native structure under the tested expression conditions. We proposed these proteins may form a complex due to their operonic organization and confirmed this hypothesis in previous work (165). FadE28-FadE29 complex was isolated upon coexpressing the proteins as a single cistronic construct from a single plasmid. In addition, in this work, we found that FadE28-FadE29 complex was also obtained upon expressing FadE28 and N-tagged FadE29 in trans on separate plasmids in *E. coli* harboring plasmid pG-KJE8, which encodes for 5 folding chaperones, dnaK-dnaJ-grpE-groES- groEL.

Coexpression and affinity purification resulted in isolation of both FadE28 and FadE29 indicating a protein complex is formed. In-gel tryptic digestion and MALDI-TOF MS indicates that FadE28 and FadE29 copurify regardless of which protein, FadE28 or FadE29, contains a His₆ tag (Figure 2-13). UV-visible spectroscopy of the FadE28-FadE29 complex shows the incorporation of a flavin cofactor in contrast to FadE29 purified from the single gene expression construct (Figure 2-12).

A. FadE28

Residue no.	Sequence
1-60	<u>MDFDPTAEQQAVADVVTSLERDISWEALVCGGVTALPVPERLGGDGVGLFEVGALLTEV</u>
61-120	<u>GRHGAVTPALATLGLGVVPLLELASAEQQDRFLAGVAKGGVLTAAALNEPGAALPDRPATS</u>
121-180	<u>FVGGRLSGTKVGVGYAEQADWMLVTADNAVVVVSPTADGVRMVRTPTSNGSDEYVMTMDG</u>
181-240	<u>VAVADCDILADVAHRVNQLALAVMGAYADGLVAGALRLTADYVANRQFGKPLSTFQTV</u>
241-300	<u>AAQLAEVYIASRTIDLVAKSIVWRLAEDLDAGDDLGVLYWVTSQAPPAMQICHHLHGGM</u>
301-339	<u>GMDVTYPMHRYSTIKDLTRLLGGPSHRELLGARCSLT</u>

B. FadE29

Residue no.	Sequence
1-60	<u>MGSSHHHHHSSGLVPRGSHMFIDLTPERQLQAEIRQYFSNLISPDERTEMEKDRHGPA</u>
61-120	<u>YRAVIRRMGRDGRLGVGWPKEFGGLGFGPIEQQIFVNEAHRADVPLPAVTLQTVGPTLQA</u>
121-180	<u>HGSELQKKKFLPAILAGEAHFAIGYTEPEAGTDLASLRTTAVRDGDHYIVNGQKVFTTGA</u>
181-240	<u>HDADYIWLACRTDPNAAKHKGISILIVDTKDPGYSWTPPIILADGAHHTNATYNDVRVPV</u>
241-300	<u>DMLVGKENDGWRLITTQLNNERVMLGPAGRFASIYDRVHAWASVPGGNGVTPIDHDDVKR</u>
301-360	<u>ALGEIRAIWRINELLNWQVASAGEDINMADAAATKVFGERVQRAGRLAEEIVGKYGNPA</u>
361-407	<u>EPDTAELLRWLDAQTKRNLVITFGGGVNEVMREMIAASGLKVPVPR</u>

Figure 2-13. Tryptic peptides from FadE28 and FadE29. Tryptic peptides, identified by MALDI-TOF MS, are mapped onto the sequence of (A) FadE28 and (B) FadE29. The sequences corresponding to the identified peptides are underlined.

Analysis of the FadE28-FadE29 complex by analytical size exclusion chromatography indicates a higher molecular weight complex is formed with an approximate molecular weight greater than 150 kDa (Figure 2-14). Further analysis by

analytical ultracentrifugation sedimentation equilibrium established that FadE28-FadE29 exists as a tetramer with a molecular weight of 156.4 kDa (Figure 2-15). The calculated molecular weight of the $\alpha_2\beta_2$ tetramer is 160.6 kDa. To confirm the stoichiometry of the complex, the assembly was analyzed by reverse phase liquid chromatography and ESI mass spectrometry (LC/UV/MS) (Figure 2-16). The absorption profile at 280 nm for the eluted peaks corresponding to FadE28 and FadE29 were integrated and using the extinction coefficient for each protein (Table 3-2), the relative concentrations of FadE28 to FadE29 were calculated to be in a 1 to 1 ratio. This ratio indicates that FadE28-FadE29 forms a $\alpha_2\beta_2$ complex. This is the first report of the identification of a heterotetrameric ACAD.

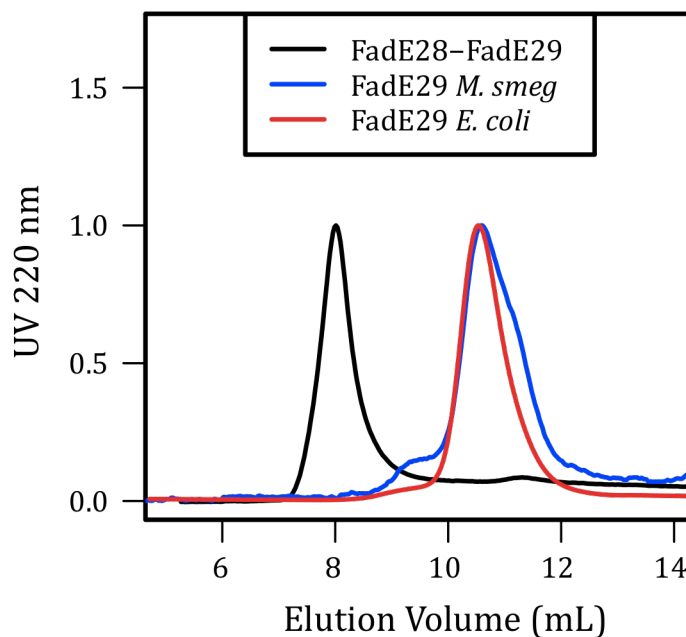


Figure 2-14. Analytical size exclusion chromatography of FadE29 and FadE28-FadE29. FadE29 expressed and purified in *E. coli* and *M. smegmatis* and the FadE28-FadE29 complex were analyzed by analytical size exclusion chromatography on a Superdex 75 column equilibrated in Buffer J. Several standard proteins were analyzed under the same conditions to generate a standard curve to estimate molecular weights of analyzed proteins. The approximated molecular weights of FadE29 and FadE28-FadE29 are 52 kDa and >150 kDa respectively. UV 220 nm signal was normalized to 1.

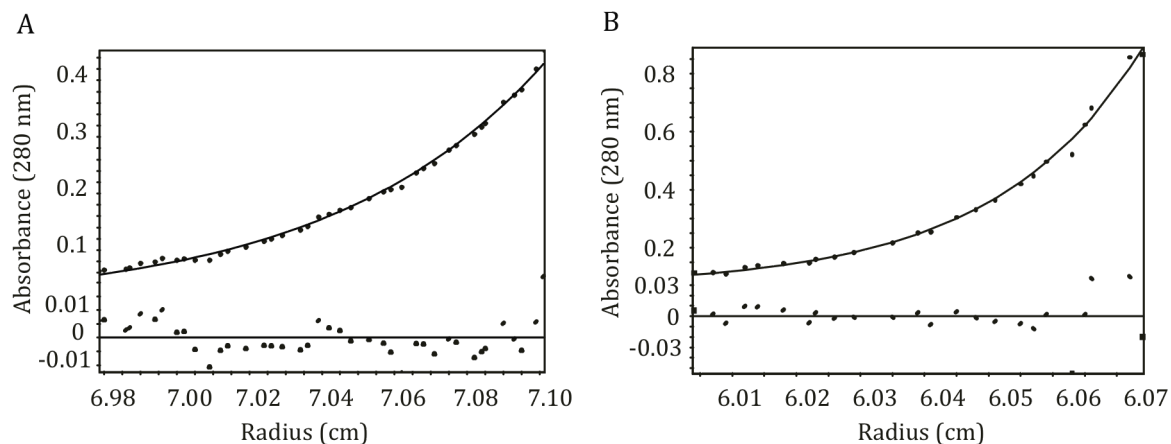


Figure 2-15. Analytical ultracentrifugation sedimentation equilibrium of FadE29 and FadE28-FadE29. (A) FadE29 (8.5, 3.4, and 1.7 μM) and (B) FadE28-FadE29 (10.6 μM , 5.3 μM , and 2.6 μM) were centrifuged at speeds of 20k, 25k, and 30k rpm at 20 $^{\circ}\text{C}$. Data were collected at 280 nm and fit globally resulting in molecular weights of 41.9 kDa and 156.4 kDa for FadE29 and FadE28-FadE29, respectively. A representative fit for each sample is shown.

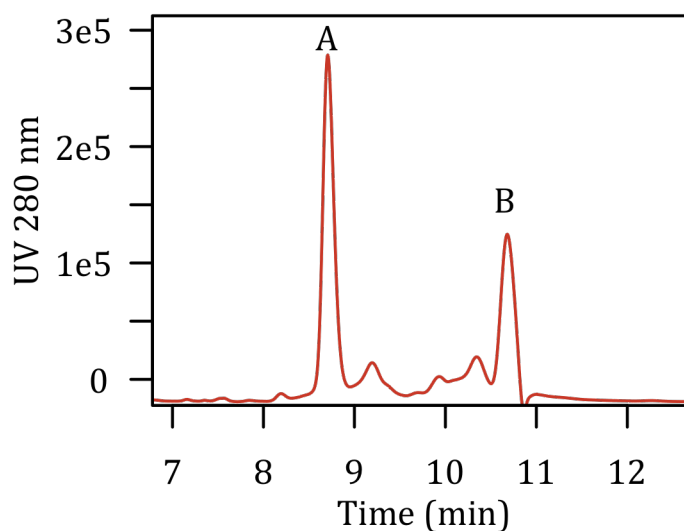


Figure 2-16. Reverse phase LC/UV chromatogram of FadE28-FadE29, monitoring at 280 nm. Peak A and peak B were identified as FadE29 and FadE28, respectively by ESI+ MS. The UV profiles of the peaks were integrated and relative concentrations were determined from the calculated extinctions coefficients, $\epsilon_{280}(\text{FadE29}) = 58900 \text{ M}^{-1} \text{ cm}^{-1}$ and $\epsilon_{280}(\text{FadE28}) = 35410 \text{ M}^{-1} \text{ cm}^{-1}$.

The flavin cofactor was isolated from FadE28-FadE29 by denaturing the protein with heat. The isolated flavin was analyzed by LC/UV/MS. The protonated FAD species at $m/z = 786$ was observed (Figure 2-17B) by positive ESI-MS. In addition the retention time and absorbance spectrum were identical to those of an FAD standard (Figure 2-17A).

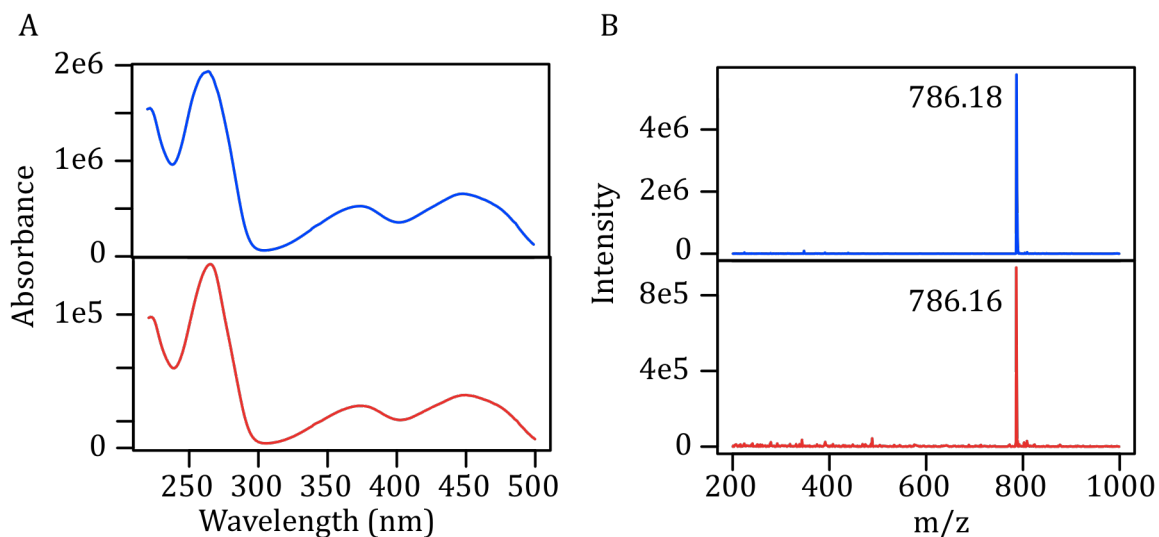


Figure 2-17. UV/LC/MS of isolated flavin from FadE28-FadE29. Shown are the (A) absorbance spectra and (B) ESI+ MS of isolated flavin in red and FAD standard in blue.

The FadE28-FadE29 $\alpha_2\beta_2$ complex has two cofactor binding sites

The FAD stoichiometry of the complex was determined spectrophotometrically. The protein concentration was determined using the calculated extinction coefficient of FadE28 and FadE29 at 280 nm and the concentration of FAD was calculated with extinction coefficient at 260nm. The FAD to protein molar ratio was calculated to be 1.5 ± 0.2 FAD molecules per $\alpha_2\beta_2$ FadE28-FadE29 tetramer. This result suggests that FadE28-FadE29 has two FAD bindings sites in contrast to typical ACAD tetramers that have four FAD binding sites.

To corroborate this unique occurrence in heterotetrameric FadE28-FadE29, we examined ligand protein interactions of FAD and ACADs in several crystal structures, including human SCAD (PDB 2VIG), MCAD (PDB 1EGC), and i3VD (PDB 1IVH). Typical tetrameric ACAD subunits are organized as dimer of dimers and each dimer has two FAD binding sites. The cofactor-binding site is comprised of both chains of the ACAD dimer and

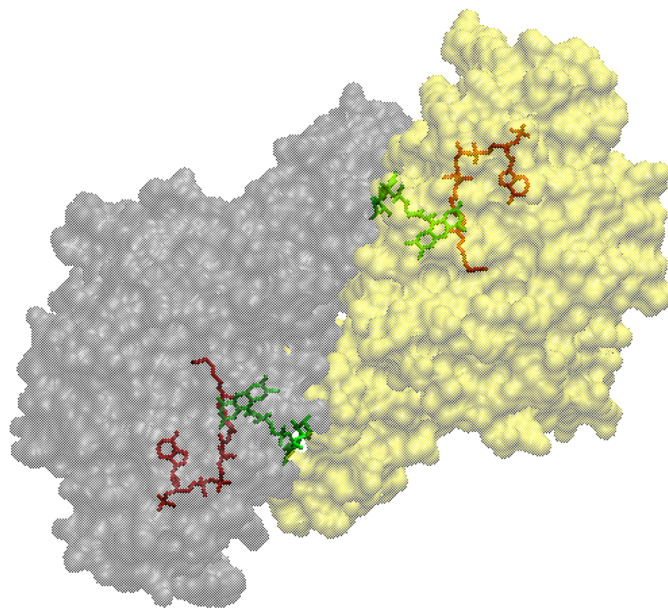
the adenosine of FAD lies at the interface. Each chain of the dimer makes interactions with adenosine in one binding pocket and interactions with riboflavin in the second binding pocket. Each monomer has one active site. Given the experimentally determined FAD stoichiometry, we expect the FAD binding residues are not fully conserved in both FadE28 and FadE29.

Polar side chains of Thr161, Ser167, Thr193, and Thr403 are involved in hydrogen bond formation with the riboflavin of FAD (numbering for MCAD). These residues are conserved Ser or Thr residues within the human ACAD family. Sequence analysis of FadE29 reveals that Thr161, Ser167, Thr193 are conserved in FadE29 as residues Thr126, Thr132, and Thr158. Thr403 aligns with Val367 in the sequence alignment and therefore is not conserved in FadE29. In contrast, sequence analysis of FadE28 reveals that Thr161, Ser167, Thr193, and Thr403 are not conserved in FadE28. These residues align with Asn107, Leu113, Gly134, and Pro325 in the sequence of FadE28. FadE29 is predicted to bind riboflavin. FadE28 lacks key conserved riboflavin binding residues and we predict it does not bind riboflavin.

Residues Arg306, Gln317, Gln374, and Gly378 form hydrogen bonds with the adenosine moiety of FAD. These residues are from the second chain of the homodimer of SCAD, MCAD, and i3VD. Sequence alignment with the other six human ACADS reveals they are highly conserved. Sequence analysis of FadE28 reveals that Arg306, Gln317, and Gly378 are conserved in FadE28 as residues Arg227, Gln238, and Gly299. Arg306, Gln317, Gln374, and Gly378 are not conserved in FadE29 and instead Pro265, Asp276, Glu331, and Lys335 are found in these positions. FadE28 is therefore predicted to bind adenosine. FadE29 lacks key conserved adenosine binding residues and we predict it does not bind adenosine.

The lack of conservation of riboflavin binding residues in FadE28 and adenosine binding residues in FadE29 strongly suggests that the FadE28-FadE29 tetramer has only two FAD binding sites. This is consistent with the experimentally determined FAD stoichiometry. In addition, this suggests that FadE28 forms a dimer with FadE29 and FAD binds at the heterodimer interface. The isolation of monomeric FadE29 suggests it does not self-associate to form a homodimer and supports this conclusion. Two FAD binding sites in tetrameric FadE28-FadE29 suggests that the complex has two active sites. The

conservation of riboflavin binding residues in FadE29 and not in FadE28 suggests that FadE29 is responsible for catalysis. This hypothesis is investigated herein.



SCAD	148	KIGCFALSEP	NGSDAGA	165	182	TKAWITNAW	190	294	AENRMAFGAPLTKLQVIQ	311
i2VD	170	KVGSFCLSEAGAGS	DSFA	187	197	SKMWISSAE	212	316	IKERIQFGKRLFDQGLQ	333
IBD	132	KFASYCLTEPGSGS	DAAS	149	166	SKAFISGAG	174	277	LNVRKQFGEPLASNQYLQ	294
MCAD	154	LMCAYCVTEPGAGS	DVAG	171	188	QKMWITNGG	196	303	ALEBKTFGKLLVEHQAIS	320
i3VD	158	YIGALAMSEPNAGS	DVVS	175	192	NKFWITNGP	200	306	LHVREAFGQKIGHFQLMQ	323
VLCAD	188	TVAAFCLTEPSSGSDAAS	205	224	SKLWISNGG	232	341	ATNRTQFGEKIHNFGLIQ	358	
VLCAD2	172	HIAAFCLTEPASGSDAAS	189	210	SKVWITNGG	218	326	ACTRKQFNKRLSEFGLIQ	343	
LCAD	166	CIGAIAMTEPGAGSDLQG	183	200	SKVFISNGS	208	314	VKQRKAFGKTVAHLQTVQ	331	
GD	173	LLGCFGLTEPNSGSDPSS	190	209	TKTWTNSP	217	316	ALDRMQFGVPLARNQLIQ	333	
FADE29	119	AHFAIGYTEPEAGTDLAS	136	153	QKVFTTGAH	161	262	ASVPGGNGVTPIDHDDVK	279	
FADE28	101	VLTA-ALNEPGAALPDRP	117	129	TKVGVGYAE	137	224	VANRKQFGKPLSTFQTV	241	

SCAD	363	AIQILGGM	370	389	IYEGTSEIQRLVIAGHL	406
i2VD	385	CIEWMGGV	392	412	IYEGASNIQLNTIAKHI	428
IBD	347	ALQMHGGY	354	396	ILEGSNEVMRILISRSL	412
MCAD	372	AVQILGGM	379	398	IYEGTSQIQRLIVAREH	415
i3VD	375	GIQCFGGN	382	402	IGAGTSEVRRLVIGRAF	418
VLCAD	411	CIQIMGGM	418	460	IFEGTNDILRLFVALQG	476
VLCAD2	397	ALQILGGL	404	424	IFEGTNEILRMYIALTG	440
LCAD	383	CVQLHGGW	390	410	IYGGTNEIMKELIAREI	426
GD	385	ARDMLGGM	392	412	TYEGTHDHALILGRAI	428
FadE29	329	AEEIVGKY	336	363	FGGGVNEVMREMIAASG	379
FadE28	293	CHHLHGM	300	321	LLGGPSHRELLGARCS	337

Figure 2-18. FAD binding residues in acyl-CoA dehydrogenases. FAD binds at the interface of two monomers. Shown is human MCAD (PDB 1EGC). FAD is shown in green and octanoyl-CoA in red. The FAD binding pocket was investigated in the crystal structures of human SCAD (PDB 2VIG), MCAD (PDB 1EGC), and i3VD (PDB 1IVH) to determine the key interactions for binding. FadE28 and FadE29 were aligned with characterized mammalian acyl-CoA dehydrogenases. Residues important for binding riboflavin are shown in red and residues important for binding adenosine are shown in blue.

The FadE28 sequence is missing canonical active-site residues

Bioinformatic analysis and sequence alignments with characterized ACADs indicate that FadE28 does not contain an obvious active site base necessary for catalysis. We used the following 9 ACAD structures and sequences as representative family members to perform sequence alignments: SCAD (P16219), MCAD (P11310), LCAD (P28330), VLCAD (P49748), VLCAD2 (Q9H845), i2VD (P45954), i3VD (P26440), iBD (Q9UKU7) and GD (Q92947) (Figure 2-20). Crystallography and mutagenesis studies have elucidated that the catalytic base, most often a glutamate, is located at one of two positions in helix G or helix J/K. These two positions orientate into the active site in the three-dimensional protein structures of ACADs. FadE29 has the conserved base in helix G, glutamate 240, like that of i3VD and LCAD. The glutamate in helix J/K is not conserved and is aligned with glycine 365 of FadE29. Interestingly FadE28 does not have an apparent catalytic base in either helix G or helix J/K, making it unclear if FadE28 is an active ACAD. The base's position in helix G or helix J/K align with glycine residues in FadE28. Therefore, we proposed that FadE29 glutamate-240 is the base responsible for catalysis of the complex and that there are two active sites per tetramer, consistent with the FAD stoichiometry and binding site analysis.

Mutagenesis suggests glutamate-240 of FadE29 is the active site residue of FadE28-FadE29

Glutamate-240 of FadE29 was mutated to a glutamine and the isolated protein was assayed for activity with hexahydroindanone **2**. No activity was observed spectroscopically or by MALDI-TOF MS. This preliminary results suggests that glutamate-240 is required for FadE28-FadE29 dehydrogenase activity. Full biophysical characterization of the FadE28-FadE29 E240Q mutant protein is underway, to verify the loss of activity is due to the mutation of E240 and not differences in secondary structure, oligomeric state, or flavin stoichiometry.

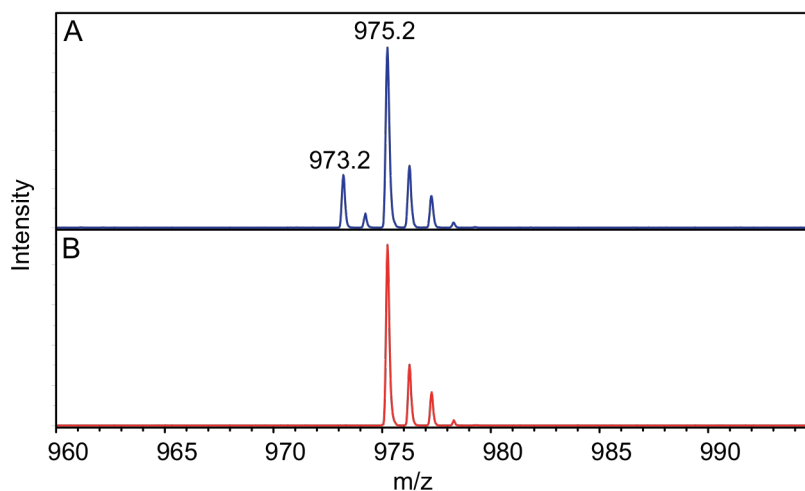


Figure 2-19. MALDI-TOF MS spectra of FadE28-FadE29 assay. 100 nM (A) FadE28-FadE29 wild-type and 100 nM (B) FadE28-FadE29_{E240Q} were assayed with hexahydroindanone **2**, m/z 975, with electron acceptor ferrocenium hexafluorophosphate at 20 °C. Dehydrogenation product, m/z 973, was not detected for E240Q mutant assay after several hours of incubation.

Our data suggest the $\alpha_2\beta_2$ tetramer has only two active sites. The ACAD family of enzymes has been extensively studied in eukaryotes and all except the VLACAD form homotetramers with 4 active sites. The very long chain classes of ACADs have two active sites though they form homodimers with an additional C terminal 180 residues that bind to the mitochondrial membrane. The heteromeric FadE complexes isolated from *M. tuberculosis* represent a third quaternary architecture for ACADs.

Bioinformatic analysis of the annotated FadE proteome of M. tuberculosis.

After our discovery of a novel heteromeric ACAD, FadE28-FadE29, with two active sites, we undertook a more extensive bioinformatic analysis of the *M. tuberculosis* ACAD proteome. The 36 annotated ACADs in the *M. tuberculosis* genome were aligned with the amino acid sequences of representative human ACAD from each of the 9 characterized classes. From the sequence alignment of the *M. tuberculosis* FadE proteome, in addition to FadE28, five additional annotated FadEs lack a conserved catalytic base in both helix G and helix J/K, including FadE3 (*Rv0215c*), FadE16 (*Rv1679*), FadE18 (*Rv1933c*), FadE27 (*Rv3505*), and FadE32 (*Rv3563*) (Figure 2-20A). The conserved ACAD glutamate has been

demonstrated to be essential for catalytic activity therefore the computational annotation of these genes as ACADs may be incorrect.

Analysis of the genomic context of *M. tuberculosis* FadEs lacking a catalytic base shows that, like FadE28, the genes encoding FadE18 and FadE27 are found within operons that encode an adjacent FadE with a conserved active site base (Figure 2-20B). FadE32 is organized within an operon encoding two additional annotated FadEs, FadE31 (*Rv3562*) and FadE33 (*Rv3564*), each with a conserved base. Interestingly, the four operons encoding FadEs without a conserved catalytic base, adjacent to an intact FadE, are all up-regulated by cholesterol (Figure 2-21). We propose that FadE18, FadE27, and FadE32 will form a heteromeric complex with the neighboring FadE17, FadE26, and FadE31 or FadE33, respectively.

A

	Helix G		Helix J/K
VLCAD	ILNNGRFGM	VLCAD	FRIFEGTND
VLCAD2	ILNSGRFSM	VLCAD2	LLIFEGTNE
i3vD	GLDLERLVL	i3vD	YEIGAGTSE
SCAD	TLDMGRIGI	SCAD	TEIYEGTSE
i2VD	SLNEGRIGI	i2VD	GTIYEGASN
MCAD	AFDKTRPVV	MCAD	YQIYEGTSQ
iBD	GLNGGRINI	iBD	HQILEGSNE
LCAD	ELPQERLLI	LCAD	QPIYGGTNE
GD	CLNNARYGI	GD	VNTYEGTHD
FadE3	KGLETGRIQ	FadE3	-----
FadE16	VLPWFNLGN	FadE16	GSVMAPTAD
FadE18	VLDVARVLL	FadE18	TEALFGSSA
FadE27	---LSTLSR	FadE27	TEFALGGAT
FadE28	LAVMG----	FadE28	TRLLGGPSH
FadE32	AYEFGALAT	FadE32	LHSAWGTPQ

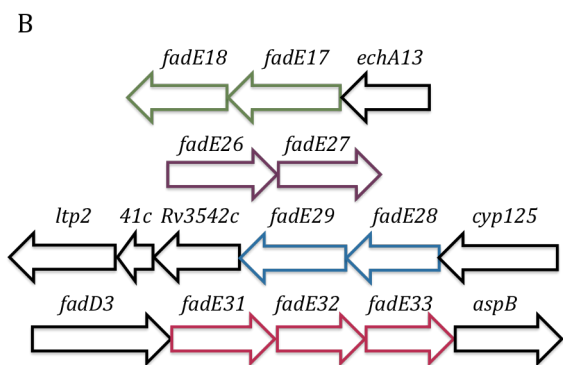


Figure 2-20. Bioinformatic analysis of the FadE proteome of *M. tuberculosis*. (A) The 36 annotated FadEs were aligned with characterized mammalian acyl-CoA dehydrogenases. The active site glutamate is highlighted in red and is located in either helix G or helix J/K. Those *M. tuberculosis* FadEs that lack a catalytic base are shown. (B) Presented is the operonic organization of those FadEs that lack a catalytic base and are adjacent to a gene encoding an intact FadE.

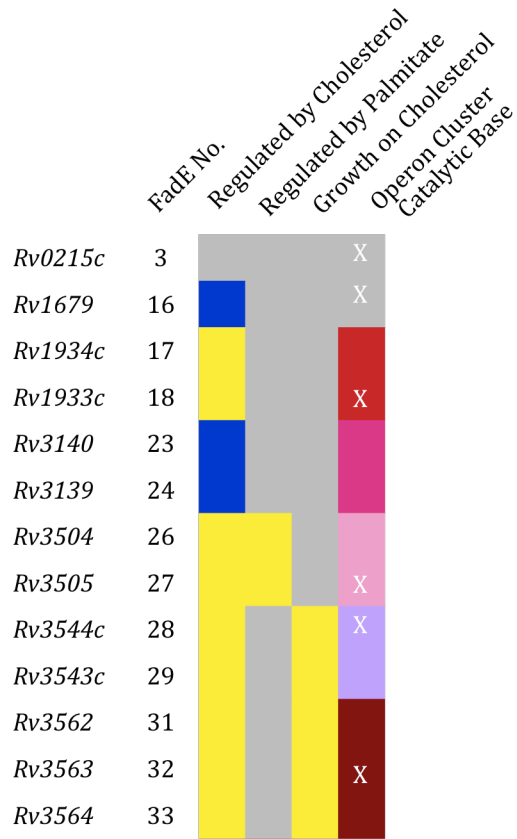


Figure 2-21. Regulation of *M. tuberculosis fadE* genes. The six *fadE* genes without a conserved catalytic base are shown along with other operonic *fadE* genes. The heat-map is clustered by regulation, phenotype, and genomic organization. cholesterol: induced-yellow, repressed-blue palmitate: induced-yellow, repressed-blue, yellow: required for growth on cholesterol. Genes marked with an “X” are lacking the canonical active-site glutamate that is the general base catalyst. FadEs in the same operon are of the same color.

Analysis of the genomic organization of the annotated *fadEs* identified two additional operonic *fadEs*, *fadE23* and *fadE24* (Figure 2-21). Both *fadE23* and *fadE24* encode for proteins with a conserved active site glutamate located in helix J/K. *fadE23* and *fadE24* are down regulated by cholesterol, are not required for growth of *M. tuberculosis* on cholesterol, and therefore are not predicted to be involved in cholesterol metabolism (Figure 2-21). It is not clear whether FadE23 and FadE24 form a heteromeric complex like FadE28 and FadE29 or two independent homotetrameric ACADs.

FadE3 and FadE16 lack the conserved active site glutamate necessary for ACAD activity, however, they are not positioned within the genome contiguous to an intact FadE

encoding gene (Figure 2-21). It is possible *fadE3* and *fadE16* are misannotated as ACADs or our sequence analysis against human ACAD representatives was not able to identify the catalytic base.

Protein-protein interactions are often important for proteins to carry out their biological function. In prokaryotes, functionally related genes are often organized into an operon and it is common for proteins encoded in the same operon to form protein complexes. However, our work is the first report of the identification and characterization of a heteromeric ACAD.

Our data indicate that despite the lack of dehydrogenase function of FadE28 it is important for ACAD activity through complex formation with FadE29. Typically, the FAD cofactor lies at the interface of two ACAD monomers that form a dimer. The lack of homodimeric interaction for FadE29 suggests that FadE28-FadE29 forms a dimer of heterodimers and we propose that FAD binds at the interface of this heterodimer. Further structural analysis is necessary to support this claim and understand the function of the heteromeric assembly. Crystallization trials are currently underway.

We predict that due to the absence of a conserved catalytic glutamate that FadE18, FadE27, and FadE32 are enzymatically inactive. The ACAD family has evolved to accommodate a variety of substrates by increasing the binding pocket size. Highly conserved residues, like the catalytic glutamate, are usually evolutionarily impervious to mutation. We propose that FadE28 has evolved, losing its catalytic function, to increase its binding pocket to accommodate polycyclic steroid substrates. Furthermore we predict that FadE18, FadE27, and FadE32 form analogous complexes that catalyze β -oxidation chemistry in the cholesterol metabolism pathway.

IV. Biochemistry of 2-enoyl-CoA hydratase Rv3541c-Rv3542c from *M. tuberculosis*

The side chain of cholesterol is proposed to be degraded by β -oxidation, the second step of which would be catalyzed by an enoyl-CoA hydratase. Sequence analysis of Rv3541c and Rv3542c, encoded by the *igr* operon suggest these genes encode for (*R*)-hydratases, typically involved in PHA biosynthesis in bacteria (145, 154). Our previous work demonstrated that Rv3541c and Rv3542c form a heteromeric complex (165). Metabolite profiling in Δ *igr* knockout *M. tuberculosis* suggest this complex catalyzes the hydration of 2-propenyl-CoA side chains of polycyclic substrates. Here we present full biophysical characterization of the Rv3541c-Rv3542c complex and demonstrate its enzymatic function as an enoyl-CoA hydratase capable of hydrating steroyl CoA thioester substrates.

Bioinformatics of Rv3541c and Rv3542c

Amino acid sequence identities amongst the (*R*)-hydratase family are typically low making it challenging to identify new members. Rv3541c and Rv3542c were analyzed by bioinformatic webservers that predict functional domains and three-dimensional structures of proteins based on sequence. Analysis of Rv3541c with Pfam (182) and InterPro (183) classify the protein as a member of the MaoC hydratase/dehydratase family, which includes several bacterial (*R*)-hydratases, and the thioesterase/thiol ester dehydrase-isomerase family, respectively. Analysis of Rv3542c by Pfam and InterPro provides similar annotations to Rv3541c, with the exception that only the N terminal domain of Rv3542c is assigned to the MaoC hydratase/dehydratase and thioesterase/thiol ester dehydrase-isomerase families. The C terminal domain is classified as a domain of unknown function.

Rv3541c is 129 amino acids in length with a molecular weight of 14 kDa. This is consistent with characterized bacterial (*R*)-hydratases. However, Rv3542c is more than twice the length of typical bacterial (*R*)-hydratases, with 311 amino acids and a molecular weight of 34 kDa. (*R*)-hydratases of this size have been identified in eukaryotic systems, and typically accept longer acyl-CoA substrates (Table 1-4). Rv3541c and Rv3542c were aligned with characterized bacterial (*R*)-hydratases in order to identify the conserved (*R*)-hydratase motif, which includes the active site residues required for hydratase activity.

The (*R*)-hydratase active site motif is also conserved in larger, eukaryotic enzymes, however, is it located in the C terminus. The (*R*)-hydratase motif is partially conserved in Rv3541c and Rv3542c (Table 2-5). Catalytic Asp and His residues are conserved in Rv3541c, while a Tyr replaces His in Rv3542c. The role of Asp and His in catalysis is to form a hydrogen bonding network with the catalytic water molecule. It is possible that Tyr-48 of Rv3542c could perform this role. The predicted active site residues of Rv3542c are located in the N terminus, unlike eukaryotic (*R*)-hydratases of similar length. Sequence alignment of Rv3542c with eukaryotic (*R*)-hydratases failed to identify an active site motif in the C terminus.

Table 2-5. The (*R*)-hydratase motif. Rv3541c and Rv3542c were aligned with characterized (*R*)-hydratases from *Aeromonas caviae* (PhaJ_{AC}), *Rhodospirillum rubrum* (PhaJ_{Rr}), and *Pseudomonas aeruginosa* (PhaJ_{Pa}). Catalytic residues are highlighted in red.

Protein	(<i>R</i>)-hydratase motif
Rv3541c	22 STALATRDFQDVH 34
Rv3542c	36 NWVEAIGDRNPIY 48
PhaJ _{AC}	25 FAALS_EDFNPLH 42
PhaJ _{Rr}	43 FAGLS_MDTNPAH 60
PhaJ _{Pa}	32 FAAVS_GDRNPVH 49

UniProt entry no. PhaJ_{AC} O32472, PhaJ_{Rr} Q9L9X2, PhaJ_{Pa} Q9LBK2.

Rv3541c and Rv3542c were analyzed with the structure prediction server Phyre2, which employ PSI-BLAST, secondary structure prediction, and profile-profile algorithms to predict three-dimensional structure (184). Rv3541c and the N-terminus of Rv3542c are predicted to have a hold dog fold, a common fold of CoA substrate utilizing enzymes, including (*R*)-hydratases. Over 20 structures with a hot dog fold were matched to Rv3541c and the N terminus of Rv3542c, with greater than 99% confidence, strongly suggesting Rv3541c and the N terminus of Rv3542c adopt a hot dog fold. Searching against the C terminus of Rv3542cs results in one hit, PDB 2GNN, with 100% confidence. PDB 2GNN is a eukaryotic protein from *Sulfolobus solfataricus* with possible involvement in lipid biosynthesis (185). Analysis of the C terminus of Rv3542c alone results in an additional hit,

with greater than 90% confidence, to a putative dehydratase from *Chloroflexus aurantiacus* (PDB 4E3E). It is possible that the C terminal domain of Rv3542c has a hot dog like fold as well, though our sequence analysis suggests the domain is less conserved than that of the N terminus.

Individual expression of Rv3541c and Rv3542c results in unstable, aggregation-prone protein

Initially Rv3541c and Rv3542c were expressed individually in *E. coli* and purified by IMAC. Under the tested conditions, isolated Rv3541c and Rv3542c precipitated after elution by imidazole from the His bind resin. The instability of Rv3541c and Rv3542c suggests they are not properly folded. Expression of Rv3541c and Rv3542c with folding chaperones dnaK-dnaJ-grpE-groES-groEL did not produce stable protein, further suggesting they are not in their native conformation. Analysis of the individual protein's activity and oligomeric state was not possible due to their instability.

FadE28-FadE29 forms an obligate $\alpha_2\beta_2$ tetramer

Our previous results demonstrated that Rv3541c-Rv3542c forms a stable, heteromeric complex. The protein complex was obtained by coexpressing the proteins as a single cistronic construct or on separate plasmids in *E. coli* with folding chaperones dnaK-dnaJ-grpE-groES-groEL. In both expression experiments only Rv3542c contained a N terminal hexahistidine tag for purification. SDS-PAGE and in-gel tryptic digestion and MALDI-TOF MS fingerprinting demonstrated that Rv3541c and Rv3542c were obtained by IMAC purification, confirming complex formation (Figure 2-22). The instability of individually isolated proteins and the stability of the Rv3541c-Rv3542c suggest this is the physiologically relevant complex.

A. Rv3541c

Residue No.	Sequence
1-50	MTVVGAVLPELKLYGDPTFIVSTALATRDFQDVHHRDKAVAQGSKDIFV
51-100	<u>NIL</u> TDTGLVQRYVTDWAGPSALIKSIGRLRGVPWYAYDTVTFSGEVTAVN
101-150	DGLITVKVVGRNTLGDHVTATVELSMRDS

B. Rv3542c

Residue No.	Sequence
1-50	MGSSHHHHHSSGLVPRGSHVTGVSDIQEAVAQIKAAGPSKPRLARDPVN
51-100	QPMINNWVEAIGDRNPIYVDGIVAPPAMIQVWTMMGLGGVRPKDDPLGPI
101-150	IKLFDDAGYIGVVATNCEQTYHRYLLPGEQGWFINQHIVWQVGDEDVAEM
151-200	<u>NWRILKFKPAGSPSSVPDDLDPDAMMRPSSSRDTAFFWDGVPVAVWQDKSV</u>
201-250	<u>PIN</u> YVSSGRGTVFSFVHHAPKVPGRTPFVIALVEEGRMLGELRG
251-261	DWSLYAWEPDE

Figure 2-22. Tryptic peptides from Rv3541c and Rv3542c. Tryptic peptides, identified by MALDI-TOF mass spectrometry, are mapped onto the sequence of (A) Rv3541c and (B) Rv3542c. The sequences corresponding to the identified peptides are underlined.

The complex was analyzed by analytical size exclusion chromatography and analytical ultracentrifugation sedimentation equilibrium experiments to determine the molecular weight. The elution time of Rv3541c-Rv3542c from an analytical size exclusion column, compared to protein standards, analyzed under the same conditions, resulted in an approximate molecular weight of 100 kDa, confirming the formation of a higher order oligomer (Figure 2-23). Analytical ultracentrifugation data acquired at three protein concentrations fit to an ideal model and resulted in a complex molecular weight of 101 kDa (Figure 2-24). The calculated molecular weight of possible complex stoichiometries of

Rv3541c (14.0 kDa) and Rv3542c (36.1 kDa) indicates Rv3541c-Rv3542c forms an $\alpha_2\beta_2$ tetramer with a calculated molecular weight of 100.2 kDa. This oligomerization state was confirmed by analysis of the complex by LC/UV/MS. The absorption profile at 280 nm for the eluted peaks corresponding to Rv3541c and Rv3542c were integrated and the extinction coefficient for each protein (Table 3-2) was used to determine the stoichiometry of Rv3541c to Rv3542c. A 1 to 1 stoichiometry of Rv3541c to Rv3542c confirmed the $\alpha_2\beta_2$ oligomer assembly.

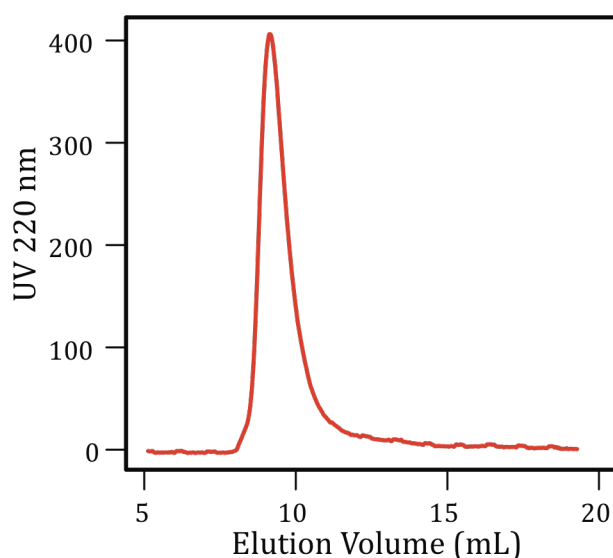


Figure 2-23. Analytical size exclusion chromatography of Rv3541c-Rv3542c. Rv3541c-Rv3542c was analyzed by analytical size exclusion chromatography on a Superdex 75 column equilibrated in Buffer J. Several standard proteins were analyzed under the same conditions to generate a standard curve to estimate molecular weights of analyzed proteins. The approximated molecular weight of Rv3541c-Rv3542c is 100 kDa.

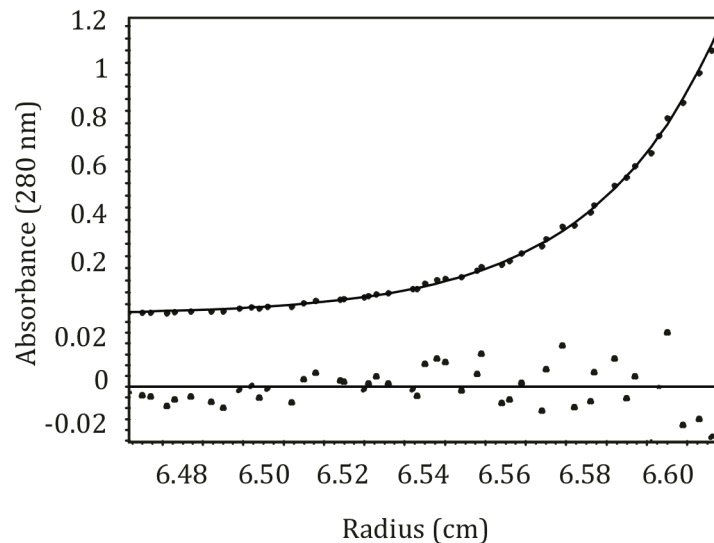


Figure 2-24. Analytical ultracentrifugation sedimentation equilibrium of Rv3541c-Rv3542c. Rv3541c-Rv3542c (5.9 μM , 3.8 μM , and 2.4 μM) was centrifuged at speeds of 10k, 15k, and 20k rpm at 4 $^{\circ}\text{C}$. Data were collected as 280 nm and fit globally resulting in a molecular weight of 101 kDa. A representative fit is shown.

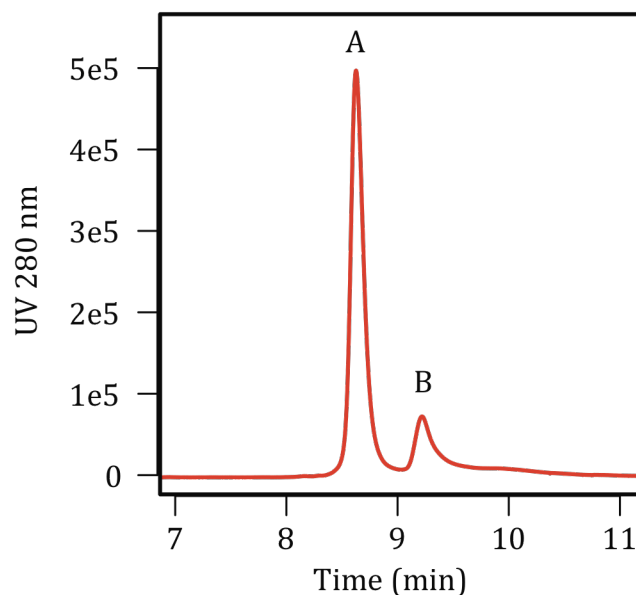


Figure 2-25. Reverse phase LC/UV chromatogram of Rv3541c-Rv3542c, monitoring at 280 nm. Peak A and peak B were identified as Rv3542c and Rv3541c, respectively, by ESI MS. The UV profiles of the peaks were integrated and relative concentrations were determined from the calculated extinctions coefficients, $\epsilon_{280}(\text{Rv3542c}) = 59,930 \text{ M}^{-1} \text{ cm}^{-1}$ and $\epsilon_{280}(\text{Rv3541c}) = 16,960 \text{ M}^{-1} \text{ cm}^{-1}$.

Rv3541c-Rv3542c catalyzes the hydration of polycyclic acyl-CoA substrates

FadE28-FadE29 is capable of catalyzing the dehydrogenation of polycyclic CoA thioesters substrates, including hexahydroindanone **2** and pregnenone **3** (Table 2-4). We previously proposed that Rv3541c-Rv3542c is required for the metabolism of the partially metabolized 2'-propanoate side chain of cholesterol and catalyzes the subsequent hydration step after dehydrogenation by FadE28-FadE29. To test this hypothesis we prepared the hypothetical substrates **4** and **5** *in situ* using FadE28-FadE29 (Figure 2-26). Reactions were initiated by the addition of pure Rv3541c-Rv3542c and monitored by MALDI-TOF MS. Product was detected in initial time points taken at 5 and 2 min, for **4** and **5** respectively, post Rv3541c-Rv3542c addition. Negative controls without enzyme were conducted and no activity was detected. Characterization of product formation by ^1H NMR spectroscopy is required for confirmation that the tertiary alcohol is formed.

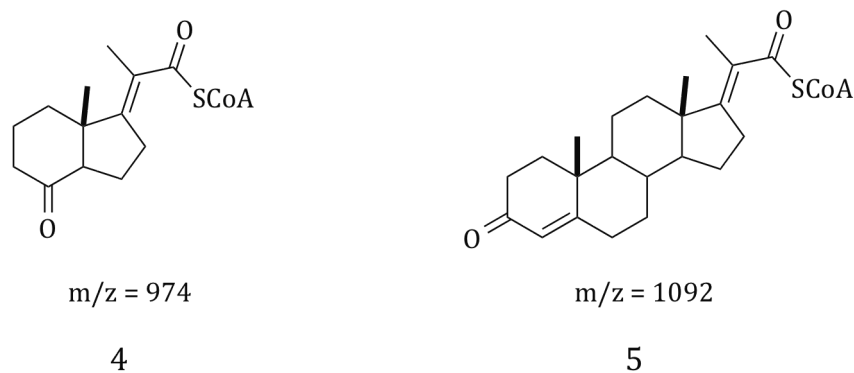


Figure 2-26. Substrates of Rv3541c-Rv3542c. Unsaturation was introduced into compounds **3** and **4** by FadE28-FadE29 *in situ*, starting with compounds **2** and **3**. The position of the alkene is presumed.

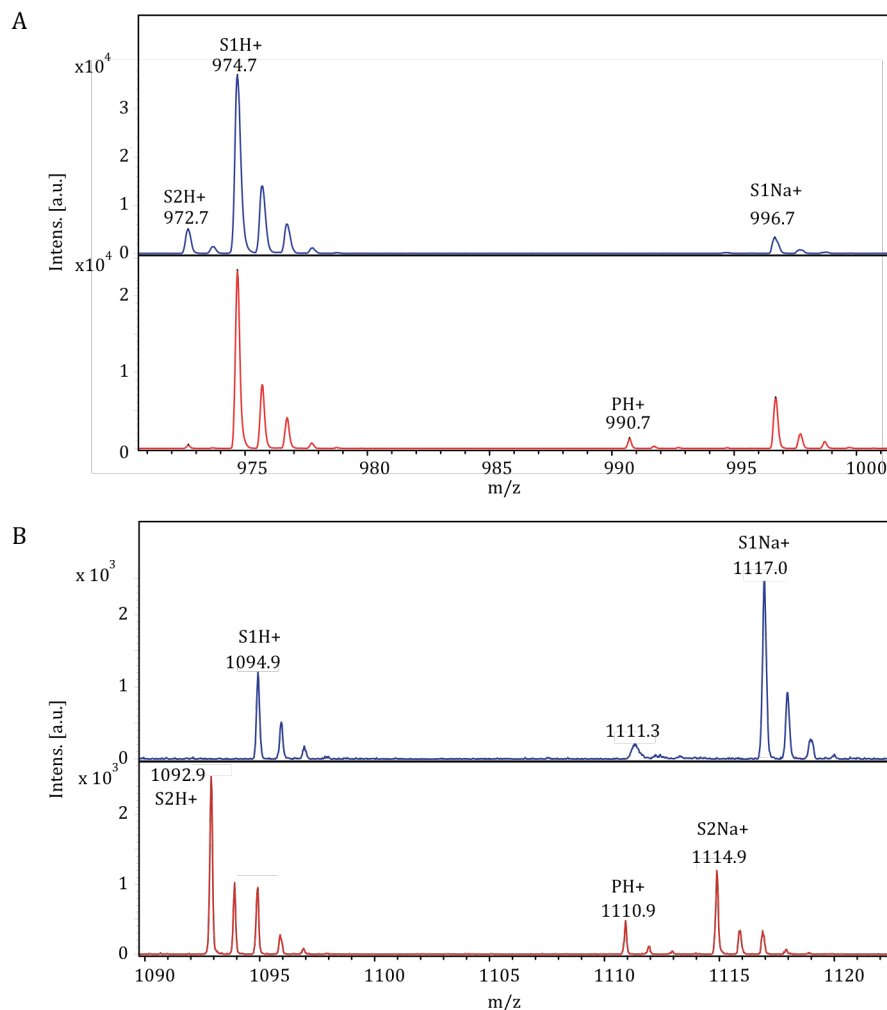


Figure 2-27. Detection of Rv3541c-Rv3542 assay products by MALDI-TOF MS. Rv3541c-Rv3542c was assayed with **4** (A) and **5** (B) and products were detected by MALDI-TOF MS. In **blue** are control spectra without Rv3541c-Rv3542c and in **red** is **5** and 2 min, for A and B respectively, after the addition of Rv3541c-Rv3542c. The m/z of hydrated products for **4** and **5** are 991 and 1111, respectively. M/z 1111 observed in the control for **5** is not predicted to be hydrated product because it was observed before generating the unsaturated substrate by FadE28-FadE29. S1 = substrate for FadE reaction, S2 = substrate for Rv3541c-Rv3542c reaction, P = product of Rv3541c-Rv3542c assay.

Functional role of the heteromeric enoyl-CoA hydratase

We propose the unique $\alpha_2\beta_2$ assembly of Rv3541c-Rv3542c has an important functional role in substrate binding, similar to eukaryotic (*R*)-hydratases. Crystallographic structures of eukaryotic (*R*)-hydratases demonstrate they are composed of an α and β

domain formed from a single polypeptide chain (155). The monomeric fold is structurally similar to the homodimeric fold of bacterial (*R*)-hydratases. One major difference however is that the $\alpha + \beta$ domain structure of eukaryotic (*R*)-hydratase contains one active site located in the C terminal domain, while the prokaryotic (*R*)-hydratase homodimer has two active sites per dimer (Figure 1-6). The lack of an active site in the N terminal domain of the eukaryotic enzyme allows it to accommodate long chain acyl-CoA thioesters.

Two heterodimeric (*R*)-hydratases have been identified in *M. tuberculosis*, encoded by three adjacent, operonic genes (186). HadB (*Rv0636*) forms a heterodimer with HadA (*Rv0635*) and also a heterodimer with HadC (*Rv0637*). HadB contains a conserved (*R*)-hydratase motif while HadA and HadC both lack catalytic Asp and His residues. These heterodimeric complexes are specific for long chain fatty acids and activities were highest for C12, C16 and C20 fatty acyl-CoAs. The specificity for long chain fatty acids and the heterodimeric assembly with one active site, suggests this unique architecture is also important for accommodating large substrates. No crystallographic evidence supports this hypothesis.

Rv3541c is predicted to fold as a single domain with one active site. Rv3542c is predicted to fold as two separate hot-dog like domains with only one active site in the N terminal domain. The heterotetrameric assembly would be composed of six domains with four active sites making it structurally unique from known (*R*)-hydratases. We predict this unique architecture is important for Rv3541c-Rv3542c to bind large, bulky steroyl CoA thioester substrates. Further mutagenesis studies and structural analysis are necessary to support this hypothesis and understand the function of the heterotetrameric assembly. Crystallization trials are currently underway in the lab and a preliminary structure has been obtained. A detailed analysis of the structure is ongoing.

Chapter 3

Experimental Methods

I.	Materials	87
II.	General Methods	87
III.	Instrumentation	88
IV.	Buffers	88
V.	Expression plasmid construction and plasmid table	89
VI.	Expression and protein purification	90
VII.	In-gel tryptic digestion and MALDI-TOF mass fingerprinting	94
VIII.	Azasteroids	94
IX.	3 β -HSD assay	94
X.	Isotopic labeling of LDL-cholesterol.	95
XI.	Metabolic labeling of <i>M. tuberculosis</i>	96
XII.	Growth of <i>M. tuberculosis</i> in the presence of free cholesterol	96
XIII.	Biochemical analysis of LDL-derived [¹⁴ C]-lipids	97
XIV.	Biochemical extraction and analysis of cholesterol-derived bacterial lipid metabolites	97
XV.	Purification and characterization of cholesterol-derived metabolite 1	98

XVI.	FadE28-FadE29 substrate synthesis	98
XVII.	Dehydrogenase assay	100
XVIII.	Hydratase assay	100
XIX.	Identification and quantification of flavin cofactor	100
XX.	Analytical size exclusion chromatography	100
XXI.	Analytical ultracentrifugation sedimentation equilibrium	101
XXII.	LC/UV/MS of intact protein complexes	101

I. Materials

Ferricinium hexafluorophosphate, ergocalciferol, coenzyme A, propionyl-CoA, and (RS)-[2-¹³C]-mevalonolactone were purchased from Sigma-Aldrich (St. Louis, MO). 2,5-dihydroxybenzoic acid and 4,22-stigmastadien-3-one were purchased from MP Biomedicals (Solon, Ohio). (RS)-[2-¹⁴C]-mevalonolactone was purchased from Perkin Elmer (Waltham, MA). Isopropyl β-D-1-thiogalactopyranoside was from Denville Scientific (Metuchen, NJ). Tryptone, HEPES, TRIS, and ampicillin were purchased from Fisher Scientific (Pittsburgh, PA). Vitamin D2, L-arabinose, chloramphenicol, and sodium chlorite were purchased from Acros Organics (New Jersey). Tetracycline is from US Biochemical Corp (Cleveland, OH) and kanamycin is from IBI Scientific (Peosta, IA). Yeast extract was purchased from Research Products International Co. (Mount Prospect, IL). iProof DNA polymerase was from Bio-Rad (Hercules, CA). Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, and protein ladder were from New England Biolabs (Beverly, MA). Chaperone plasmid pG-KJE8 was purchased from Takara Bio Inc. (Otsu, Shiga Japan). HisTrap FF columns and Superdex 200 HiLoad 16/60 and 10/300 GL columns were from GE Healthcare Biosciences Corp. (Piscataway, NJ). A reverse phase XBridge BEH 300 C4 3.5 μm column (2.1 x 100 mm) was purchased from Waters (Milford, MA).

II. General Methods

DNA sequencing with an Applied Biosystems 3730 DNA analyzer was performed at Stony Brook University Sequencing Facility to verify the coding sequence of the expression plasmids in Table 3-1. Total genomic DNA of *M. tuberculosis* H37Rv was obtained from the TB Research Materials Facility at Colorado State University (Fort Collins, CO) (NIAD N01-AI40091). Oligonucleotides were from Eurofins mwg Operon (Huntsville, AL). *M. tuberculosis* H37Rv Δ *igr*, and complemented Δ *igr* strains were obtained from David R. Sherman and were constructed as described in (63). *M. tuberculosis* was grown at 37 °C in Middlebrook 7H9 liquid media, supplemented with 10% oleate-albumin-dextrose-NaCl-catalase (OADC). 2xYT is composed of 16 g tryptone, 10 g yeast extract and 5 g NaCl per liter.

III. Instrumentation

Matrix-assisted laser desorption ionization (MALDI) MS were acquired on a Bruker Autoflex II TOF/TOF spectrometer at the Institute of Chemical Biology and Drug Discovery, Stony Brook University. High-resolution mass spectrometry was performed on an LTQ-Orbitrap from Thermo Scientific with an electrospray source at the Proteomics Center at Stony Brook University. Metabolite profiling data was acquired in positive ion mode with an ionization voltage of 1.8 kV, capillary voltage of 43 V, and tube lens of 150 V. LC/UV/MS analysis of azasteroids and protein samples was performed on a Waters ACQUITY Ultra Performance LC system equipped with a photodiode array detector and a single quadrupole detector. Protein purification and analytical size exclusion chromatography were performed on a Pharmacia ÄKta explorer. NMR spectra were acquired on Bruker 800 or 900 MHz microcryoprobe or cryoprobe NMR at the New York Structural Biology Center. UV-visible spectra and kinetic assays were performed on a Shimadzu UV-2550 UV-visible spectrophotometer. Analytical ultracentrifugation sedimentation equilibrium experiments were conducted with a Beckman Coulter Optima XL-A ultracentrifuge.

IV. Buffers

Buffer A	50 mM triethanolamine hydrochloride buffer, pH 8.5
Buffer B	20 mM Tris-HCl buffer, pH 8.0, supplemented with 300 mM NaCl and 10 mM imidazole
Buffer C	20 mM Tris-HCl buffer, pH 8.0, supplemented with 300 mM NaCl and 250 mM imidazole
Buffer D	50 mM Tris-HCl buffer, pH 8.0
Buffer E	50 mM Tris-HCl buffer, pH 8.0, supplemented with 200 mM NaCl
Buffer F	20 mM Tris-HCl buffer, pH 8.0, supplemented with 300 mM NaCl, 0.5 M imidazole
Buffer G	50 mM Tris-HCl buffer, pH 8.0, 200 mM NaCl and 1 mM TCEP
Buffer H	100 mM TAPS buffer, pH 8.5, supplemented with 150 mM NaCl and 30 mM MgCl ₂
Buffer I	100 mM HEPES buffer, pH 7.4

Buffer J 20 mM sodium phosphate buffer, pH 7.5, supplemented with 200 mM NaCl

Buffer K 5 mM sodium phosphate buffer, pH 7.5

V. Expression plasmid construction and plasmid table

The desired genes were amplified from *M. tuberculosis* H37Rv total genomic DNA by PCR using forward and reverse primers. The PCR product was digested with the appropriate restriction endonuclease and ligated into a similarly digested vector. DNA sequencing of the plasmids confirmed that the sequence was correct and that no mutations were introduced during the cloning procedures.

Table 3-1. Plasmids used in this work.

Construct name	Plasmid	Gene	Fusion Tag	Restriction sites used	Antibiotic marker
^a <i>pET28b-1106c</i>	pET28b	<i>Rv1106c</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtbigr-6</i>	pET28b	<i>Rv3545c-Rv3540c</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtbigr-5</i>	pET28b	<i>Rv3544c-Rv3540c</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtbigr-3</i>	pET28b	<i>Rv3542c-Rv3540c</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtbigr-1</i>	pET28b	<i>Rv3540c</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtb28N</i>	pET28b	<i>fadE28</i>	N-terminal His ₆	NdeI/NotI	kan
<i>pMtb29N</i>	pET28b	<i>fadE29</i>	N-terminal His ₆	NdeI/XhoI	kan
<i>pMtb28</i>	pET20b	<i>fadE28</i>	---	NdeI /NotI	amp
<i>pMtb29Ms</i>	pSD31	<i>fadE29</i>	N-terminal His ₆	EcoRV	hygro
<i>pMtb2829</i>	pET28b	<i>fadE28-fadE29</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtb2829_{E240Q}</i>	pET28b	<i>fadE28-fadE29</i>	N-terminal His ₆	NdeI/HindIII	kan
<i>pMtb41cN</i>	pET28b	<i>Rv3541c</i>	N-terminal His ₆	NdeI/XhoI	kan
<i>pMtb42cN</i>	pET28b	<i>Rv3542c</i>	N-terminal His ₆	NdeI/NotI	kan
<i>pMtb41c</i>	pET20b	<i>Rv3541c</i>	---	NdeI/XhoI	amp
^b <i>pG-KJE8</i>					cam

^a (73)

^b Takara Bio Inc.

VI. Expression and protein purification

M. tuberculosis 3 β -HSD

Expression and purification of 3 β -HSD was carried out as previously described (73). Construct *pET28b-1106c* was transformed into competent BL21(DE3) *E. coli* cells. Colonies were grown overnight in 10 mL LB medium supplemented with 30 μ g mL⁻¹ kanamycin. The starter culture was used to inoculate 2 \times YT media and the cells were grown to *OD*₆₀₀ ~ 0.6 – 0.8. Expression of the gene was induced by isopropyl β -D-thiogalactoside (IPTG) (0.4 mM) and was allowed to proceed for 20 h at 22 °C. Cells were harvested by centrifugation

at 5000 rpm for 30 min at 4 °C. The pellet (5g) was suspended in Buffer A and lysed by French press, 3 times at 1000 psig. The lysed cells were centrifuged at 125,000 × g for 1 h at 4 °C. The supernatant was collected and subjected to ammonium sulfate precipitation (5-15% w/v). After incubation at 4 °C, the precipitated protein was collected by centrifugation at 5,000 rpm for 20 min at 4 °C. The pellet was suspended in 5 mL of Buffer A and dialyzed against the same buffer with MWCO 6,000-8,000 molecular porous membrane tubing. The sample was loaded onto a diethylaminoethyl cellulose column DE-52 (30 mm x 25 cm), equilibrated with Buffer A, and the column was developed using a 0 – 1 M NaCl linear gradient (300 mL of Buffer A and 300 mL Buffer A with 1 M NaCl). Fractions with activity were combined, desalted, and concentrated by ultrafiltration with 30 kDa membrane, to a volume of 10 mL. The protein was then loaded onto a Q-sepharose column (10 mL bed volume) equilibrated in Buffer A. The column was developed using Buffer A and pure fractions with dehydrogenase activity were combined and concentrated.

igr operon

The *igr* operon (*Rv3545c-Rv3540c*) was cloned from genomic DNA into pET28b (Novagen) at *NdeI* and *HindIII* restriction sites to give construct pMtbigr-6. Constructs pMtbigr-5 and pMtbigr-3 were prepared by deletion of *Rv3545c* or deletion of *Rv3545c/Rv3544c/Rv3543c* respectively by PCR. Construct pMtbigr-1 was prepared by cloning *Rv3540c* from genomic DNA into pET28b at *NdeI* and *XhoI* restriction sites. Each construct introduced an N-terminal hexahistidine tag. Constructs were introduced into BL21(DE3) *E. coli* and single colonies were selected on LB plates supplemented with 30 µg/mL kanamycin and cultured in 2 × YT media at 37 °C. Expression was induced at $OD_{600} \sim 0.6 - 0.8$ by the addition of 1 mM IPTG and cells were grown for 18 to 20 h at 25 °C. Harvested cells were suspended in Buffer B. Cells were lysed by French press and cellular debris was removed by centrifugation at 125,000 × g for 1 h. Proteins were purified by immobilized metal affinity chromatography (IMAC) using Hisbind resin. Resin was thoroughly washed with Buffer B and protein was eluted with Buffer C. Eluted protein was analyzed by SDS-PAGE and observed protein band identities were confirmed by in-gel tryptic digestion and MALDI-TOF mass fingerprinting.

M. tuberculosis FaeE28 and FaeE29 in Escherichia coli

Construct pMtb28N or pMtb29N was transformed into BL21(DE3) *E. coli*, single colonies were selected on LB plates containing 30 µg/mL kanamycin, and cultured in 2 x YT media at 37 °C. Expression was induced at $OD_{600} \sim 0.6 - 0.8$ by the addition of 50 µM to 1 mM IPTG, and cells were grown 20 h at 16-25 °C. Cells were lysed by French press or sonication in Buffer B, and cellular debris was removed by centrifugation at $125,000 \times g$ for 1 h. Proteins were purified by IMAC using Hisbind resin following the manufacturer's protocol using Buffers B and C and analyzed by reducing SDS-PAGE and UV-visible spectroscopy. FadE29 was further purified by anion exchange chromatography on a Pharmacia MonoQ column (1 mL) equilibrated in Buffer D. Protein was eluted at a flow rate of 0.5 mL/min with a linear gradient from 100% Buffer D to 100% Buffer D supplemented with 1 M NaCl. After injection, the column was washed with five column volumes (CV) of Buffer D, then changed to 80% Buffer D over 20 CV. Next, the gradient was changed from 80% to 25% Buffer D over 10 CV, then to 0% Buffer D over 5 CV. Then size exclusion chromatography on a Superdex 200 HiLoad 16/60 column was performed with Buffer E at a flow rate of 0.5 mL/min.

M. tuberculosis FadE29 in M. smegmatis

Construct pMtb29Ms was electroporated into *M. smegmatis* mc²155 and single colonies were selected on 7H10 plates supplemented with 100 µg/mL ampicillin, 10 µg/mL cycloheximide, and 50 µg/mL hygromycin and grown in 7H9 media supplemented with 0.2% glycerol. After 2 days, expression was induced with 0.2% L-arabinose and cultures were grown for an additional 24 h. Cells were lysed by French press in Buffer B, and cellular debris was removed by centrifugation at $125,000 \times g$ for 1 h. Protein was purified as detailed above for FadE29.

M. tuberculosis Rv3541c and Rv3542c

Construct pMtb41cN or pMtb42cN was transformed into BL21(DE3) *E. coli*, single colonies were selected on LB plates containing 30 µg/mL kanamycin and cultured in 2xYT media at 37 °C. Expression was induced at $OD_{600} \sim 0.6 - 0.8$ by the addition of 50 µM to 1 mM IPTG, and cells were grown 20 h at 16-25 °C. Cells were lysed by French press or sonication in Buffer B, and cellular debris was removed by centrifugation at $125,000 \times g$ for 1 h. Proteins were purified by IMAC using Hisbind resin following the manufacturer's protocol using Buffers B and C and analyzed by reducing SDS-PAGE.

Complex expression

FadE28-FadE29

Constructs pMtb29N and pMtb28 were co-transformed into BL21(DE3) *E. coli*::pG-KJE8 (chaperone plasmid, Takara). Single colonies were selected on LB plates containing the appropriate antibiotics (100 µg/mL ampicillin for pET20b, 30 µg/mL kanamycin for pET28b, and 20 µg/mL chloramphenicol for pG-KJE8) and cultured in 2 x YT media at 37 °C. Chaperone expression was induced upon inoculation with 2 mg/mL L-arabinose and 5 ng/mL tetracycline. FadE expression was induced at $OD_{600} \sim 0.6 - 0.8$ with the addition of 1 mM IPTG, and cells were grown 20 h at 25 °C. Similarly, FadE28-FadE29 complex and FadE28-FadE29E240Q mutant protein were obtained with construct pMtb2829 or pMtb2829E240Q respectively, expressed in BL21(DE3) *E. coli*::pG-KJE8. Cells were lysed by French press or sonication in Buffer B, and cellular debris was removed by centrifugation at $125,000 \times g$ for 1 h. Proteins were purified by IMAC using HisTrap FF columns (GE Healthcare), eluting with a linear gradient from 100% Buffer B to 100% Buffer F over 5 CV. Eluted protein was further purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with Buffer G. Samples were analyzed by reducing SDS-PAGE and observed protein band identities were confirmed by in-gel tryptic digestion and MALDI-TOF mass fingerprinting.

Rv3541c-Rv3542c

Rv3541c-Rv3542c complex was obtained using construct pMtbigr3. Expression was carried out as described above for pMtbigr3, above. For analytical studies, protein was purified in Buffer B by immobilized metal affinity chromatography using HisTrap FF columns, eluting with a linear gradient from 100% Buffer B to 100% Buffer F over 5 CV. Eluted protein was further purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with Buffer E.

Rv3541c-Rv3542c complex can also be obtained by coexpression with constructs pMtb41c and pMtb42cN. Constructs were co-transformed into BL21(DE3) *E. coli*::pG-KJE8. Single colonies were selected on LB plates containing the appropriate antibiotics (100 µg/mL ampicillin for pET20b, 30 µg/mL kanamycin for pET28b, and 20 µg/mL chloramphenicol for pG-KJE8) and cultured in 2 x YT media at 37 °C. Chaperone expression was induced upon inoculation with 2 mg/mL L-arabinose and 5 ng/mL tetracycline.

Rv3541c and Rv3542c expression was induced at $OD_{600} \sim 0.6 - 0.8$ with the addition of 1 mM IPTG, and cells were grown 20 h at 25 °C. Protein was purified as detailed above for Rv3541c-Rv3542c complex.

VII. In-gel tryptic digestion and MALDI-TOF mass fingerprinting

All protein identities were confirmed by in-gel tryptic digestion and MALDI-TOF mass fingerprinting. Briefly, excised bands were washed with 200 μ L of CH₃CN/H₂O followed by reduction with 100 μ L of 45 mM DTT in 100 mM ammonium bicarbonate at 56 °C for 45 min. Gel pieces were then covered with 100 μ L of 55 mM iodoacetamide in 100 mM ammonium bicarbonate and incubated in the dark for 30 min. The samples were dried and digested with trypsin according to the manufacturer's protocol in 50 mM ammonium bicarbonate overnight at 37 °C. Peptides were then extracted with 60% CH₃CN/0.1% TFA and dried. The MALDI sample was prepared using a C18-Zip-tip (Millipore). The sample was loaded on to the zip-tip equilibrated in 0.1% TFA, washed with 3 x 10 μ L of 0.1% TFA, and eluted with 10 μ L of 0.1%TFA/49.9% H₂O/50% ACN. A saturated solution of matrix 3,5-dimethoxy-4-hydroxy-cinnamic acid was prepared in 29.9% EtOH/70% H₂O/0.1% TFA. MALDI-TOF mass spectra were acquired in positive ion mode and data were analyzed using Flex-Analysis software and MS-Bridge.

VIII. Azasteroids

6-azasteroids were provided by GlaxoSmithKline (Research Triangle Park, NC). Frye and coworkers describe the synthesis and purity confirmation by elemental analysis of the 6-azasteroids (172, 173, 187). 4-azasteroids were purchased from Sigma-Aldrich. The purity (>98%) and identity of azasteroids were confirmed by LC/UV/MS (Waters UPLC/diode array/SQD) for use in these assays. The LC was performed with a 1.7 μ M C18 (2.1 x 100 mm) column and the gradient was from 100% H₂O to 100% methanol over 10 minutes at a flow rate of 0.5 mL/min.

IX. 3 β -HSD assay

3 β -HSD was assayed with substrates DHEA (120 μ M) and NAD⁺ (400 μ M), with or without inhibitor, in Buffer H. Reactions were initiated by the addition of 3 β -HSD (125 nM). Inhibitor stock solutions were prepared in DMSO. The final DMSO concentration was held fixed at 2%, which does not affect enzyme activity (73). The formation of NADH was

monitored at 340 nm for the first 150 s of reaction at 30 °C. Assays were performed in duplicate and IC₅₀ values were determined from 8 concentrations by fitting to Equation 3-1. Mechanism of inhibition was determined by fitting initial velocity data to Equation 3-2 using Grafit 4.0 (Erithacus software, Sussex UK).

Equation 3-1

$$v = V_m / (1 + [I] / IC_{50})$$

Equation 3-2

$$v = V_m [S] / \{ K_m (1 + [I] / K_{ic}) + [S] (1 + [I] / K_{iu}) \}$$

where v is the initial velocity, V_m is the maximum velocity, S is the varied substrate, K_M is the Michaelis-Menten constant for the varied substrate, and K_{ic} is the competitive inhibition constant and K_{iu} is the uncompetitive inhibition constant.

X. Isotopic labeling and characterization of LDL-cholesterol.*

HepG2 human liver cells were grown to 80% confluence in HepG2 growth media (DMEM, 10% fetal calf serum, 20 mM L-glutamine, 100 IU/mL penicillin, 10 µg/mL streptomycin, and 10 mM HEPES). Media was then replaced with HepG2 growth media supplemented with 5 µM mevastatin and 0.8 µg/mL (*RS*)-mevalonolactone (natural abundance), or 0.8 µg/mL (*RS*)-[2-¹³C]-mevalonolactone or 6.25 µCi (*RS*)-[2-¹⁴C]-mevalonolactone. Cells were grown for 4-5 days. Culture supernatants were then harvested and sterilized by filtration. Culture supernatant-derived LDL particles were concentrated and unincorporated mevalonolactone was removed by ultrafiltration through a 100 kDa MWCO filter. Removal of unincorporated mevalonolactone was performed by concentrating the supernatants 10-fold followed by two 100 mL washes with sterile PBS. The filter flow-through was monitored by scintillation counting and the LDL was deemed free of ¹⁴C-mevalonolactone contamination when radioactive counts in the filter flow-through were no higher than background. LDL-[¹³C]-cholesterol was isolated in the same manner as the LDL-[¹⁴C]-cholesterol preparation and the LDL-[¹³C]-cholesterol sample underwent the ultrafiltration procedure to remove unincorporated [¹³C]-mevalonolactone.

Isolated LDL-[¹³C]-cholesterol was analyzed by MALDI-TOF MS in reflectron, positive ion mode with 2,5-dihydroxybenzoic acid matrix prepared in 0.1% TFA/50%

* These procedures were performed by Brain C. VanderVen, Cornell University.

CH₃CN. A ¹³C-DEPT135 NMR spectrum was acquired on a 800 MHz microcryoprobe in CDCl₃ to confirm the positions of ¹³C labeling.

XI. Metabolic labeling of *M. tuberculosis**

LDL-[¹⁴C]-cholesterol

The *M. tuberculosis* WT, *Δigr*, and complemented *Δigr* mutant cultures were grown in 7H9 OADC for six days and the bacteria were resuspended in 7H9 OADC medium supplemented with LDL-[¹⁴C]-cholesterol (5,000 cpm/mL). After 2 weeks of growth in the presence of LDL-[¹⁴C]-cholesterol the bacterial cultures were centrifuged at 3,000 rpm for 20 min to yield a cell pellet and culture supernatant. The supernatants were harvested by filtration through 0.22 μm filter unit and the resulting cell pellets and culture supernatants were used for lipid metabolite isolation.

LDL-[¹³C]-cholesterol

The *M. tuberculosis* WT, *Δigr*, and complemented *Δigr* mutant cultures were grown in 7H9 OADC for six days and the cell pellet was resuspended in 7H9 OADC medium supplemented containing LDL-[¹³C]-cholesterol (50-60 μg/mL) of culture. Total cholesterol in the LDL-[¹³C]-cholesterol preparations was determined using the Amplex® Red cholesterol determination assay (Invitrogen, Carlsbad, CA). After 2 weeks of growth in the presence of LDL-[¹³C]-cholesterol, the bacterial cultures were centrifuged at 3,000 rpm for 20 min to yield a cell pellet and culture supernatant. The supernatants were harvested by filtration through a 0.22 μm filter unit and the resulting cell pellets and culture supernatants were used for lipid metabolite isolation.

XII. Growth of *M. tuberculosis* in the presence of free cholesterol†

The *M. tuberculosis* WT, *Δigr*, and complemented *Δigr* mutant cultures were grown in 7H9 OADC for six days and the cell pellet was resuspended in 7H9 OADC medium supplemented with 0.1 mM cholesterol added in a 1:1 tyloxapol:EtOH (v/v) solution as described (37). After 2 weeks of growth in the presence of free cholesterol, the bacterial cultures were centrifuged at 3,000 rpm for 20 min to yield a cell pellet and culture

* These procedures were performed by Brain C. VanderVen, Cornell University.

† These procedures were performed by Brain C. VanderVen, Cornell University.

supernatant. The supernatants were harvested by filtration through a 0.22 μm filter unit and the resulting cell pellets and culture supernatants were used for lipid metabolite isolation.

XIII. Biochemical extraction and analysis of LDL-derived [^{14}C]-lipids*

Total lipids were extracted by the Bligh-Dyer method from the aqueous-soluble LDL particles (188). Briefly, the aqueous lipid solution was phase partitioned with 2:1 CHCl_3 :MeOH (v/v) and the organic solvent extractable material was dried by evaporation under a N_2 stream. The dried lipids were resuspended in 25:24:4 CHCl_3 :MeOH:H $_2\text{O}$ (v/v/v), the phases partitioned, the resulting organic solvent layer saved to a new tube, and dried by evaporation under a N_2 stream.

To analyze the apolar LDL-lipids by TLC, the LDL lipid extracts were resuspended in CHCl_3 and radioactivity was determined by scintillation counting. 2500 cpm of the apolar LDL-derived lipids was resolved by TLC on pre-coated F_{254} aluminum backed silica plates (Sigma-Aldrich, Saint Louis, MO) in toluene:acetone (99:1, v/v). The TLC plates were visualized via molybdophosphoric acid staining and charring as described in (189). Radiolabeled lipids were visualized by autoradiography of the TLC plates using phosphoimaging.

XIV. Biochemical extraction and LC/MS analysis of cholesterol-derived bacterial lipid metabolites

The bacterial cell pellets were washed once in distilled H $_2\text{O}$ and the pellet was extracted twice with 75 mL EtOAc for 24 h. The organic layers were pooled and dried by evaporation under a N_2 stream. The cell-free culture supernatants were extracted twice with 75 mL EtOAc for 24 h. The organic layers were pooled and dried by evaporation under a N_2 stream.

Extracts dissolved in MeOH were analyzed by microcapillary liquid chromatography-tandem mass spectrometry with a Dionex 3000 HPLC and a Thermo LTQ Orbitrap mass spectrometer equipped with a custom nanoLC electrospray ionization source. Analytes were separated on a column packed with 10 cm of 5 μm Magic C18 material (Agilent, Santa Clara, CA). A flow rate of 300 nL/min was used with a gradient from 0.1% formic acid/H $_2\text{O}$ (Solvent A) to 0.1% formic acid/98% CH_3CN (Solvent B). After analyte loading, the gradient was held constant at 100% Solvent A for 5 min followed by a

30 min gradient from to 40% Solvent B. Then, the gradient was switched from 40% to 80% Solvent B over 5 min and held constant for 3 min. Finally, the gradient was changed from 80% Solvent B to 100% Solvent A over 1 min, and then held constant at 100% Solvent A for 15 more min. Application of a 1.8 kV distal voltage electrosprayed the eluted analytes directly into the ion trap mass spectrometer. Masses were recorded over a 50-1000 m/z range in positive mode.

Data were analyzed with the XCMS software package (190) and by manual inspection. An aligned peak list was generated by XCMS for each sample, including the integrated ion counts. Those ions with integrated intensity less than 1e5 were filtered from the sample. The integrated peak areas were then used to calculate fold changes for pair wise comparisons of H37Rv:*Δigr* and complement:*Δigr*. Ions unique to H37Rv, complement or *Δigr* were identified by generating a subset of those ions with fold changes greater than 2.5. The subset was then reduced to only those ions derived from cholesterol by comparing the isotope distributions for the ion in natural abundance and LDL-[¹³C]-cholesterol samples.

XV. Purification and characterization of cholesterol-derived metabolite 1

Metabolite **1** was purified by reverse phase HPLC on a Phenomenex Luna C18 column (5 μm, 250 × 10 mm) from extracts of a 500 mL *Δigr M. tuberculosis* culture grown with unlabeled cholesterol. A flow rate of 3 mL/min was used with a gradient from H₂O (Solvent C) to MeOH (Solvent D). The gradient was held at 100% C for 5 min, then changed to 50% C over 15 min. Next a linear change from 50% C to 20% C over 25 min, then to 0% C over 5 min. Finally the gradient was held at 0% C for 10 min. Fractions containing the desired metabolite were combined and dried for NMR analysis in CD₃OD on a 900 MHz cryoprobe spectrometer.

XVI. FadE28-FadE29 substrate synthesis

Carboxylic acids, 1β-(2'-propanoic acid)-3α-H-7αβ-methylhexahydro-4-indanone (191-193) and 3-oxo-4-pregnene-20-carboxylic acid (194) were prepared by ozonolysis of vitamin D₂ and 4,22-stigmastadien-3-one, respectively. Briefly, the ozonolysis was performed in CH₂Cl₂ at -78 °C. The reaction was purged with ozone until the solution turned blue in color, then purged with O₂ until the blue dissipated. 1% pyridine was added

to the ozonolysis of 4,22-stigmastadiene-3-one to selectively ozonolyse the 22-ene and the reaction was monitored by TLC (195). Dimethyl sulfide (10 eq) was added and the reaction was allowed to warm to rt and stirred for 12 to 15 h. Solvent was removed under reduced pressure and the product was redissolved in 10% H₂O/CH₃CN and chilled on ice. Sodium chlorite (10 eq) was added and the reaction stirred for 15 h at rt. The reaction was dried under reduced pressure and the product extracted from acidified brine with CH₂Cl₂. The products were purified by silica gel chromatography in 1:4 or 2:3 EtOAc:hexanes. **3-oxo-4-pregnene-20-carboxylic acid**: ¹H NMR (500 MHz, CDCl₃) δ 9.5 (br, 1H, COOH), 5.8 (s, 1H, C4-H), 1.22 (d, 3H, *J* = 6.5 Hz, C21-H), 1.18 (s, 3H, C19-H), 0.73 (s, 3H, C18-H); ¹³C NMR δ 203 (C3), 126 (C4), 175 (C5), 180 (C22). **1β-(2'-propanoic acid)-3α-H-7αβ-methylhexahydro-4-indanone**: ¹H NMR (500 MHz, CDCl₃) δ 1.30 (d, 3H, *J* = 6.5 Hz, C3'-H), δ 0.69 (s, 3H, C8-H) ¹³C NMR δ 213 (C4), 184 (C1'). LC/ESI-MS indicated a purity of greater than 95%.

The carboxylic acids prepared above as well as butanoic acid, 3-methylbutanoic acid, and 2-methylpropanoic acid were converted to their corresponding acyl-CoA thioesters via the mixed anhydride method. Briefly, Et₃N (2 eq) was added to the acid in CH₂Cl₂ and stirred for 10 min. Then ethyl chloroformate (2 eq) was added dropwise to the reaction on ice. The reaction was allowed to warm to rt and stirred for 2 h, then dried under a stream of N₂. The mixed anhydride was dissolved in THF and filtered through glass wool into coenzyme A in H₂O:THF (1:4), pH 8.0. 5,5'-dithiobis-(2-nitrobenzoic acid was used to monitor the disappearance of free thiol, the reaction was then acidified to pH 4, and THF was removed. Remaining free acid was extracted into Et₂O, and the product in the aqueous layer was purified by reverse phase HPLC with a Phenomenex Luna C18 column (5 μm, 250 × 10 mm). A flow rate of 3 mL/min was used with a gradient from 100% 20 mM ammonium acetate to 100% MeOH. The identities of CoA thioesters were confirmed by LC/MS.

XVII. Dehydrogenase assay

FadE28-FadE29 (3.2 $\mu\text{g/mL}$) was assayed for dehydrogenase activity with ferrocenium hexafluorophosphate (250 μM) as the artificial electron acceptor and 100 μM substrate in Buffer I (196). The formation of product was followed spectrophotometrically at 300 nm and 25 $^{\circ}\text{C}$. Specific activities were calculated from the initial velocity over the first 100 s of reaction and the extinction coefficient of ferrocenium hexafluorophosphate, 3.4 $\text{mM}^{-1} \text{cm}^{-1}$. Assays were conducted in triplicate and substrates tested include 1 β -(2'-propanoyl-CoA)-3 α -H-7 α β -methylhexahydro-4-indanone **2**, 3-oxo-4-pregnene-20-carboxyl-CoA **3**, propanoyl-CoA, butyryl-CoA, isovaleryl-CoA, and isobutyryl-CoA. Product formation was confirmed by MALDI-TOF MS in positive ion mode with 2,5-dihydroxybenzoic acid matrix. Negative controls without substrate or without enzyme were conducted to check for background reduction of ferrocenium hexafluorophosphate.

XVIII. Hydratase assay

FadE28-FadE29 was used to synthesize **4** and **5** (Figure 2-26) from **2** and **3** (Table 2-4) *in situ*. 100 μM of **4** or **5** was incubated with 500 nM FadE28-FadE29 in Buffer I with 250 μM ferrocenium hexafluorophosphate at 20 $^{\circ}\text{C}$. After 30 min 500 nM Rv3541c-Rv3542c was added. The reaction was quenched for analysis by spotting 1 μL of assay mixture with 1 μL of 2,5-dihydroxybenzoic acid matrix prepared in 0.1% TFA/49.9% H_2O /50% ACN. Product formation was confirmed by MALDI-TOF MS in positive ion mode. Negative controls were conducted without Rv3541c-Rv3542c.

XIX. Identification and quantification of flavin cofactor

FadE28-FadE29 was denatured by heating in boiling water. The sample was chilled on ice and centrifuged to pellet precipitated protein. The supernatant was analyzed by reverse phase LC/UV/MS in ESI positive mode and compared to FAD standard. Flavin was quantified using the absorbance and extinction coefficient of FAD at 260 nm. The protein was solubilized in 6 M guanidine-HCl and the concentration determined using the sum of the extinction coefficients for FadE28 and FadE29 at 280 nm (Table 3-2).

XX. Analytical size exclusion chromatography

FadE29 and FadE28-FadE29 samples were analyzed by analytical size exclusion chromatography on a Superdex 75 (10/300 GL) column (GE Healthcare). The column was equilibrated with Buffer J. Samples were eluted isocratically at a flow rate of 0.5 mL/min in

Buffer J, monitoring at 220 and 280 nm. Several standard proteins, including ribonuclease A, lysozyme, TEV protease, carbonic anhydride, aprotinin, BSA, and conalbumin, were analyzed under the same conditions to generate standard curves to estimate molecular weights of analyzed proteins.

XXI. Analytical ultracentrifugation sedimentation equilibrium

Molecular weights were determined using analytical ultracentrifugation sedimentation equilibrium. FadE29 (8.5, 3.4, and 1.7 μM) in Buffer K and FadE28-FadE29 (10.6 μM , 5.3 μM , and 2.6 μM) in Buffer G, were centrifuged at speeds of 20k, 25k, and 30k at 20 °C. Rv3541c-Rv3542c (5.9 μM , 3.8 μM , and 2.4 μM) were centrifuged at speeds of 10k, 15k, and 20 k at 4°C in Buffer J. Scans were acquired after 18 and 20 h of centrifugation at each speed monitoring at 280 nm. The protein partial-specific volume of 0.7336 for FadE29, 0.7350 for FadE28-FadE29, and 0.7380 for Rv3541c-Rv3542c and the solvent density 0.9994 for FadE29, 1.0079 for FadE28-FadE29, and 1.0092 for Rv3541c-Rv3542c were calculated using SEDNTERP (Biomolecular Interaction Technologies Center at the University of New Hampshire). Data were fit globally to the ideal, single species model using Heteroanalysis (University of Connecticut Analytical Ultracentrifugation Facility) to determine the molecular weight.

XXII. LC/UV/MS of intact protein complexes

Protein complex stoichiometries of FadE28-FadE29 and Rv3541c-Rv3542c were confirmed by LC/UV/MS. Samples were separated on a XBridge BEH 300 C4 3.5 μm column (2.1 x 100 mm) at 40 °C with a linear gradient from 95% A to 95% B over 15 min, where A is 5% isopropanol/0.1% trifluoroacetic acid and B is 99.9% isopropanol/0.1% trifluoroacetic acid. MS spectra were collected in ESI positive ion mode with a cone voltage of 40 V, a capillary voltage of 4.5 kV, and source temperature of 150 °C. MS spectra were deconvolved using ESIprot 1.0 (197) and peaks in the UV 280 nm chromatograms were integrated using R. The integrated peak areas of each protein were divided by the corresponding molar extinction coefficient for the protein to yield the molar concentrations (Table 3-2). Protein stoichiometries were determined from the ratios of these molar concentrations.

Table 3-2. Calculated extinction coefficients for proteins studied in this work.

Protein	Extinction coefficient 280 nm ($M^{-1} \text{ cm}^{-1}$)
FadE28	35,410
FadE29	58,900
Rv3541c	16,960
Rv3542c	59,930

Chapter 4

Conclusions and Future Perspectives

I.	6-azasteroid inhibitors of <i>M. tuberculosis</i> cholesterol metabolism	104
II.	Cholesterol derived metabolite profiling with [1,7,15,22,26- ¹³ C]-cholesterol	105
III.	Investigating the FadE proteome of <i>M. tuberculosis</i>	107
IV.	Studies of heteromeric (<i>R</i>)-hydratases	108
V.	Final conclusions	110

I. 6-azasteroid inhibitors of *M. tuberculosis* cholesterol metabolism

The rise in multidrug resistant strains of *M. tuberculosis* has made antibiotics with novel targets a necessity. The cholesterol metabolism pathway is an attractive area of research because cholesterol is important for growth and persistence of *M. tuberculosis*. Several cholesterol-regulated genes are upregulated in macrophages and have been shown to be important for growth *in vivo*. For example, the cholesterol transport mutant $\Delta mce4$ shows growth attenuation in the mouse model, indicating cholesterol import is important for *M. tuberculosis* survival during infection.

The enzyme 3β -HSD is proposed to be the first step of cholesterol metabolism and inhibitors of 3β -HSD are important for targeting the cholesterol metabolic pathway. In chapter 2.1 we tested a series of azasteroids as inhibitors of *M. tuberculosis* 3β -HSD. *In vitro*, 3β -HSD catalyzes the oxidation of DHEA, pregnenolone, and cholesterol. The substrate preference could not be established during steady-state kinetic analysis because detergent was required to solubilize the hydrophobic sterol substrates (Scheme 2-1). Our SAR study identified a 100 nM 6-azasteroid inhibitor, azasteroid **7** (Table 2-1). This compound contains an 8-carbon side chain analogous to the side chain of cholesterol. Analysis of the inhibition mechanism indicates azasteroid **7** acts as a competitive inhibitor to DHEA and therefore binds to the steroid substrate pocket of 3β -HSD. Inhibitors with shorter substituents at the C17 position showed reduced inhibition, indicating that 3β -HSD likely prefers C27 sterols, consistent with cholesterol being the preferred substrate of *M. tuberculosis* 3β -HSD and 3β -HSD being the first enzymatic step of cholesterol metabolism. *In vivo* inhibition studies with a small molecule inhibitor of cholesterol metabolism are necessary to provide insight into the potential of targeting *M. tuberculosis* cholesterol metabolism.

Reverse genetics experiments give conflicting results as to whether cholesterol metabolism by *M. tuberculosis* is a good drug target, since several cholesterol-metabolizing gene knockouts do not display an *in vivo* phenotype. For instance, recent evaluation of the 3β -HSD knockout strain in guinea pigs showed this enzyme is not required for *M. tuberculosis* growth, though we demonstrate it is the first step of the cholesterol

metabolism pathway. This suggests that cholesterol is not essential as a source of energy for the bacterium *in vivo* but does not exclude additional roles of cholesterol metabolism.

Evaluation of enzymes as potential drug targets using knockout phenotypes can be problematic. Reverse genetics experiments do not rule out the possibility of redundant enzymes or enzyme specificity overlap, which can compensate when a gene knockout removes a single enzyme's activity. This specificity overlap can lead to misinterpretation of reverse genetic phenotypes. While 3 β -HSD appears to not be a viable target in treating *M. tuberculosis* infection the reverse genetics experiments, from which these conclusions are drawn may not be sufficient to identify functions required for *in vivo* survival.

6-azasteroid inhibitors targeting *M. tuberculosis* 3 β -HSD may provide further insight into the impact on blocking cholesterol metabolism *in vivo*. Furthermore, 6-azasteroids likely target multiple enzymes of the cholesterol metabolism pathway. Several enzymes are predicted to bind structurally similar substrates and therefore may have similar substrate binding pockets.

6-azasteroids will soon be tested in *M. tuberculosis* infected mice. Their impact on bacterial load will be investigated over a several week time period. The results of this experiment may help evaluate whether targeting the cholesterol metabolic pathway by small molecule inhibitors is beneficial for the treatment of *M. tuberculosis* infection.

II. Cholesterol-derived metabolite profiling with [1,7,15,22,26-¹³C]-cholesterol

Understanding cholesterol metabolism at the molecular level is required for developing cholesterol metabolism targeting small molecules. This requires fully elucidating the cholesterol metabolism pathway. Transcriptional profiling and global phenotypic profiling experiments have identified 52 genes upregulated by cholesterol and 96 genes important for growth of *M. tuberculosis* on cholesterol, respectively. The challenge remains to determine the biochemical function of these genes in the cholesterol metabolism pathway. Several genes involved in cholesterol metabolism have been studied using recombinant enzymes and/or mutant strains and the functions of several others have been proposed based on sequence similarity to characterized actinomycete sterol metabolizing enzymes. Elucidating the cholesterol metabolism pathway is necessary in order to understand the role cholesterol plays in *M. tuberculosis* pathogenesis.

Tracing of ^{13}C isotopically labeled carbon sources by MS or NMR is a powerful technique for establishing pathways and metabolic networks within cells. Commercially available, isotopically labeled, [4- ^{14}C] and [26- ^{14}C] cholesterol has been previously used to study cholesterol metabolism in *M. tuberculosis*. However, the small number of labeled carbons limits the usefulness of this tool. In chapter 2.2 we developed [1,7,15,22,26- ^{13}C]-cholesterol as a tool for MS metabolite-profiling to trace cholesterol carbons through the metabolic pathway of *M. tuberculosis*. Our isotopically labeled cholesterol contains five ^{13}C labels located in the ring system and side chain and allows for tracing additional carbons compared to the commercially available isotopes (Figure 2-4). We applied this tool to study the metabolite profile of H37Rv Δ *igr* strain (Figure 2-7). The *igr* operon is important because *M. tuberculosis* growth is attenuated in mice infected with the Δ *igr* strain. We elucidated the structure of an accumulating metabolite in the Δ *igr* strain by multidimensional NMR and mass spectral analysis. In addition, we proposed structures of additional cholesterol derived metabolites from MS/MS data (Figure 2-8 and Figure 2-9). Metabolites with fully degraded side chains were not identified in the Δ *igr* strain and we conclude the *igr* operon is required for side-chain metabolism. We demonstrate the successful application [1,7,15,22,26- ^{13}C]-cholesterol in identifying cholesterol-derived metabolites and elucidating gene function.

Assigning gene function from sequence data alone is complicated because the *M. tuberculosis* genome encodes large number of genes regulated by cholesterol and many functions are encoded by more than one gene. In the literature, enzymes have been assigned roles in the pathway based on computational annotation from sequence data, without the necessary *in vitro* enzyme characterization or evaluation of *in vivo* metabolic phenotypes.

One possible extension of the work described here is the application of [1,7,15,22,26- ^{13}C]-cholesterol to study additional knockout strains of *M. tuberculosis* to discover gene function and aid in elucidation of the cholesterol metabolism pathway. We predict that cholesterol-derived metabolites will accumulate upstream of the function of the deleted gene, aiding in functional assignment.

The fact that the number of genes upregulated by cholesterol is more than the proposed degradation pathway requires further complicates functional assignment. It is

possible that multiple pathways exist to transform cholesterol. Cholesterol is a known precursor for the biosynthesis of bile acids, vitamin D, and steroid hormones. 3β -HSD of *vaccinia virus* has been shown to be important for the synthesis of glucocorticoid corticosterone, an immunomodulatory, anti-inflammatory compound with the ability to suppress the host immune response. It is possible that *M. tuberculosis* produces similar secondary metabolites from cholesterol, allowing the bacteria to persist in the granuloma by interfering with the host immune response.

Our metabolic [1,7,15,22,26- ^{13}C]-cholesterol in combination with high resolution MS could be applied to studying genes of unknown function regulated by cholesterol but not assigned to the degradative pathway. These genes may be involved in additional cholesterol pathways not yet established. In addition, identifying cholesterol metabolites in wild type strains could help identify additional roles for cholesterol *in vivo*. Extensive metabolic tracing studies with several sampling times may be required to get a complete picture of the cholesterol metabolome in *M. tuberculosis*. Identifying the cholesterol-derived metabolome may help establish the role of cholesterol in *M. tuberculosis* persistence.

III. Investigating the FadE proteome of *M. tuberculosis*

In chapters 2.2 and 2.3 we identified and studied the unique architecture of an isolated protein complex encoded by the *igr* operon, FadE28-FadE29. FadE28-FadE29 was isolated upon coexpressing the proteins in *E. coli*. Analytical size exclusion chromatography, analytical ultracentrifugation sedimentation equilibrium, and reverse phase LC/UV/MS experiments determined that FadE28-FadE29 forms a novel $\alpha_2\beta_2$ tetramer (Figure 2-14, Figure 2-15, and Figure 2-16). We demonstrated that FadE28-FadE29 catalyzes the dehydrogenation of sterol substrates analogous to the characterized metabolite from *M. tuberculosis* Δ *igr* (Table 2-4). From the metabolite profiling data of Δ *igr* and the catalytic activity of FadE28-FadE29, we confirm the protein complex is required for the metabolism of the 2'-propanoate side chain of cholesterol.

It was determined that FadE28-FadE29 noncovalently binds two FAD molecules per tetramer, compared to typical acyl-CoA dehydrogenases (ACADs) that bind four FAD molecules per tetramer. In addition, FadE28 lacks the active site base necessary for

catalysis (Figure 2-19 and Figure 2-20). These results indicate tetrameric FadE28-FadE29 contains two active sites. We propose the evolutionary loss of catalytic function of FadE28 allows for FadE28-FadE29 to bind large steroyl-CoA substrates. β -Oxidation enzymes involved in bacterial sterol metabolism have been mostly unidentified due to the difficulty of assigning function to the large number of β -oxidation encoding genes in the genome. The studies described in this work are the first enzymatic demonstration of an ACAD from *M. tuberculosis* with activity for sterol substrates.

The identification of heterotetrameric FadE28-FadE29 has led to the prediction that additional heterotetrameric FadE complexes are encoded in the *M. tuberculosis* genome. Recent experiments investigating this hypothesis found that FadE23 forms a complex with FadE24, FadE26 forms a complex with FadE27, and FadE31 forms a complex with FadE32 (Meng Yang, Matthew Wipperman, Nicole S. Sampson, unpublished results). Adjacent *fadE* genes of a single operon encode each isolated protein complex. FadE26, FadE27, FadE31 and FadE32 are upregulated by cholesterol.

Predictions based on phenotypic profiling experiments have assigned several *fadE* genes to the *M. tuberculosis* cholesterol metabolism. Our work, however, demonstrates that several *fadE* gene products form and presumably function as heteromeric complexes *in vivo*. The assignment of individual FadEs to individual steps in the cholesterol metabolism pathway before further experimental analysis is therefore problematic.

We suggest the FadE heterotetramer architecture is important for the ability to bind sterol substrates. Further studies are necessary to substantiate this hypothesis. Experiments are in progress to identify additional heteromeric FadE complexes in *M. tuberculosis*. We will explore the substrate specificity of each complex and investigate the structural assembly of this unique ACAD motif. Crystallization trials are underway to obtain a three dimensional structure of the FadE heteroassociation and will provide insight into the role of this unique architecture in binding of sterol substrates.

IV. Studies of heteromeric (R)-hydratases

In chapters 2.2 and 2.4 we identified and studied the unique architecture of an isolated protein complex, encoded by the *igr* operon, Rv3541c-Rv3542c. Rv3541c-Rv3542c was isolated upon coexpressing the proteins in *E. coli*. Analytical size exclusion

chromatography, analytical ultracentrifugation sedimentation equilibrium, and reverse phase LC/UV/MS experiments determined that Rv3541c-Rv3542c forms a novel $\alpha_2\beta_2$ tetramer (Figure 2-23, Figure 2-25, and Figure 2-24). We demonstrated that Rv3541c-Rv3542c catalyzes hydration reactions of sterol substrates (Figure 2-26 and Figure 2-27). From the metabolite profiling data of *Δigr* and the catalytic activity of Rv3541c-Rv3542c, we confirm the protein complex is required for the metabolism of the 2'-propanoate side chain of cholesterol.

The primary sequence and three dimensional structure of (*R*)- and (*S*)-hydratases are unrelated. Sequence analysis of Rv3541 and Rv3542c indicate they are (*R*)-hydratases. Typically (*S*)-hydratases are involved in β -oxidation in bacteria and (*R*)-hydratases are involved in polyhydroxyalkanoate (PHA) biosynthesis. Our results indicate a novel role of (*R*)-hydratases in bacterial sterol metabolism.

Sequence analysis and biophysical characterization of the Rv3541c-Rv3542c $\alpha_2\beta_2$ tetramer suggests that it is structurally unique. Rv3541c has a predicted single hot dog fold and Rv3542c presumably folds as a two separate domains each with a hot dog fold. This 6-domain assembly has four active sites, based on sequence analysis, and this architecture has not been observed previously for a (*R*)-hydratase. Two heteromeric (*R*)-hydratases have been identified in the *M. tuberculosis* genome, HadA-HadB and HadB-HadC, which demonstrate specificity for long chain fatty acids. Interestingly only HadB has the conserved (*R*)-hydratase catalytic motif, suggesting HadA and HadC are inactive enzymes. This supports the hypothesis that evolutionary loss of the catalytic function increases the binding pocket to accommodate large substrates. Rv3541c and the N terminal domain of Rv3542c each are proposed to contain the residues necessary for catalytic activity. The role of the second domain of Rv3542c, predicted to be catalytically inactive, may be important for large substrate binding. Further studies are necessary to substantiate this claim.

We have recently obtained a crystal structure of truncated Rv3541c-Rv3542c, where 130 amino acids were proteolytically removed from the C terminus. From the three-dimensional structure, the proposed catalytic residues in Rv3541c and Rv3542c are topologically conserved to those of (*R*)-hydratase from *A. caviae* (PDB 1IQ6). Co-

crystallization with pregnenone **3** is currently underway as well as the testing of catalytic mutants Rv3541c_(D29N)-Rv3542c and Rv3541c-Rv3542c_(D43N). These experiments will be key in defining the substrate binding pockets and determining the number of active sites. Investigation of the substrate specificity is required to establish whether Rv3541c-Rv3542c is specific for sterol substrates.

Our results and the results of others suggest heteromeric enzyme complexes are encoded by cistronic genes. One possible evolutionary path involves acquiring redundant functions by gene duplication. The duplicated gene then evolves losing catalytic function to increase the binding pocket to accommodate polycyclic steroid substrates. Several additional (*R*)-hydratases have been proposed in *M. tuberculosis*, but no two are operonically organized in the genome. Due to the low sequence similarity of the (*R*)-hydratase family it is possible additional (*R*)-hydratase genes are present in the genome but are not annotated.

V. Final conclusions

The role of cholesterol metabolism in *M. tuberculosis* pathogenesis has not been fully established. Elucidation of the cholesterol metabolism pathway is essential to determine the molecular cause of *in vivo* phenotypes observed in cholesterol-metabolizing gene knockout strains. The work presented here has addressed this challenge. We demonstrate the application of isotopically labeled cholesterol to investigate the cholesterol metabolome of *M. tuberculosis* Δ *igr* strain. The *igr* operon is important because it is required for intracellular growth of the bacterium *in vivo*. Our approach can be applied to studying the metabolite profile of additional knockout strains to determine gene function.

In addition we identified novel enzyme architectures of oxidation enzymes involved in cholesterol side-chain metabolism. Knowledge of these assemblies has already aided in identify additional acyl-CoA dehydrogenase enzyme complexes in *M. tuberculosis*. Metabolic and enzymatic characterization of cholesterol metabolizing genes is required to evaluate cholesterol metabolism as a drug target and develop antibiotics that inhibit the pathway.

Chapter 5

References

1. Zhang, Y. (2005) The magic bullets and tuberculosis drug targets, *Annu Rev Pharmacol Toxicol* 45, 529-564.
2. Mwinga, A., and Bernard Fourie, P. (2004) Prospects for new tuberculosis treatment in Africa, *Trop Med Int Health* 9, 827-832.
3. Rattan, A., Kalia, A., and Ahmad, N. (1998) Multidrug-resistant *Mycobacterium tuberculosis*: molecular perspectives, *Emerg Infect Dis* 4, 195-209.
4. Fischbach, M. A., and Walsh, C. T. (2009) Antibiotics for emerging pathogens, *Science* 325, 1089-1093.
5. Pieters, J. (2008) *Mycobacterium tuberculosis* and the macrophage: maintaining a balance, *Cell Host Microbe* 3, 399-407.
6. Park, J. S., Tamayo, M. H., Gonzalez-Juarrero, M., Orme, I. M., and Ordway, D. J. (2006) Virulent clinical isolates of *Mycobacterium tuberculosis* grow rapidly and induce cellular necrosis but minimal apoptosis in murine macrophages, *J Leukoc Biol* 79, 80-86.
7. Russell, D. G., Cardona, P. J., Kim, M. J., Allain, S., and Altare, F. (2009) Foamy macrophages and the progression of the human tuberculosis granuloma, *Nat Immunol* 10, 943-948.
8. Caceres, N., Tapia, G., Ojanguren, I., Altare, F., Gil, O., Pinto, S., Vilaplana, C., and Cardona, P. J. (2009) Evolution of foamy macrophages in the pulmonary granulomas of experimental tuberculosis models, *Tuberculosis (Edinb)* 89, 175-182.
9. Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffe, M., Emile, J. F., Marchou, B., Cardona, P. J., de Chastellier, C., and Altare, F. (2008) Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for *M. tuberculosis* persistence, *PLoS Pathog* 4, e1000204.
10. Barry, C. E., 3rd, Boshoff, H. I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson, R. J., and Young, D. (2009) The spectrum of latent tuberculosis: rethinking the biology and intervention strategies, *Nat Rev Microbiol* 7, 845-855.
11. Jarlier, V., and Nikaido, H. (1994) Mycobacterial cell wall: structure and role in natural resistance to antibiotics, *FEMS Microbiol Lett* 123, 11-18.

12. De Rossi, E., Ainsa, J. A., and Riccardi, G. (2006) Role of mycobacterial efflux transporters in drug resistance: an unresolved question, *FEMS Microbiol Rev* 30, 36-52.
13. Kochi, A., Vareldzis, B., and Styblo, K. (1993) Multidrug-resistant tuberculosis and its control, *Res Microbiol* 144, 104-110.
14. Mandavilli, A. (2007) Virtually incurable TB warns of impending disaster, *Nat Med* 13, 271.
15. Marris, E. (2006) Extreme TB strain threatens HIV victims worldwide, *Nature* 443, 131.
16. Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011) The challenge of new drug discovery for tuberculosis, *Nature* 469, 483-490.
17. (2011) World Health Organization. Global Tuberculosis Control.
18. Villemagne, B., Crauste, C., Flipo, M., Baulard, A. R., Déprez, B., and Willand, N. (2012) Tuberculosis: the drug development pipeline at a glance, *Eur J Med* 51, 1-16.
19. Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H. W., Neefs, J. M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N., and Jarlier, V. (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*, *Science* 307, 223-227.
20. Sita Lumsden, E. G., and Swoboda, J. A. (1952) Isoniazid in the treatment of pulmonary tuberculosis, *Tubercle* 33, 322-329.
21. Thomas, J. P., Baughn, C. O., Wilkinson, R. G., and Shepherd, R. G. (1961) A new synthetic compound with antituberculous activity in mice: ethambutol (dextro-2,2'-(ethylenediimino)-di-l-butanol), *Am Rev Respir Dis* 83, 891-893.
22. Makarov, V., Manina, G., Mikusova, K., Mollmann, U., Ryabova, O., Saint-Joanis, B., Dhar, N., Pasca, M. R., Buroni, S., Lucarelli, A. P., Milano, A., De Rossi, E., Belanova, M., Bobovska, A., Dianiskova, P., Kordulakova, J., Sala, C., Fullam, E., Schneider, P., McKinney, J. D., Brodin, P., Christophe, T., Waddell, S., Butcher, P., Albrethsen, J., Rosenkrands, I., Brosch, R., Nandi, V., Bharath, S., Gaonkar, S., Shandil, R. K., Balasubramanian, V., Balganes, T., Tyagi, S., Grosset, J., Riccardi, G., and Cole, S. T. (2009) Benzothiazinones kill *Mycobacterium tuberculosis* by blocking arabinan synthesis, *Science* 324, 801-804.

23. Singh, R., Manjunatha, U., Boshoff, H. I., Ha, Y. H., Niyomrattanakit, P., Ledwidge, R., Dowd, C. S., Lee, I. Y., Kim, P., Zhang, L., Kang, S., Keller, T. H., Jiricek, J., and Barry, C. E., 3rd. (2008) PA-824 kills nonreplicating *Mycobacterium tuberculosis* by intracellular NO release, *Science* 322, 1392-1395.
24. Matsumoto, M., Hashizume, H., Tomishige, T., Kawasaki, M., Tsubouchi, H., Sasaki, H., Shimokawa, Y., and Komatsu, M. (2006) OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice, *PLoS Med* 3, e466.
25. Hosaka, M., Yasue, T., Fukuda, H., Tomizawa, H., Aoyama, H., and Hirai, K. (1992) In vitro and in vivo antibacterial activities of AM-1155, a new 6-fluoro-8-methoxy quinolone, *Antimicrob Agents Chemother* 36, 2108-2117.
26. Ji, B., Lounis, N., Maslo, C., Truffot-Pernot, C., Bonnafous, P., and Grosset, J. (1998) In vitro and in vivo activities of moxifloxacin and clinafloxacin against *Mycobacterium tuberculosis*, *Antimicrob Agents Chemother* 42, 2066-2069.
27. Garrod, L. P. (1950) The nature of the action of streptomycin on tubercle bacilli, *Amer Rev Tuberculosis* 62, 582-585.
28. Kushner, S., and Dalalian, H. (1948) Experimental chemotherapy of tuberculosis; substituted nicotinamides, *J Org Chem* 13, 834-836.
29. Barbachyn, M. R., Hutchinson, D. K., Brickner, S. J., Cynamon, M. H., Kilburn, J. O., Klemens, S. P., Glickman, S. E., Grega, K. C., Hedges, S. K., Toops, D. S., Ford, C. W., and Zurenko, G. E. (1996) Identification of a novel oxazolidinone (U-100480) with potent antimycobacterial activity, *J Med Chem* 39, 680-685.
30. Calvori, C., Frontali, L., Leoni, L., and Tecce, G. (1965) Effect of rifamycin on protein synthesis, *Nature* 207, 417-418.
31. Tahlan, K., Wilson, R., Kastrinsky, D. B., Arora, K., Nair, V., Fischer, E., Barnes, S. W., Walker, J. R., Alland, D., Barry, C. E., 3rd, and Boshoff, H. I. (2012) SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*, *Antimicrobial agents and chemotherapy* 56, 1797-1809.
32. Koul, A., Vranckx, L., Dendouga, N., Balemans, W., Van den Wyngaert, I., Vergauwen, K., Gohlmann, H. W., Willebrords, R., Poncelet, A., Guillemont, J., Bald, D., and Andries, K. (2008) Diarylquinolines are bactericidal for dormant mycobacteria as a result of disturbed ATP homeostasis, *J Biol Chem* 283, 25273-25280.

33. Rao, S. P., Alonso, S., Rand, L., Dick, T., and Pethe, K. (2008) The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis*, *Proc Natl Acad Sci U S A* 105, 11945-11950.
34. Cole, S. T., Brosch, R., Parkhil, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry 3rd, C. E., Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature* 393, 537-544.
35. Camus, J. C., Pryor, M. J., Medigue, C., and Cole, S. T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv, *Microbiology* 148, 2967-2973.
36. Segal, W., and Bloch, H. (1956) Biochemical differentiation of *Mycobacterium tuberculosis* grown *in vivo* and *in vitro*, *J Bacteriol* 72, 132-151.
37. Pandey, A. K., and Sasseti, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol, *Proc Natl Acad Sci U S A* 105, 4376-4380.
38. Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B., and Guilhot, C. (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis, *Mol Microbiol* 34, 257-267.
39. Kim, M. J., Wainwright, H. C., Locketz, M., Bekker, L. G., Walther, G. B., Dittrich, C., Visser, A., Wang, W., Hsu, F. F., Wiehart, U., Tsenova, L., Kaplan, G., and Russell, D. G. (2010) Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism, *EMBO Mol Med* 2, 258-274.
40. Smith, C. V., Huang, C. C., Miczak, A., Russell, D. G., Sacchettini, J. C., and Honer zu Bentrup, K. (2003) Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*, *J Biol Chem* 278, 1735-1743.
41. Munoz-Elias, E. J., and McKinney, J. D. (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence, *Nature Med* 11, 638-644.
42. McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., Jr., and Russell, D. G. (2000)

Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase, *Nature* 406, 735-738.

43. Gould, T. A., van de Langemheen, H., Munoz-Elias, E. J., McKinney, J. D., and Sacchettini, J. C. (2006) Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in *Mycobacterium tuberculosis*, *Mol Microbiol* 61, 940-947.
44. Munoz-Elias, E. J., Upton, A. M., Cherian, J., and McKinney, J. D. (2006) Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence, *Mol Microbiol* 60, 1109-1122.
45. Eisenreich, W., Dandekar, T., Heesemann, J., and Goebel, W. (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence, *Nat Rev Microbiol* 8, 401-412.
46. Rhee, K. Y., de Carvalho, L. P., Bryk, R., Ehrt, S., Marrero, J., Park, S. W., Schnappinger, D., Venugopal, A., and Nathan, C. (2011) Central carbon metabolism in *Mycobacterium tuberculosis*: an unexpected frontier, *Trends Microbiol* 19, 307-314.
47. Orlova, E. V., Sherman, M. B., Chiu, W., Mowri, H., Smith, L. C., and Gotto, A. M., Jr. (1999) Three-dimensional structure of low density lipoproteins by electron cryomicroscopy, *Proc Natl Acad Sci U S A* 96, 8420-8425.
48. Mohn, W. W., van der Geize, R., Stewart, G. R., Okamoto, S., Liu, J., Dijkhuizen, L., and Eltis, L. D. (2008) The actinobacterial *mce4* locus encodes a steroid transporter, *J Biol Chem* 283, 35368-35374.
49. Miner, M. D., Chang, J. C., Pandey, A. K., Sasseti, C. M., and Sherman, D. R. (2009) Role of cholesterol in *Mycobacterium tuberculosis* infection, *Indian J Exp Biol* 47, 407-411.
50. Arima, K., Nagasawa, M., Bae, M., and Tamura, G. (1969) Microbial transformation of sterols Part I. Decomposition of cholesterol by microorganisms, *Agr Biol Chem* 33, 1636-1643.
51. van der Geize, R., and Dijkhuizen, L. (2004) Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications, *Curr Opin Microbiol* 7, 255-261.
52. Malaviya, A., and Gomes, J. (2008) Androstenedione production by biotransformation of phytosterols, *Bioresource Technol* 99, 6725-6737.

53. Sih, C. J., Tai, H. H., and Tsong, Y. Y. (1967) The mechanism of microbial conversion of cholesterol into 17-keto steroids, *J Am Chem Soc* 89, 1957-1958.
54. Sih, C. J., Tai, H. H., Tsong, Y. Y., Lee, S. S., and Coombe, R. G. (1968) Mechanisms of steroid oxidation by microorganisms. XIV. Pathway of cholesterol side-chain degradation, *Biochemistry* 7, 808-818.
55. Sih, C. J., Wang, K. C., and Tai, H. H. (1968) Mechanisms of steroid oxidation by microorganisms. XIII. C22 acid intermediates in degradation of cholesterol side chain, *Biochemistry* 7, 796-807.
56. Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M. C., Sim, E., Dijkhuizen, L., Davies, J. E., Mohn, W. W., and Eltis, L. D. (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages, *Proc Natl Acad Sci U S A* 104, 1947-1952.
57. Rostoniec, K. Z., Wilbrink, M. H., Capyk, J. K., Mohn, W. W., Ostendorf, M., van der Geize, R., Dijkhuizen, L., and Eltis, L. D. (2009) Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1, *Mol Microbiol* 74, 1031-1043.
58. Nesbitt, N., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N. S., and Dubnau, E. (2010) A thiolase of *M. tuberculosis* is required for virulence and for production of androstenedione and androstadienedione from cholesterol, *Infect Immun* 78, 275-282.
59. Griffin, J. W., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., and Sasseti, C. M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism, *PLoS Pathog* 7, e1002251.
60. Capyk, J. K., Kalscheuer, R., Stewart, G. R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs Jr., W. R., Eltis, L. D., and Mohn, W. W. (2009) Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27-steroids, *J Biol Chem* 284, 35534-35542.
61. McLean, K. J., Lafite, P., Levy, C., Cheesman, M. R., Mast, N., Pikuleva, I. A., Leys, D., and Munro, A. W. (2009) The structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection, *J Biol Chem* 284, 35524-35533.
62. Ouellet, H., Guan, S., Johnston, J. B., Chow, E. D., Kells, P. M., Burlingame, A. L., Cox, J. S., Podust, L. M., and de Montellano, P. R. (2010) *Mycobacterium tuberculosis*

CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one, *Mol Microbiol* 77, 730-742.

63. Chang, J. C., Harik, N. S., Liao, R. P., and Sherman, D. R. (2007) Identification of Mycobacterial genes that alter growth and pathology in macrophages and in mice, *J Infect Dis* 196, 788-795.
64. Chang, J. C., Miner, M. D., Pandey, A. K., Gill, W. P., Harik, N. S., Sasseti, C. M., and Sherman, D. R. (2009) *igr* Genes and *Mycobacterium tuberculosis* cholesterol metabolism, *J Bacteriol* 191, 5232-5239.
65. Driscoll, M. D., McLean, K. J., Levy, C., Mast, N., Pikuleva, I. A., Lafite, P., Rigby, S. E., Leys, D., and Munro, A. W. (2010) Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen, *J Biol Chem* 285, 38270-38282.
66. Johnston, J. B., Ouellet, H., and Ortiz de Montellano, P. R. (2010) Functional redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses, *J Biol Chem* 285, 36352-36360.
67. Horinouchi, S., Ishizuka, H., and Beppu, T. (1991) Cloning, nucleotide sequence, and transcriptional analysis of the NAD(P)-dependent cholesterol dehydrogenase gene from a *Nocardia* sp. and its hyperexpression in *Streptomyces* spp, *Appl Environ Microbiol* 57, 1386-1393.
68. Horinouchi, M., Hayashi, T., and Kudo, T. (2012) Steroid degradation in *Comamonas testosteroni*, *J Steroid Biochem Mol Biol* 129, 4-14.
69. Rostoniec, K. Z., Wilbrink, M., Capyk, J. K., Mohn, W. W., Ostendorf, M., van der Geize, R., Dijkhuizen, L., and Eltis, L. D. (2009) Cytochrome P450 125 (CYP125) catalyzes C26-hydroxylation to initiate sterol side chain degradation in *Rhodococcus jostii* RHA1, *Mol Microbiol* 74, 1031-1043.
70. Ishizaki, T., Hirayama, N., Shinkawa, H., Nimi, O., and Murooka, Y. (1989) Nucleotide sequence of the gene for cholesterol oxidase from a *Streptomyces* sp., *J Bacteriol* 171, 596-601.
71. Machang'u, R. S., and Prescott, J. F. (1991) Purification and properties of cholesterol oxidase and choline phosphohydrolase from *Rhodococcus equi*, *Can J Vet Res* 55, 332-340.

72. Drzyzga, O., Fernández De Las Heras, L., Morales, V., Navarro Llorens, J. M., and Perera, J. (2011) Cholesterol degradation by *Gordonia cholesterolivorans*, *Appl Environ Microbiol* 77, 4802-4810.
73. Yang, X., Dubnau, E., Smith, I., and Sampson, N. S. (2007) Rv1106c from *Mycobacterium tuberculosis* is a 3 β -hydroxysteroid dehydrogenase, *Biochemistry* 46, 9058-9067.
74. Yang, X., Gao, J., Smith, I., Dubnau, E., and Sampson, N. S. (2011) Cholesterol is not an essential source of nutrition for *Mycobacterium tuberculosis* during infection, *J Bacteriol* 193, 1473-1476.
75. Sih, C. J., and Bennett, R. E. (1962) Steroid 1-dehydrogenase of *Nocardia restrictus*, *Biochim Biophys Acta* 56, 584-592.
76. Kaufmann, G., Thole, H., Kraft, R., and Atrat, P. (1992) Steroid-1-dehydrogenase of *Rhodococcus erythropolis*: purification and N-terminal amino acid sequence, *J Steroid Biochem Mol Biol* 43, 297-301.
77. Horhold, C., Groh, H., Danhardt, S., Lestrovaja, N. N., and Schubert, K. (1975) Steroid-transforming enzymes from microorganisms. I. Production and purification of Δ^4 -3-ketosteroid-5 α -reductase from *Mycobacterium smegmatis*, *Z Allg Mikrobiol* 15, 563-564.
78. Knol, J., Bodewits, K., Hessels, G. I., Dijkhuizen, L., and van der Geize, R. (2008) 3-Keto-5 α -steroid Δ^1 -dehydrogenase from *Rhodococcus erythropolis* SQ1 and its orthologue in *Mycobacterium tuberculosis* H37Rv are highly specific enzymes that function in cholesterol catabolism, *Biochem J* 410, 339-346.
79. Brzostek, A. (2005) Identification and targeted disruption of the gene encoding the main 3-ketosteroid dehydrogenase in *Mycobacterium smegmatis*, *Microbiology* 151, 2393-2402.
80. van der Geize, R., Hessels, G. I., and Dijkhuizen, L. (2002) Molecular and functional characterization of the *kstD2* gene of *Rhodococcus erythropolis* SQ1 encoding a second 3-ketosteroid Δ^1 -dehydrogenase isoenzyme, *Microbiology* 148, 3285-3292.
81. Brzostek, A., Pawelczyk, J., Rumijowska-Galewicz, A., Dziadek, B., and Dziadek, J. (2009) *Mycobacterium tuberculosis* is able to accumulate and utilize cholesterol, *J Bacteriol* 191, 6584-6591.
82. Andor, A., Jekkel, A., Hopwood, D. A., Jeanplong, F., Ilkoy, E., Konya, A., Kurucz, I., and Ambrus, G. (2006) Generation of useful insertionally blocked sterol degradation

- pathway mutants of fast-growing mycobacteria and cloning, characterization, and expression of the terminal oxygenase of the 3-ketosteroid 9 α -hydroxylase in *Mycobacterium smegmatis* mc(2)155, *Appl Environ Microbiol* 72, 6554-6559.
83. van der Geize, R., Hessels, G. I., van Gerwen, R., van der Meijden, P., and Dijkhuizen, L. (2002) Molecular and functional characterization of *kshA* and *kshB*, encoding two components of 3-ketosteroid 9 α -hydroxylase, a class IA monooxygenase, in *Rhodococcus erythropolis* strain SQ1, *Mol Microbiol* 45, 1007-1018.
 84. Strijewski, A. (1982) The steroid-9 α -hydroxylation system from *Nocardia* species, *Eur J Biochem* 128, 125-135.
 85. Petrusma, M., Dijkhuizen, L., and Van Der Geize, R. (2009) *Rhodococcus rhodochrous* DSM 43269 3-ketosteroid 9 α -hydroxylase, a two-component iron-sulfur-containing monooxygenase with subtle steroid substrate specificity, *Appl Environ Microbiol* 75, 5300-5307.
 86. Capyk, J. K., D'Angelo, I., Strynadka, N. C., and Eltis, L. D. (2009) Characterization of 3-ketosteroid 9 α -hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*, *J Biol Chem* 284, 9937-9946.
 87. Capyk, J. K., Casabon, I., Gruninger, R., Strynadka, N. C., and Eltis, L. D. (2011) Activity of 3-ketosteroid 9 α -hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*, *J Biol Chem* 286, 40717-40724.
 88. Horinouchi, M., Hayashi, T., and Kudo, T. (2004) The genes encoding the hydroxylase of 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione in steroid degradation in *Comamonas testosteroni* TA441, *J Steroid Biochem Mol Biol* 92, 143-154.
 89. Dresen, C., Lin, L. Y., D'Angelo, I., Tocheva, E. I., Strynadka, N., and Eltis, L. D. (2010) A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism, *J Biol Chem* 285, 22264-22275.
 90. Rengarajan, J., Bloom, B. R., and Rubin, E. J. (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages, *Proc Natl Acad Sci U S A* 102, 8327-8332.
 91. Horinouchi, M., Kurita, T., Yamamoto, T., Hatori, E., Hayashi, T., and Kudo, T. (2004) Steroid degradation gene cluster of *Comamonas testosteroni* consisting of 18 putative genes from meta-cleavage enzyme gene *tesB* to regulator gene *tesR*, *Biochem Biophys Res Commun* 324, 597-604.

92. Horinouchi, M., Yamamoto, T., Taguchi, K., Arai, H., and Kudo, T. (2001) Meta-cleavage enzyme gene *tesB* is necessary for testosterone degradation in *Comamonas testosteroni* TA441, *Microbiology* 147, 3367-3375.
93. Yam, K. C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.-X., Sneickus, V., Ly, L. H., Converse, P. J., Jacobs Jr., W. R., Strynadka, N., and Eltis, L. D. (2009) Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*, *PLoS Pathog* 5, e1000344.
94. Lack, N. A., Yam, K. C., Lowe, E. D., Horsman, G. P., Owen, R. L., Sim, E., and Eltis, L. D. (2010) Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism, *J Biol Chem* 285, 434-443.
95. Horinouchi, M., Hayashi, T., Koshino, H., Kurita, T., and Kudo, T. (2005) Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in *Comamonas testosteroni* TA441, *Appl Environ Microbiol* 71, 5275-5281.
96. Hayakawa, S., Kanematsu, Y., Fujiwara, T., and Kako, H. (1976) Microbiological degradation of bile acids. The preparation of some hypothetical metabolites involved in cholic acid degradation, *Biochem J* 154, 577-587.
97. Hayakawa, S., Hashimoto, S., and Onaka, T. (1976) Microbiological degradation of bile acids. Nitrogenous hexahydroindane derivatives formed from cholic acid by *Streptomyces rubescens*, *Biochem J* 160, 745-755.
98. Schomer, U., Sheldrick, W. S., and Wagner, F. (1978) Microbial transformation of β -sitosterol: accumulation of 3-(5 α -Hydroxy-7 $\alpha\beta$ -methyl-1-oxo-3 α H-hexahydroindan-4 α -yl)propionic acid and the X-ray structural identification of its 6-lactone, *J Chem Soc Perkin Trans 1*, 336.
99. Nakamatsu, T., Beppu, T., and Arima, K. (1980) Microbial degradation of steroids to hexahydroindanone derivatives, *Agric Biol Chem* 44, 1469-1474.
100. Miclo, A., and Germain, P. (1992) Hexahydroindanone derivatives of steroids formed by *Rhodococcus equi*, *Appl Microbiol Biotechnol* 36, 456-460.
101. Hashimoto, S., and Hayakawa, S. (1977) Microbiological degradation of bile acids. Metabolites formed from 3-(3 α -hexahydro-7 $\alpha\beta$ -methyl-1,5-dioxoindan-4 α -yl)propionic acid by *Streptomyces rubescens*, *Biochem J* 164, 715-726.

102. Hayakawa, S., Fujiwara, T., and Kako, H. (1976) Microbiological degradation of bile acids. The conjugation of a certain cholic acid metabolite with amino acids in *Corynebacterium equi*, *Biochem J* 160, 757-768.
103. Hayakawa, S., and Fujiwara, T. (1977) Microbiological degradation of bile acids, further degradation of a cholic acid metabolite containing the hexahydroindane nucleus by *Corynebacterium equi*, *Biochem J* 162, 387-397.
104. Fried, J., thoma, R. W., and Klingsberg, A. (1953) Oxidation of steroids by microorganisms. III. side chain degradation, ring D-cleavage and dehydrogenation in ring A, *J Am Chem Soc* 75, 5764-5765.
105. Prairie, R. L., and Talalay, P. (1963) Enzymatic formation of testololactone, *Biochemistry* 2, 203-208.
106. Kołek, T., Szpineter, A., and Swizdor, A. (2008) Baeyer-Villiger oxidation of DHEA, pregnenolone, and androstenedione by *Penicillium lilacinum* AM111, *Steroids* 73, 1441-1445.
107. Hunter, A. C., and Carragher, N. E. (2003) Flexibility of the endogenous progesterone lactonisation pathway in *Aspergillus tamarii* KITA: transformation of a series of cortical steroid analogues, *J Steroid Biochem Mol Biol* 87, 301-308.
108. Hunter, A. C., Coyle, E., Morse, F., Dedi, C., Dodd, H. T., and Koussoroplis, S. J. (2009) Transformation of 5-ene steroids by the fungus *Aspergillus tamarii* KITA: mixed molecular fate in lactonization and hydroxylation pathways with identification of a putative 3 β -hydroxy-steroid dehydrogenase/ Δ 5- Δ 4 isomerase pathway, *Biochim Biophys Acta* 1791, 110-117.
109. van der Geize, R., Grommen, A. W., Hessels, G. I., Jacobs, A. A., and Dijkhuizen, L. (2011) The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development, *PLoS Pathog* 7, e1002181.
110. Hu, Y., van der Geize, R., Besra, G. S., Gurucha, S. S., Liu, A., Rohde, M., Singh, M., and Coates, A. (2010) 3-Ketosteroid 9 α -hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*, *Mol Microbiol* 75, 107-121.
111. Lack, N., Lowe, E. D., Liu, J., Eltis, L. D., Noble, M. E., Sim, E., and Westwood, I. M. (2008) Structure of HsaD, a steroid-degrading hydrolase, from *Mycobacterium tuberculosis*, *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64, 2-7.

112. Simard, J., Ricketts, M. L., Gingras, S., Soucy, P., Feltus, F. A., and Melner, M. H. (2005) Molecular biology of the 3β -hydroxysteroid dehydrogenase/ $\Delta 5$ - $\Delta 4$ isomerase gene family, *Endocr Rev* 26, 525-582.
113. Thomas, J. L., Duax, W. L., Addlagatta, A., Brandt, S., Fuller, R. R., and Norris, W. (2003) Structure/function relationships responsible for coenzyme specificity and the isomerase activity of human type 1 3β -hydroxysteroid dehydrogenase/isomerase, *J Biol Chem* 278, 35483-35490.
114. Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden, H. M. (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase, *Biochemistry* 39, 5691-5701.
115. Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B., and Jornvall, H. (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update, *Chem Biol Interact* 143-144, 247-253.
116. Whiteley, J. M., Xuong, N. H., and Varughese, K. I. (1993) Is dihydropteridine reductase an anomalous dihydrofolate reductase, a flavin-like enzyme, or a short-chain dehydrogenase?, *Adv Exp Med Biol* 338, 115-121.
117. Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L., and Orr, J. C. (1991) Three-dimensional structure of holo 3α , 20β -hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family, *Proc Natl Acad Sci U S A* 88, 10064-10068.
118. Filling, C., Nordling, E., Benach, J., Berndt, K. D., Ladenstein, R., Jornvall, H., and Oppermann, U. (2001) Structural role of conserved Asn179 in the short-chain dehydrogenase/reductase scaffold, *Biochem Biophys Res Commun* 289, 712-717.
119. Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J., and Dziadek, J. (2007) Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*, *FEMS Microbiol Lett* 275, 106-112.
120. Uhía, I., Galán, B., Kendall, S. L., Stoker, N. G., and García, J. L. (2012) Cholesterol metabolism in *Mycobacterium smegmatis*, *Environ Microbiol Rep* 4, 168-182.
121. Ivashina, T. V., Nikolayeva, V. M., Dovbnya, D. V., and Donova, M. V. (2012) Cholesterol oxidase ChoD is not a critical enzyme accounting for oxidation of sterols to 3-keto-4-ene steroids in fast-growing *Mycobacterium sp.* VKM Ac-1815D, *J Steroid Biochem Mol Biol* 129, 47-53.

122. de Carvalho, L. P., Fischer, S. M., Marrero, J., Nathan, C., Ehrt, S., and Rhee, K. Y. (2010) Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon substrates, *Chem Biol* 17, 1122-1131.
123. Chang, J. C., Harik, N. S., Liao, R. P., and Sherman, D. R. (2007) Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice, *J Infect Dis* 196, 788-795.
124. Chang, J. C., Miner, M. D., Pandey, A. K., Gill, W. P., Harik, N. S., Sasseti, C. M., and Sherman, D. R. (2009) *igr* Genes and *Mycobacterium tuberculosis* cholesterol metabolism, *J Bacteriol* 191, 5232-5239.
125. Pandey, A. K., and Sasseti, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol, *Proc Natl Acad Sci USA* 105, 4376-4380.
126. McLean, K. J., Belcher, J., Driscoll, M. D., Fernandez, C. C., Le Van, D., Bui, S., Golovanova, M., and Munro, A. W. (2010) The *Mycobacterium tuberculosis* cytochromes P450: physiology, biochemistry & molecular intervention, *Future Med Chem* 2, 1339-1353.
127. McLean, K. J., Dunford, A. J., Neeli, R., Driscoll, M. D., and Munro, A. W. (2007) Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems, *Arch Biochem Biophys* 464, 228-240.
128. Ouellet, H., Johnston, J. B., and Ortiz de Montellano, P. R. (2010) The *Mycobacterium tuberculosis* cytochrome P450 system, *Arch Biochem Biophys* 493, 82-95.
129. McLean, K. J., Hans, M., and Munro, A. W. (2012) Cholesterol, an essential molecule: diverse roles involving cytochrome P450 enzymes, *Biochem Soc Trans* 40, 587-593.
130. Hudson, S. A., McLean, K. J., Munro, A. W., and Abell, C. (2012) *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets, *Biochem Soc Trans* 40, 573-579.
131. Thorpe, C., and Kim, J. J. P. (1995) Structure and mechanism of action of the acyl-CoA dehydrogenases, *The FASEB Journal* 9, 718-725.
132. Ghisla, S., and Thorpe, C. (2004) Acyl-CoA dehydrogenase: A mechanistic overview *Eur J Biochem* 271, 494-508.

133. Bode, K., Hooks, M. A., and Couee, I. I. (1999) Identification, separation, and characterization of acyl-coenzyme A dehydrogenases involved in mitochondrial β -oxidation in higher plants, *Plant Physiol* 119, 1305-1314.
134. Komuniecki, R., Fekete, S., and Thissen-Parra, J. (1985) Purification and characterization of the 2-methyl branched-chain Acyl-CoA dehydrogenase, an enzyme involved in NADH-dependent enoyl-CoA reduction in anaerobic mitochondria of the nematode, *Ascaris suum*, *J Biol Chem* 260, 4770-4777.
135. Kionka, C., and Kunau, W. H. (1985) Inducible β -oxidation pathway in *Neurospora crassa*, *J Bacteriol* 161, 153-157.
136. Kunau, W. H., Dommès, V., and Schulz, H. (1995) β -oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress, *Prog Lipid Res* 34, 267-342.
137. He, M., Pei, Z., Mohsen, A. W., Watkins, P., Murdoch, G., Van Veldhoven, P. P., Ensenauer, R., and Vockley, J. (2011) Identification and characterization of new long chain acyl-CoA dehydrogenases, *Mol Genet Metab* 102, 418-429.
138. Ensenauer, R., He, M., Willard, J. M., Goetzman, E. S., Corydon, T. J., Vandahl, B. B., Mohsen, A. W., Isaya, G., and Vockley, J. (2005) Human acyl-CoA dehydrogenase-9 plays a novel role in the mitochondrial β -oxidation of unsaturated fatty acids, *J Biol Chem* 280, 32309-32316.
139. Tiffany, K. A., Roberts, D. L., Wang, M., Paschke, R., Mohsen, A. W., Vockley, J., and Kim, J. J. (1997) Structure of human isovaleryl-CoA dehydrogenase at 2.6 Å resolution: structural basis for substrate specificity, *Biochemistry* 36, 8455-8464.
140. Bross, P., Engst, S., Strauss, A. W., Kelly, D. P., Rasched, I., and Ghisla, S. (1990) Characterization of wild-type and an active site mutant of human medium chain acyl-CoA dehydrogenase after expression in *Escherichia coli*, *J Biol Chem* 265, 7116-7119.
141. Eder, M., Krautle, F., Dong, Y., Vock, P., Kieweg, V., Kim, J. J., Strauss, A. W., and Ghisla, S. (1997) Characterization of human and pig kidney long-chain-acyl-CoA dehydrogenases and their role in β -oxidation, *Eur J Biochem* 245, 600-607.
142. Binzak, B., Willard, J., and Vockley, J. (1998) Identification of the catalytic residue of human short/branched chain acyl-CoA dehydrogenase by in vitro mutagenesis, *Biochim Biophys Acta* 1382, 137-142.

143. Stern, J. R., and Del Campillo, A. (1956) Enzymes of fatty acid metabolism. II. Properties of crystalline crotonase, *J Biol Chem* 218, 985-1002.
144. Fukui, T., Shiomi, N., and Doi, Y. (1998) Expression and characterization of (*R*)-specific enoyl coenzyme A hydratase involved in polyhydroxyalkanoate biosynthesis by *Aeromonas caviae*, *J Bacteriol* 180, 667-673.
145. Tsuge, T., Fukui, T., Matsusaki, H., Taguchi, S., Kobayashi, G., Ishizaki, A., and Doi, Y. (2000) Molecular cloning of two (*R*)-specific enoyl-CoA hydratase genes from *Pseudomonas aeruginosa* and their use for polyhydroxyalkanoate synthesis, *FEMS Microbiol Lett* 184, 193-198.
146. Reiser, S. E., Mitsky, T. A., and Gruys, K. J. (2000) Characterization and cloning of an (*R*)-specific trans-2,3-enoylacyl-CoA hydratase from *Rhodospirillum rubrum* and use of this enzyme for PHA production in *Escherichia coli*, *Appl Microbiol Biotechnol* 53, 209-218.
147. Park, S. J., and Lee, S. Y. (2003) Identification and characterization of a new enoyl coenzyme A hydratase involved in biosynthesis of medium-chain-length polyhydroxyalkanoates in recombinant *Escherichia coli*, *J Bacteriol* 185, 5391-5397.
148. Tsuge, T., Taguchi, K., Seiichi, T., and Doi, Y. (2003) Molecular characterization and properties of (*R*)-specific enoyl-CoA hydratases from *Pseudomonas aeruginosa*: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid β -oxidation, *Int J Biol Macromol* 31, 195-205.
149. Koski, K. M., Haapalainen, A. M., Hiltunen, J. K., and Glumoff, T. (2005) Crystal structure of 2-enoyl-CoA hydratase 2 from human peroxisomal multifunctional enzyme type 2, *J Mol Biol* 345, 1157-1169.
150. Qin, Y. M., Haapalainen, A. M., Conry, D., Cuebas, D. A., Hiltunen, J. K., and Novikov, D. K. (1997) Recombinant 2-enoyl-CoA hydratase derived from rat peroxisomal multifunctional enzyme 2: role of the hydratase reaction in bile acid synthesis, *Biochem J* 328 (Pt 2), 377-382.
151. Baes, M. (2000) Inactivation of the peroxisomal multifunctional protein-2 in mice impedes the degradation of not only 2-methyl-branched fatty acids and bile acid intermediates but also of very long chain fatty acids, *J Biol Chem* 275, 16329-16336.
152. Hiltunen, J. K., Wenzel, B., Beyer, A., Erdmann, R., Fossa, A., and Kunau, W. H. (1992) Peroxisomal multifunctional β -oxidation protein of *Saccharomyces cerevisiae*. Molecular analysis of the fox2 gene and gene product, *J Biol Chem* 267, 6646-6653.

153. Goepfert, S., Hiltunen, J. K., and Poirier, Y. (2006) Identification and functional characterization of a monofunctional peroxisomal enoyl-CoA hydratase 2 that participates in the degradation of even cis-unsaturated fatty acids in *Arabidopsis thaliana*, *J Biol Chem* 281, 35894-35903.
154. Hisano, T., Tsuge, T., Fukui, T., Iwata, T., Miki, K., and Doi, Y. (2003) Crystal structure of the (*R*)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis, *J Biol Chem* 278, 617-624.
155. Koski, M. K. (2004) A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2, *J Biol Chem* 279, 24666-24672.
156. Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N., and Suguna, K. (2009) Analysis of proteins with the 'hot dog' fold: prediction of function and identification of catalytic residues of hypothetical proteins, *BMC Struct Biol* 9, 37.
157. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site, *Structure* 4, 253-264.
158. Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z. S. (2000) Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme, *Nat Struct Biol* 7, 555-559.
159. Kostrewa, D., Winkler, F. K., Folkers, G., Scapozza, L., and Perozzo, R. (2005) The crystal structure of PffabZ, the unique β -hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of *Plasmodium falciparum*, *Protein Sci* 14, 1570-1580.
160. Yokoyama, T., Choi, K. J., Bosch, A. M., and Yeo, H. J. (2009) Structure and function of a *Campylobacter jejuni* thioesterase Cj0915, a hexameric hot dog fold enzyme, *Biochim Biophys Acta* 1794, 1073-1081.
161. Wang, F., Langley, R., Gulten, G., Wang, L., and Sacchettini, J. C. (2007) Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence, *Chem Biol* 14, 543-551.
162. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas sp.* Strain CBS-3, *J Biol Chem* 273, 33572-33579.

163. Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D., and Holden, H. M. (2003) The structure of 4-hydroxybenzoyl-CoA thioesterase from *arthrobacter sp.* strain SU, *J Biol Chem* 278, 43709-43716.
164. Thomas, S. T., Yang, X., and Sampson, N. S. (2011) Inhibition of the *M. tuberculosis* 3 β -hydroxysteroid dehydrogenase by azasteroids, *Bioorg Med Chem Lett* 21, 2216-2219.
165. Thomas, S. T., VanderVen, B. C., Sherman, D. R., Russell, D. G., and Sampson, N. S. (2011) Pathway profiling in *Mycobacterium tuberculosis*: elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism, *J Biol Chem* 286, 43668-43678.
166. Baker, M. E., and Blasco, R. (1992) Expansion of the mammalian 3 β -hydroxysteroid dehydrogenase/plant dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a bacterial UDP-galactose-4-epimerase, and open reading frames in *vaccinia virus* and *fish lymphocystis disease virus*, *FEBS Lett* 301, 89-93.
167. Reading, P. C., Moore, J. B., and Smith, G. L. (2003) Steroid hormone synthesis by *vaccinia virus* suppresses the inflammatory response to infection, *J Exp Med* 197, 1269-1278.
168. Sakai, N., Tanaka, M., Takahashi, M., Fukada, S., Mason, J. I., and Nagahama, Y. (1994) Ovarian 3 β -hydroxysteroid dehydrogenase/ Δ 5-4-isomerase of rainbow trout: its cDNA cloning and properties of the enzyme expressed in a mammalian cell, *FEBS Lett* 350, 309-313.
169. Hsu, C. Y., Yu, N. W., and Chen, S. J. (1980) Development of Δ 5-3 β -hydroxysteroid dehydrogenase activity in the interrenal gland of *Rana catesbeiana*, *Gen Comp Endocrinol* 42, 167-170.
170. Finsterbusch, A., Lindemann, P., Grimm, R., Eckerskorn, C., and Luckner, M. (1999) Δ 5-3 β -hydroxysteroid dehydrogenase from *Digitalis lanata Ehrh.* - a multifunctional enzyme in steroid metabolism?, *Planta* 209, 478-486.
171. Frye, S. V., Haffner, C. D., Maloney, P. R., Mook, R. A., Jr., Dorsey, G. F., Jr., Hiner, R. N., Batchelor, K. W., Bramson, H. N., Stuart, J. D., Schweiker, S. L., and et al. (1993) 6-Azasteroids: potent dual inhibitors of human type 1 and 2 steroid 5 α -reductase, *J Med Chem* 36, 4313-4315.
172. Frye, S. V., Haffner, C. D., Maloney, P. R., Mook, R. A., Jr., Dorsey, G. F., Jr., Hiner, R. N., Cribbs, C. M., Wheeler, T. N., Ray, J. A., Andrews, R. C., and et al. (1994) 6-Azasteroids:

- structure-activity relationships for inhibition of type 1 and 2 human 5 α -reductase and human adrenal 3 β -hydroxy- Δ 5-steroid dehydrogenase/3-keto- Δ 5-steroid isomerase, *J Med Chem* 37, 2352-2360.
173. Frye, S. V., Haffner, C. D., Maloney, P. R., Hiner, R. N., Dorsey, G. F., Noe, R. A., Unwalla, R. J., Batchelor, K. W., Bramson, H. N., Stuart, J. D., and et al. (1995) Structure-activity relationships for inhibition of type 1 and 2 human 5 α -reductase and human adrenal 3 β -hydroxy- Δ 5-steroid dehydrogenase/3-keto- Δ 5-steroid isomerase by 6-azaandrost-4-en-3-ones: optimization of the C17 substituent, *J Med Chem* 38, 2621-2627.
 174. Frye, S. V. (2006) Discovery and clinical development of dutasteride, a potent dual 5 α -reductase inhibitor, *Current topics in medicinal chemistry* 6, 405-421.
 175. Isler, O., Ruegg, R., Wursch, J., Gey, K. F., and Pletscher, A. (1957) Zur Biosynthese Des Cholesterins Aus β , Δ -Dihydroxy- β -Methyl-Valeriansaure, *Helv Chim Acta* 40, 2369-2373.
 176. Bloch, K. (1965) Biological Synthesis of Cholesterol, *Science* 150, 19-28.
 177. Popják, G., and Cornforth, J. W. (1966) Substrate stereochemistry in squalene biosynthesis: The first Ciba medal lecture, *Biochem J* 101, 553 b554-568.
 178. Popják, G., Cornforth, R. H., Ryhage, R., Goodman, D. S., and Cornforth, J. W. (1961) Studies on Biosynthesis of Cholesterol XV. Mechanism of Squalene Biosynthesis from Farnesyl Pyrophosphate and from Mevalonate, *J Biol Chem* 236, 1934-1947.
 179. Popjak, G., Edmond, J., Anet, F. A., and Easton Jr., N. R. (1977) Carbon-13 NMR studies on cholesterol biosynthesized from [^{13}C]mevalonates, *J Am Chem Soc* 99, 931-935.
 180. Hayakawa, S., Kanematsu, Y., and Fujiwara, T. (1967) New dehydroxylation reaction observed in the microbiological degradation pathway of cholic acid, *Nature* 214, 520-521.
 181. Kieslich, K. (1985) Microbial side-chain degradation of sterols, *J Basic Microbiol* 25, 461-474.
 182. Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) The Pfam protein families database, *Nucleic Acids Res* 36, D281-288.

183. Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, C. J., Scheremetjew, M., Tate, J., Thimmajananathan, M., Thomas, P. D., Wu, C. H., Yeats, C., and Yong, S. Y. (2012) InterPro in 2011: new developments in the family and domain prediction database, *Nucleic Acids Res* 40, D306-312.
184. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server, *Nat Protoc* 4, 363-371.
185. Krishna, S. S., Aravind, L., Bakolitsa, C., Caruthers, J., Carlton, D., Miller, M. D., Abdubek, P., Astakhova, T., Axelrod, H. L., Chiu, H. J., Clayton, T., Deller, M. C., Duan, L., Feuerhelm, J., Grant, J. C., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kumar, A., Marciano, D., McMullan, D., Morse, A. T., Nigoghossian, E., Okach, L., Reyes, R., Rife, C. L., van den Bedem, H., Weekes, D., Xu, Q., Hodgson, K. O., Wooley, J., Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2010) The structure of SSO2064, the first representative of Pfam family PF01796, reveals a novel two-domain zinc-ribbon OB-fold architecture with a potential acyl-CoA-binding role, *Acta Crystallogr F Struct Biol Crystal Comm* 66, 1160-1166.
186. Sacco, E., Covarrubias, A. S., O'Hare, H. M., Carroll, P., Eynard, N., Jones, T. A., Parish, T., Daffe, M., Backbro, K., and Quemard, A. (2007) The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*, *Proc Natl Acad Sci U S A* 104, 14628-14633.
187. Haffner, C. (1995) Synthesis of 6-azacholesten-3-ones - potent inhibitors of 5 α -reductase, *Tetrahedron Lett* 36, 4039-4042.
188. Bligh, E. G., and Dyer, W. J. (1959) A rapid method of total lipid extraction and purification, *Can J Biochem Physiol* 37, 911-917.
189. Dobson, G., Minnikin, D. E., Minniken, S. M., Parlett, J. H., and Goodfellow, M. (1985) Systematic analysis of complex mycobacterial lipids, In *Chemical methods in bacterial systematics* (Goodfellow, M., and Minnikin, D. E., Eds.), pp 237-265, Academic Press, London.
190. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal Chem* 78, 779-787.

191. Inhoffen, H. H., Quinkert, G., Schutz, S., Friedrich, G., and Tober, E. (1958) Studies in the vitamin D series. XXV. Degradation of vitamins D2 and D3 with the formation of 8-methyl-trans-hydrindan-4-ol-1-one, *Chem Ber* 91, 781-791.
192. Inhoffen, H. H., Bruckner, K., and Grundel, R. (1954) The vitamin D series. I. Rearrangement of vitamin D2 to an isotachysterol and partial synthesis of isovitamin D2, *Chem Ber* 87, 1-13.
193. Mincione, E., Bovicelli, P., and Forcellese, M. L. (1989) Improved conversion of vitamin D2 into the Windaus ketone and its regioselective hydroxylation via organoboranes at C-26, *Synth Commun* 19, 723-735.
194. Leppik, R. A. (1982) Deoxycholic acid degradation by a *Pseudomonas* species. Acidic intermediates from the initial part of the catabolic pathway, *Biochem J* 202, 747-751.
195. Slomp Jr., G., and Johnson, J. L. (1958) Ozonolysis .II. The effect of pyridine on the ozonolysis of 4,22-stigmastadien-3-one, *J Am Chem Soc* 80, 915-921.
196. Lehman, T. C., and Thorpe, C. (1990) Alternate electron acceptors for medium-chain acyl-CoA dehydrogenase: use of ferricenium salts, *Biochemistry* 29, 10594-10602.
197. Winkler, R. (2010) ESIprot: a universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data, *Rapid Commun Mass Spectrom* 24, 285-294.

Appendix

Expression study of *vaccinia virus* 3 β -HSD

133

Expression study of *vaccinia virus* 3 β -HSD

Vaccinia virus, of the poxvirus family, was used as the vaccine that eradicated smallpox in the 1980s. The gene *A44L* of *vaccinia virus* encodes for a 3 β -hydroxysteroid dehydrogenase (3 β -HSD). 3 β -HSD contributes to the virulence of *vaccinia virus* through steroid hormone synthesis (166). Deletion of the gene *A44L* results in a decrease in the levels of the glucocorticoid, corticosterone, in the mouse host, a more rapid recruitment of key lymphocytes, CD4⁺ and CD8⁺ in the lungs, and an increase in IFN- γ production. Production of corticosterone, an anti-inflammatory agent, leads to the suppression of the host immune system in response to infection and may increase the vulnerability of a potential host to infection. This enzyme has been shown in cell assays to catalyze the conversion of progesterone to pregnenolone. In this study we sought to characterize the recombinant enzyme *in vitro*.

Several constructs for expression were prepared using *E. coli* codon optimized *vaccinia virus* *A44L* (Table A1). Under all expression conditions investigated with constructs pVVHSD-1 – pVVHSD-4, *vaccinia virus* 3 β -HSD was localized to inclusion bodies. Attempts to refold insoluble protein solubilized in 6 M guanidine hydrochloride by dilution and dialysis were not successful. Expression temperature (22 °C and 16 °C) and IPTG concentration (50, 100, 200, and 400 μ M) were varied but did not increase *vaccinia virus* 3 β -HSD solubility. *Vaccinia virus* 3 β -HSD was also expressed with folding chaperones dnaK-dnaJ-grpE-groES-groEL, but no soluble protein was obtained.

Expression as the maltose binding protein (MBP) fusion with construct pVVHSD-5 resulted in soluble protein. Protein was purified by affinity purification with amylose resin (Figure A-1). Fusion protein was assayed for 3 β -HSD activity with DHEA, pregnenolone, and cholesterol as substrates and NAD⁺ and NADP⁺ as cofactors. No activity was detected. Attempts to cleave MBP with enterokinase were not successful. In conclusion, soluble, active *vaccinia virus* 3 β -HSD could not be obtained.

Table A-1. Expression constructs for *vaccinia virus* 3 β -HSD.

Construct name	Plasmid	Gene	Fusion tag	Restriction sites	Antibiotic marker
pVVHSD-1	pET28b	VV <i>A44L</i>	N-terminal His ₆	NdeI/HindIII	kan
pVVHSD-2	pET28b	VV <i>A44L</i>	N and C-terminal His ₆	NdeI/HindIII	kan
pVVHSD-3	pET29a	VV <i>A44L</i>	No tag	NdeI/HindIII	kan
pVVHSD-4	pET29a	VV <i>A44L</i>	C-terminal His ₆	NdeI/HindIII	kan
pVVHSD-5	^a pMAL-c4e	VV <i>A44L</i>	N-terminal MBP	EcoRI/HindIII	amp

^a New England BioLabs
 VV: *vaccinia virus*

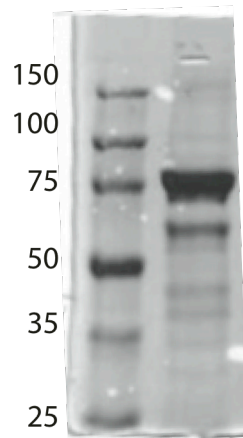


Figure A-1. SDS-PAGE of purified *vaccinia virus* 3 β -HSD- MBP.

ATGGCGGTGTATGCGGTGACCGGCGGCGGGC'TTTC'GGGTTCGCTATATTGTGAAAC'TGCTGATTAG
CGCGGATGATGTT'CAGGAAATTCGTGTGATTGATATTGTTGAAGATCCGCAGCCGATTACCAGCAAAG
TTAAAGTTATTAATTATATTCAGTGCATATTAATGATTTTGATAAAGTGCGCGAAGCGCTGGATGGC
GTGAACCTGATTATTCATACCGCGGCGCTGGTGGATGTGTTTGGCAAATATACCGATAACGAAATTAT
GAAAGTGAATTATTTATGGCACCCAGACCATTCTGGCGGCGTGGTGGATCTGGGCATTAATATCTGA
TTTATACCAGCAGCATGGAAGCGATTGGCCCCGAATAAACATGGTGATCCGTTTATTGGCCATGAACAT
ACCC'TGTATGATATTAGCCCGGGCCATGTGTATGCGAAAAGCAAACGCATGGCGGAACAGCTGGTGAT
GAAAGCGAACAACAGCGTGATTATGAACGGCGCGAAAC'TGTATACCTGCTGCCTGCGTCCGACCGGTA
TTTATGGTGAAGGCGATAAACTGACCAAAGTGT'TTTATGAACAGTGCAAACAGCATGGCAATATTTATG
TATCGTACCGTGGATGATAATGCGGTGCATAGCCGTGTGTATGTTGGTAATGCGGCGTGGATGCATGT
TCTGGCGGCGAAATATATTCAGTATCCGGGTAGCAAATTAAGGTAACGCGTATTTTTGCTATGATT
ATAGCCCGAGCTGTAGCTATGATATGTTTAACCTGCTGCTGATGAAACCGCTGGGCATTAACAGGGT
AGCCGCATTCGCGCTGGATGCTGAAAATGTATGCGTGCAAAAACGATATGAAACGCATTC'TGTTTTCG
CAAACCGAGCTGCTGAATAACTATACCC'TGAAAATTAGCAATACCACCTTTGAAGTGCACCAACA
ATGCGGAAC'TGGATTTTAAC'TATAGCCCGATTTTTGATGTTGATGTGGCGTTTAAACGCACCCGCAA
TGGCTGGAAGAAAGCGAATAA

Figure A-2. Codon optimized gene sequence of *vaccinia virus A44L*.