Stony Brook University

Identification and characterization of cholesterol metabolism related genes and gene products from *Mycobacterium tuberculosis*

A Dissertation Presented by

Meng Yang

to

The Graduate School

In Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

In

Chemistry

Stony Brook University

August 2015

Stony Brook University

The Graduate School

Meng Yang

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
Professor and chair of Chemistry, Stony Brook University

Daniel P. Raleigh, Ph.D., Chairperson of Defense
Professor of Chemistry, Stony Brook University

Jarrod B. French, Ph.D., Third Member of Defense
Professor of Chemistry, Stony Brook University

Miguel Garcia-Diaz, Ph.D., Outside Member of Defense
Professor of Pharmacological Sciences, Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Identification and characterization of cholesterol metabolism related genes and gene products from *Mycobacterium tuberculosis*

by

Meng Yang

Doctor of Philosophy

in

Chemistry

Stony Brook University

2015

Mycobacterium tuberculosis (Mtb), as an intracellular pathogen, preferentially respires on lipids when surviving *in vivo*. Cholesterol metabolism is important for *Mtb*'s persistence and virulence. We demonstrate through experiment and bioinformatic analysis the existence of an architecturally distinct subfamily of acyl-CoA dehydrogenase (ACAD) enzymes that are $\alpha_2\beta_2$ heterotetramers with two active sites. These enzymes are encoded by two adjacent ACAD (*fadE*) genes that are regulated by cholesterol. Their structures and genomic locations suggest that the $\alpha_2\beta_2$ heterotetrameric structural motif has evolved to enable catalysis of dehydrogenation of steroid- or polycyclic-CoA substrates and that they function in four sub-pathways of cholesterol metabolism.

Cholesterol side chain degradation is proposed to proceed through three β-oxidation cycles. ChsE1-ChsE2 (FadE28-FadE29) has been elucidated to function in the last cycle of β-oxidation through gene knockout studies and biochemical analysis. We identify and assign the substrate

specificities of another two enzymes, ChsE4-ChsE5 (FadE26-FadE27) and ChsE3 (FadE34), that carry out cholesterol side chain oxidation in *Mtb.* Steady-state assays demonstrate that ChsE4-ChsE5 preferentially catalyzes the oxidation of 3-oxo-cholest-4-en-26-oyl CoA in the first cycle of cholesterol side-chain β-oxidation that ultimately yields propionyl-CoA, whereas ChsE3 specifically catalyzes the oxidation of 3-oxo-chol-4-en-24-oyl CoA in the second cycle of β-oxidation that generates acetyl-CoA. However, ChsE4-ChsE5 can catalyze the oxidation of 3-oxo-chol-4-en-24-oyl CoA as well as 3-oxo-4-pregnene-20-carboxyl-CoA. The X-ray crystallographic structure of ChsE4-ChsE5 was determined to a resolution of 2.0 Å and represents the first high-resolution structure of a heterotetrameric acyl-CoA dehydrogenase (ACAD). Unlike typical homotetrameric ACADs that bind four flavin adenine dinucleotide (FAD) cofactors, ChsE4-ChsE5 binds one FAD at each dimer interface, resulting in only two substrate-binding sites rather than the classical four active sites. Comparison of the ChsE4-ChsE5 substrate-binding site to those of known mammalian ACADs reveals an enlarged binding cavity that accommodates steroid substrates and highlights novel prospects for designing inhibitors against the committed β-oxidation step in the first cycle of cholesterol side-chain degradation by *Mtb*.

Besides ACADs, we also characterized the enoyl-CoA hydratase function in the second step of the third cycle of β-oxidation. We solved the first structures of a heterotetrameric MaoC-like enoyl-CoA hydratase, ChsH1-ChsH2, which is encoded by two adjacent genes from the *igr* operon. We demonstrate that ChsH1-ChsH2 catalyzes the hydration of a steroid enoyl-CoA, 3-oxo-4,17-pregnadiene-20-carboxyl-CoA, in the β-oxidation pathway for cholesterol side chain degradation. The ligand-bound and apoenzyme structures of ChsH1-ChsH2^{N} reveal an unusual, modified hot-dog fold with a severely truncated central α-helix that creates an expanded binding

site to accommodate the bulkier steroid ring system. The structures show quaternary structure shifts that accommodate the four rings of the steroid substrate and offer an explanation for why the unusual heterotetrameric assembly is utilized for hydration of this steroid. The unique αβ heterodimer architecture utilized by ChsH1-ChsH2 to bind its distinctive substrate further highlights an opportunity for the development of new anti-mycobacterial drugs that target cholesterol metabolism pathway specific to *Mtb*.

In addition, we investigated post-translational modification as a regulatory mechanism in the cholesterol metabolism pathway. We present a possible N-succinylase that is active with ChsE4-ChsE5 as a substrate, and demonstrate that introduction of a negative charge on lysine 238 reduces the catalytic activity of ChsE4-ChsE5. This modification in combination with other post-translational modifications in the pathway may suppress cholesterol metabolism under hypoxic conditions. We suggest that succinylation is a strategy for dynamically controlling this metabolic pathway to enable *Mtb* to quickly adapt to the changing environment via rapid sensing of the cellular redox status to flexibly alter reaction rates and directions. This demonstration provides us a new perspective to understand the metabolism of *Mtb* and adds an additional level to the complexity of the proteome.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

2xYT	2xyeast tryptone broth
3β-HSD	3β-hydroxysteroid dehydrogenase
9-OH-ADD	9α-hydroxy-androsta-1,4-diene-3,17-dione
ACAD	acyl-CoA dehydrogenase
AD	androst-4-ene-3,17-dione
A. caviea	*Aeromonas caviae*
AcRH	*Aeromonas caviae* MaoC-like enoyl-CoA hydratase
amp	ampicillin
ATP	adenosine triphosphate
cam	chloramphenicol
cfu	colony forming units
CoA	Coenzyme A
C. testosteroni	*Comamonas testosteroni*
CHD	complete hot-dog fold
C-CtRH	C-terminal domain of *Candida tropicalis* MaoC-like enoyl-CoA hydratase
C-HuRH	C-terminal domain of *Homo sapiens* MaoC-like enoyl-CoA hydratase
CtRH	*Candida tropicalis* MaoC-like enoyl-CoA hydratase
3,4-DHSA	3,4-dihydroxy-9,10-seconandrost-1,3,5(10)triene-7,19-dione
4,9-DHSA	4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10), 2-diene-4-oic aicd
E. coli	*Escherichia coli*
Ech	enoyl-CoA hydratase
ESI	electrospray ionization
Fad	fatty acid degrading
FAD	flavin adenine dinucleotide
GD	glutaryl-CoA dehydrogenase
HSA	3-hydroxy-9,10-seconandrost-1,3,5(10)triene-7,19-dione
HEPES	hydroxyethyl piperazine ethanesulfonic acid
His_6	hexahistidine
HIV	human immunodeficiency virus

HPLC	high performance liquid chromatography
HuRH	*homo sapiens* MaoC-like enoyl-CoA hydratase
igr	intracellular growth
IMAC	immobilized-metal ion affinity chromatography
IPTG	ssopropyl β-D-thiogalactoside
iBD	isobutyryl-CoA dehydrogenase
i2VD	iso-2-valeryl-CoA dehydrogenase
i3VD	iso-3-valeryl-CoA dehydrogenase
ICHD	incomplete hot-dog fold
ICL	isocitrate lyase
k_{cat}	Michaelis-Menten catalytic rate constant
K_M	Michaelis-Menten constant
kDa	kilo-Dalton
kan	kanamycin
LB	luria broth
LDL	low density lipoprotein
Ltp	lipid transfer protein
LCAD	long chain acyl-CoA dehydrogenase
MALDI	matrix-assisted laser desorption ionization
Mtb	*Mycobacterium tuberculosis*
MDR-TB	multi-drug resistant *Mycobacterium tuberculosis*
MCAD	medium chain acyl-CoA dehydrogenase
M.smegmatis	*Mycobacterium smegmatis*
MaoC	monoamine oxidase C
MS	mass spectrometry
N-CtRH	N-terminal domain of *Candida tropicalis* MaoC-like enoyl-CoA hydratase
N-HuRH	N-terminal domain of *Homo sapiens* MaoC-like enoyl-CoA hydratase
NMR	nuclear magnetic resonance
OD	optical density
PCR	polymerase chain reaction
PDB	protein data bank

PDIM	phthiocerol dimycocerosate
PCR	polymerase chain reaction
R.equi	*Rhodococcus equi*
R.jostii	*Rhodococcus jostii*
rpm	revolutions per minute
rt	room temperature
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SCAD	short chain acyl-CoA dehydrogenase
TAPS	n-tri(hydroxylmethyl)methyl-3-aminopropanesulfonic acid
TB	Tuberculosis
TDR-TB	totally-drug resistant *Mycobacterium tuberculosis*
TAPS	n-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
TOF	time of flight
UV	ultraviolet
Vi	initial velocity
VLCAD	very long chain acyl-CoA dehydrogenase
WT	wild type
XDR-TB	extremely-drug resistant *Mycobacterium tuberculosis*

Chapter I. Introduction

1.1 Tuberculosis disease and treatment

Tuberculosis (TB) is one of the world's deadliest infectious diseases; about 9 million people developed TB and 1.5 million people died from TB in 2013. TB is the leading direct cause of HIV deaths and there are about one million people suffering from HIV/TB co-infection.[1] TB is caused by the pathogen called *Mycobacterium tuberculosis (Mtb)*. Existing drugs only target a small fraction of essential growth pathways, and it is not difficult for *Mtb* to develop resistance by activating compensatory pathways. Among the new TB cases, 3.5 % of them are infected with multidrug-resistant TB (MDR-TB). Even though TB is slowly declining every year, the high mortality of MDR-TB, extensively drug-resistant TB (XDR-TB) and totally drug-resistant TB (TDR-TB) are seriously threating people's lives all over the world.[2]

Table 1-1. TB drugs: First line drugs, second line drugs and drugs on trial.

Drugs	Lines	Developed year	Pathway targetted
Isoniazid[3]	First	1952	Cell wall synthesis
Rifampicin[4]	First	1963	RNA synthesis
Pyrazinamide[5]	First	1954	Protein synthesis
Ethambutol[6]	First	1962	Cell wall synthesis
Streptomycin[7]	First	1944	Protein synthesis
Ethionamide[8]	Second	1956	Cell wall synthesis
Kanamycin[9]	Second	1957	Protein synthesis
Capreomycin[10]	Second	1960	Protein synthesis
Levofloxacin[11]	Second	1987	DNA amplification
Moxifloxacin[12]	Second	1999	DNA amplification
Ofloxacin[13]	Second	1990	DNA amplification
Cycloserine[14]	Second	1972	Cell wall synthesis
Terizidone[15]	Second	1991	Cell wall synthesis
Protionamide[16]	Second	2006	Cell wall synthesis
Bedaquiline[17]	Phase III	2012	ATP synthesis
SQ-109[18]	Phase II	2007	Cell wall synthesis
OPC-67683[19]	Phase III	2006	Cell wall synthesis
PA-824[20]	Phase II	2009	Cell wall synthesis

AZD5847[21]	Phase II	2012	Protein synthesis

Treatment of TB is much more complicated than for other bacterial infections. A two-month treatment using a combination of all the five first line drugs (Table 1-1), followed by a four-month treatment using rifampicin and isoniazid is the well-accepted standard protocol to treat TB patients with non-resistant *Mtb* infection who are HIV-negative.[22] Treatment of MDR-TB requires at least 20 months by including treatment with second line drugs (Table 1-1) that are toxic, poorly tolerated and less effective.[23] Bedaquiline, approved at the end of 2012 for the treatment of adult MDR-TB, targets ATP synthetase and depletes cellular energy storage. Bedaquiline works via a different mechanism relative to other available drugs that target cell wall, DNA, or protein synthesis, so it retains its activity better against *Mtb* that has acquired resistance for other TB drugs.[24] However, its severe adverse effects limit its use to when other effective treatments are not available. No effective treatments exist yet for XDR-TB and TDR-TB. New drugs based on novel pathways or targets that can shorten the therapy, treat MDR & XDR-TB, and eliminate adverse side effects are urgently needed.

1.2 *Mtb* infection process and its living environment

Mtb is transmitted through aerosol to a new host. As an intracellular pathogen, *Mtb* travels to the alveolus of the lung where it encounters alveolar macrophages. Macrophages are responsible for defending against invasions by foreign pathogens. Typically, when a pathogen is engulfed by a macrophage, it is trapped in a phagosome, which then fuses with a lysosome and digestion occurs. However, *Mtb* can reside in macrophages and replicate within them by actively preventing the phagosome-lysosome fusion. It is believed that *Mtb* will replicate unrestrictedly for a period by migrating to the local lymph nodes and establishing a systemic infection. After

this period, *Mtb* replication will be controlled by cellular immunity.[25] The success of *Mtb* as a pathogen is due to the fact that it has developed multiple strategies to counteract attacks from host macrophages. As one of *Mtb*'s tricks to hide from the host's immune system, *Mtb* enters a non-growing, but metabolically active status. This change could potentially act as a preventive mechanism to decrease *Mtb*'s nutritional requirements and to provide a long-term state of tolerance of multiple stresses.[26, 27] The result is a chronic infectious condition that can remain unnoticed for decades. Macrophages constitute the largest cellular reservoir of *Mtb*'s chronic infection. More recent studies showed that *Mtb* could even escape the trapping from phagosome and reside in cytosol,[28] which is another *Mtb*'s strategy to survive *in vivo*.

A localized inflammatory response will be triggered by *Mtb* infected alveolar macrophages, resulting in the release of tumor necrosis factor (TNF)-α, IFN-γ and inflammatory chemokines. The initiated immune response induces the formation of a granuloma, which is the pathological hallmark of tuberculosis.[29] Interestingly, the granuloma differentiates over time, generating multinucleated giant cells and foamy macrophages, and forming into a multicellular structure.[30] As granuloma formation progresses, the number of foamy macrophages increases and the hypoxia increases.[29] Viable *Mtb* in a fully developed granuloma can be released into the airway through sputum aerosol, and transmitted to a new host to start a new life cycle.

1.3 Nutrition sources of *Mtb* when residing in macrophages

Mtb infected macrophages form a foamy phenotype that is characterized by accumulated lipid bodies (LBs).[31] Virulent species such as *Mtb* and *M. avium* can actively induce foamy macrophage formation when they infect pheripheral blood mononuclear cells (PBMCs) *in vitro*, whereas, non-virulent strains such as *M. smegmatis* cannot, suggesting foamy macrophages may

be important for *Mtb*'s survival *in vivo*.[32] Inside of the foamy macrophage, *Mtb*-containing phagosomes migrate towards accumulated LBs over time, and then engulf lipid bodies. Even though the mechanism of this process has not been established yet, free *Mtb* will be transferred into lipid bodies following the engulfment process.[32] Microscopic analysis also revealed accumulated LBs inside of the pathogen. After its translocation into LBs, *Mtb* will slow down its replication and start persisting in the host, suggesting a potential biochemical connectivity between the pathogen and LBs.

A large fraction of *Mtb* bacteria isolated from TB patients' sputum samples contain LBs, which is consistent with non-replicating *Mtb* bacteria observed in hypoxic environment.[33] *Mtb* isolated from the lungs of infected animals and macrophages preferentially metabolizes fatty acids rather than carbohydrates,[34] which is a distinct metabolic state compared to typical *in vitro* cultures.

Among a total of 4000 *Mtb* genes identified through genome sequencing in 1998, at least 250 genes are potentially involved in lipid metabolism,[35] which is 5 times more lipid metabolism genes than *E. coli*. Many of the lipid-metabolism genes are induced during intracellular growth and infection.[35] Foamy macrophages play an essential role during the dormant phase of infection,[32] suggesting lipid metabolism is essential for the virulence of *Mtb in vivo*. Biochemical analysis of the major species of lipid droplets within *Mtb* infected macrophages has revealed a high abundance of cholesterol, cholesterol ester and triacylglycerides (TAG).[31]

β-oxidation and glyoxylate shunt are the two major pathways that are applied by bacteria in order to use fatty acid. Isocitrate lyases (ICLs) are involved in both the glyoxylate shunt pathway and the methylcitrate cycle (Figure 1-1). The expression of ICL1 (Rv0467) is upregulated during

infection of macrophages by *Mtb*. Δ*icl* strain of *Mtb* showed attenuated growth during its chronic phase of infection in immune-competent mice, while failed to affect its stationary phase survival *in vitro* with 7H9 broth. These observations suggestted ICL is essential for *Mtb in vivo*.[36] Deletion of both *icl1* and *icl2* genes eliminates *Mtb*'s ability to grow on fatty acid instead of carbohydrates.[37] Together, these experiments implicate that fatty acid metabolism is essential to *Mtb in vivo*.

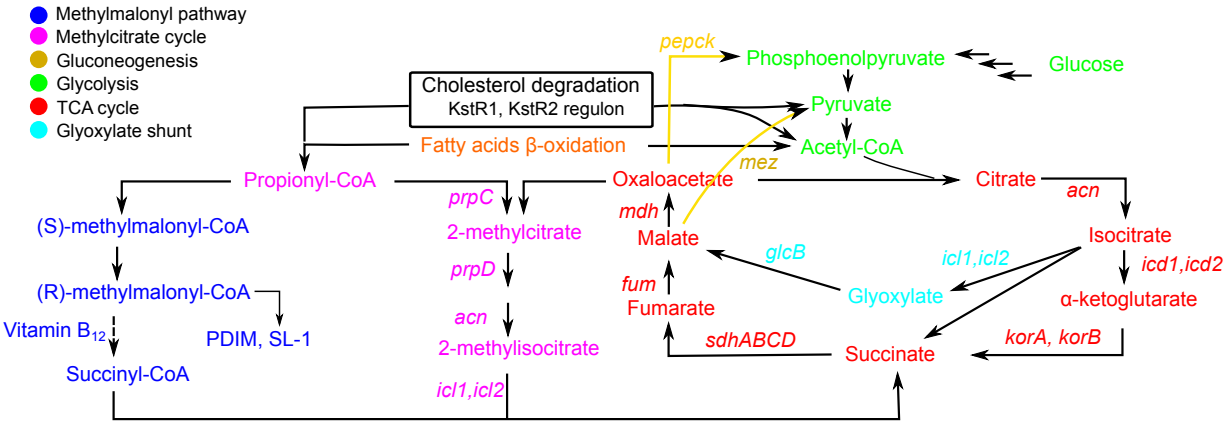


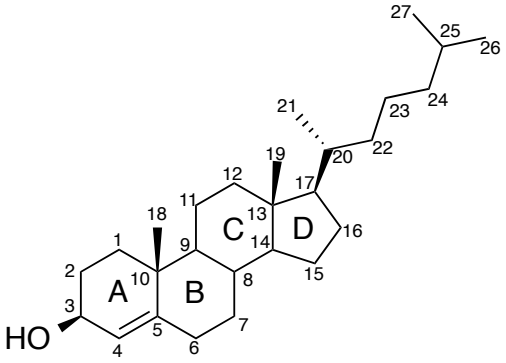
Figure 1-1. Central carbon flow network in *Mtb*. Cholesterol metabolism (black) generates pyruvate, acetyl-CoA, and propionyl-CoA. Pyruvate and acetyl-CoA feed into the TCA cycle (red) and propionyl-CoA feed into the methylcitrate cycle (purple) and the methylmalonyl pathway (blue).

Phosphoenolpyruvate carboxykinase (PEPCK), catalyzing the first committed step in gluconeogenesis, is required for *Mtb*'s growth on fatty acids. *Mtb* lacking PEPCK failed to replicate and survive during both the acute and chronic phases of infection in mouse.[38] Thus, gluconeogenesis is essential for *Mtb* throughout its infection.

1.4 The significance of cholesterol utilization and metabolism to *Mtb*

Cholesterol is a major component of LBs that accumulate in foamy macrophages. *Mtb* possesses a sophisticated cholesterol uptake system and a complete cholesterol degradation

pathway in order to utilize it.[39] In addition, *Mtb* can utilize cholesterol as a sole carbon source in *vitro* for replication.[40] Many steroid-utilizing genes have been shown to be critical *in vivo*.[41] In the process of identifying *Mtb* genes that are required for its growth in macrophages, about 100 out of the total identified 126 genes are annotated to be involved in β-oxidation and fatty acid degradation.[42] As a structurally odd lipid, the side chain and part of the ring system of cholesterol have been proposed to go through β-oxidation to be degraded.



Scheme 1-1. Cholesterol numbering scheme.

The Mce4 system is a steroid uptake sytem in Rhodococcus *jostti* RHA1. Bioinformatic analysis revealed that all *mce4* loci are linked to cholesterol metabolism.[43] *Mtb* lacking the *mce4* operon displayed a severe restriction of cholesterol-dependent growth, while no influence was observed when growing in media supplemented with glycerol, suggesting Mce4 is required for cholesterol utilization. *Mtb*: Δ *mce4* mutant displayed a growth and survival defect during chronic infection in the lung, which suggests that the host cholesterol is critical for *Mtb*'s chronic infection.[40, 44]

The cholesterol degradation pathway in mycobacteria is apparently controlled by at least two highly conserved TetR-type transcriptional repressors: KstR1 (Rv3574) and KstR2 (Rv3557c).[45]

KstR1 is induced by 3-oxo-4-cholestenoic acid, one of the first cholesterol metabolic intermediates in *Mtb*, and controls the expression of 83 cholesterol catabolic genes.[46] KstR2[47] is induced by 3aα-H-4α(3'-propanoate)-7aβ-methylhexahydro-1,5-indanedione-CoA (HIP-CoA), a C&D ring intact cholesterol intermediate, and controls the expression of 15 cholesterol catabolic genes (Figure 1-2).

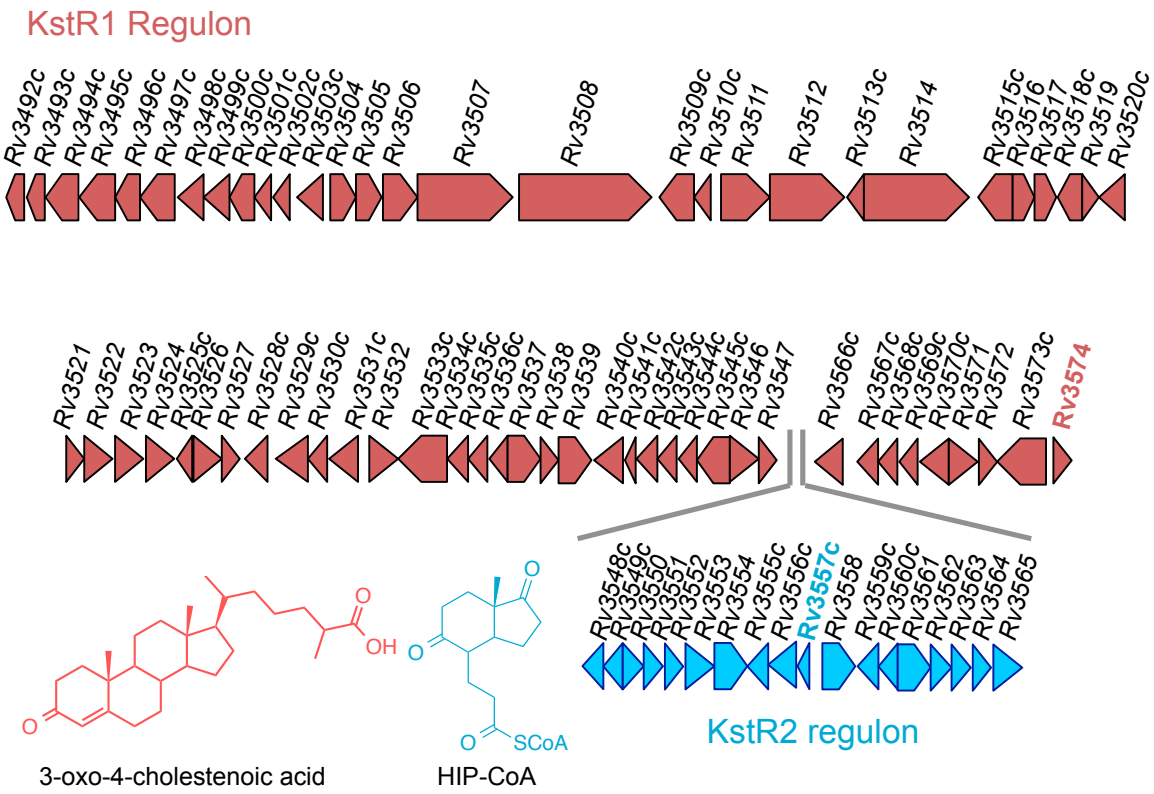


Figure 1-2. KstR1 and KstR2 in the *Mtb* geonome. KstR1 regulated genes are colored in red, and the inducer of KstR1 (Rv3574) is 3-oxo-4-cholestenoic acid. KstR2 regulated genes are colored in blue and the inducer of KstR2 (Rv3557c) is HIP-CoA.

The *igr* operon (*Rv3440c-Rv3545c*) is important for *Mtb*'s growth in both macrophages and in mice.[48] *Mtb*: Δ*igr* mutant lost its ability to grow on cholesterol by accumulating toxic metabolites when growing on cholesterol, meanwhile, this phenotype can be relieved by further knocking out the cholesteol importing system encoded by the *mce4* operon. In addition, *Mtb*: Δ*igr* mutant displayed attenuated growth during the early phase of infection in mice.[48, 49]

Recently, the cholesterol-derived metabolome of *Mtb*: Δ*igr* mutant was analyzed by employing isotopically labeled low-density lipoprotein (LDL) containing either [1, 7, 15, 22, 26-^{14}C] cholesterol or [1, 7, 15, 22, 26-^{13}C] cholesterol. The results showed that gene products of the *igr* operon are required for complete degradation of the 2'-propanoate side chain of cholesterol.[50]

The gene of *fadA5* (*Rv3546*) is adjacent to the *igr* operon, and is essential for *Mtb*'s growth on cholesterol as a sole carbon source. A *Mtb*: Δ*fadA5* mutant showed attenuated growth during the chronic phase of infection in mice, and this attenuation was not due to accumulation of toxic metabolites.[51] Therefore, *fadA5* is required for virulence in the mouse model of infection. Recently, the function of FadA5 was elucidated biochemically and structurally. FadA5 functions in the cholesterol degradation pathway by catalyzing the cleavage of 3,22-dioxo-chol-4-ene-24-oyl-CoA, a cholesterol metabolite, generating 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA) and acetyl-CoA. This function assignment further solidified the essential role of cholesterol metabolism-related genes in *Mtb*.[52]

1.5 Elucidating cholesterol metabolism pathway in *Mtb*

The cholesterol degradation pathway can be divided into ring degradation and side chain degradation. 3β-hydroxylsteroid dehydrogenase (3β-HSD, Rv1106c) is the first enzyme in this pathway by catalyzing the conversion of cholesterol to cholest-4-en-3-one (Scheme 1-2). 3β-HSD is required for the growth of *Mtb* on cholesterol as a sole carbon source, but it is not required for growth in activated macrophage or in the guinea pig model.[53]

Cyp125 (Rv3545c) is a cytochrome P450 enzyme that catalyzes the terminal hydroxylation of C27 cholesterol. Its location in the *igr* operon highlights its imporatence for *Mtb*'s pathogenesis. *Mtb* H37Rv strain lacking the *cyp125* gene still grew normally on cholesterol,

suggesting that H37Rv may contain another gene that can compensate the function of Cyp125. In contrast, when the CDC1551**:** Δ*cyp125* strain was grown on cholesterol, cholest-4-en-3-one accumulated. The accumulation of cholest-4-en-3-one demonstrated that cholesterol ring degradation can not proceed with its 8-carbon side chain intact.[54] Based on other characterized P450 enzymes, further oxidation is very likely to occur after the transformation of cholesterol into 27-hydroxyl-cholest-4-en-3-one to generate 3-oxo-cholest-4-en-26-oic acid (Scheme 1-2).

Table 1-2. Characterized genes involved in cholesterol metabolism pathway.

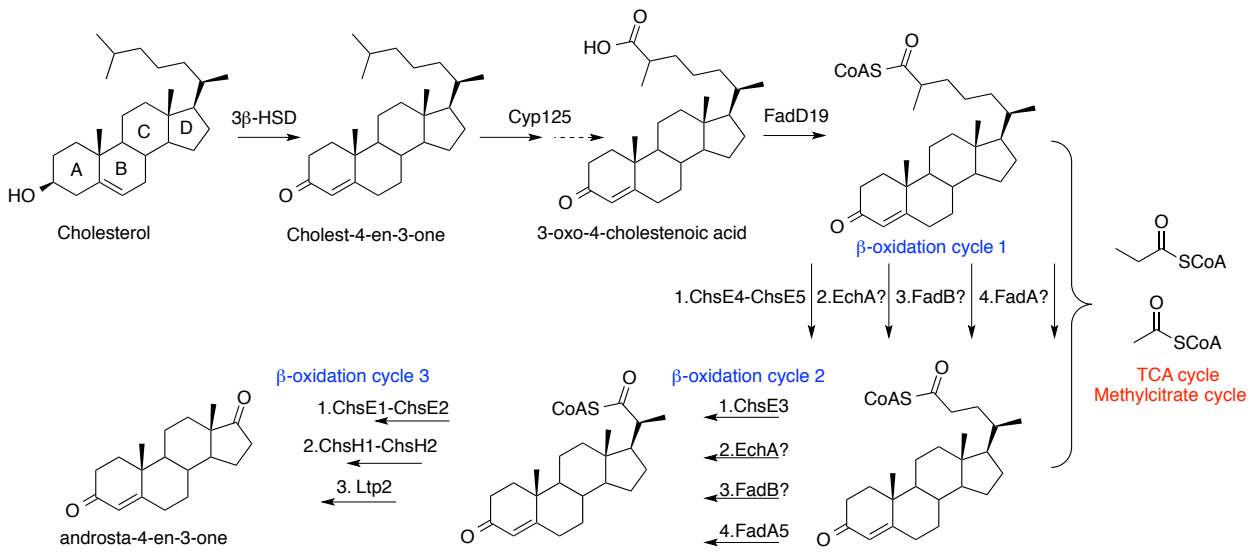
Rv number	Gene product	Function	Regulator
Rv1106c	3β-HSD	3β-hydroxysteroid dehydrogenase	KstR1(only in *M.smeg*)
Rv3494c	Mce4F	Cholesterol uptake	KstR1
Rv3495c	Mce4E	Cholesterol uptake	KstR1
Rv3496c	Mce4D	Cholesterol uptake	KstR1
Rv3497c	Mce4C	Cholesterol uptake	KstR1
Rv3498c	Mce4B	Cholesterol uptake	KstR1
Rv3499c	Mce4A	Cholesterol uptake	KstR1
Rv3504	ChsE4	Acyl-CoA dehydrogenase α	KstR1
Rv3505	ChsE5	Acyl-CoA dehydrogenase β	KstR1
Rv3506	FadD17	Acyl-CoA synthetase	KstR1
Rv3515c	FadD19	Acyl-CoA synthetase	KstR1
Rv3518c	Cyp142	Cytochrome P450 monooxygenase	KstR1
Rv3526	KshA	3-ketosteroid 9α-hydroxylase-reductase	KstR1
Rv3571	KshB	3-ketosteroid 9α-hydroxylase-oxygenase	KstR1
Rv3534c	HsaF	Probable 4-hydroxy-2-oxovalerate aldolase	KstR1
Rv3535c	HsaG	Probable acetaldehyde dehydrogenase	KstR1
Rv3536c	HsaE	Probable hydratase	KstR1
Rv3537	KstD	3-keto-Δ4-steroid-Δ1-dehydrogenase	KstR1
Rv3540c	Ltp2	Lipid transfer protein	KstR1
Rv3541c	ChsH1	Enoyl-CoA hydratase α	KstR1
Rv3542c	ChsH2	Enoyl-CoA hydratase β	KstR1
Rv3543c	ChsE2	Acyl-CoA dehydrogenase β	KstR1
Rv3544c	ChsE1	Acyl-CoA dehydrogenase α	KstR1
Rv3545c	Cyp125	Cytochrome P450 enzyme	KstR1
Rv3546	FadA5	Thiolase	KstR1
Rv3560c	FadE30	Acyl-CoA dehydrogenase	KstR2
Rv3561	FadD3	Acyl-CoA synthetase	KstR2
Rv3567c	HsaB	Monooxygenase- reducatase component	KstR1
Rv3568c	HsaC	Iron-dependent extradiol dioxygenase.	KstR1
Rv3569c	HsaD	4,9-DHSA hydrolase	KstR1
Rv3570c	HsaA	Monoxoygenase- oxygenase component	KstR1

Acyl-Coenzyme A (CoA) synthetases catalyze the ligation of CoA onto the fatty acid, forming CoA thioester. FadD19 (Rv3515c), regulated by KstR1 regulon, is responsible for transforming 3-oxo-4-cholesten-26-oate into 3-oxo-4-cholesten-26-oyl CoA (Scheme 1-2).[55]

Typically, β-oxidation in mitochondria comprises four sequential enzyme-catalyzed reactions to generate acetyl-CoA, which is fed into the citric acid cycle. Acyl-CoA dehydrogenases (ACADs, or FadEs in bacteria) catalyze the formation of a double bond between C-2 and C-3 of an acyl-CoA substrate, generating an enoyl-CoA. This reaction requires flavin adenine dinucleotide (FAD) cofactor as the electron acceptor. The next step is the hydration of the double bond to form 3-(S)-β-hydroxyl-acyl-CoA, and is catalyzed by an enoyl-CoA hydatase (EchA). This reaction is stereospecific, forming only the S isomer in mitochondrial β-oxidation. The third step is the oxidation of 3-(*S*)- β-hydroxyl-acyl-CoAs into a 3-keto-acyl-CoA with NAD^+ as the electron acceptor and is catalyzed by a 3-(*S*)-β-hydroxyl-acyl-CoA dehydrogenase (FadB). The final step is the carbon-carbon cleavage in a reverse claisen condensation reaction catalyzed by a thiolase that generates acetyl-CoA (FadA) and a shorter acyl-CoA (Scheme 1-3).

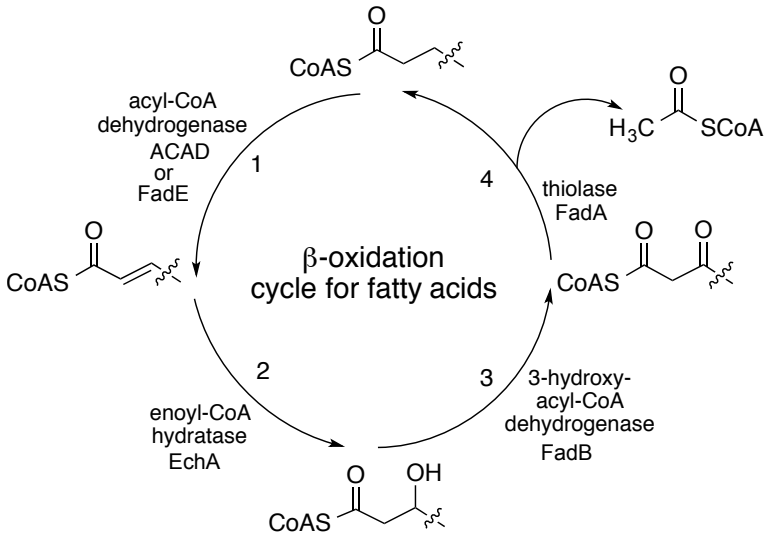
The cholesterol 8-carbon side chain requires three β-oxidation cycles to be fully degraded. As aforementioned, the gene products of the *igr* operon are responsible for degradation of the 2'-propanoate side chain of cholesterol in the last cycle of β-oxidation.[50, 56] However, until the work in this thesis, enzymes involved in the other two β-oxidation cycles remained unassigned.

The novel obligate heteromeric assembly and substrate preference of ChsE1 (Rv3543c)-ChsE2 (Rv3544c) shed light on the structural requirements for the other two ACADs in cholesterol side chain degradation. Rv3541c and Rv3542c from the *igr* operon are both annotated as (*R*)-specific enoyl-CoA hydratases, and thus also form an obligate protein complex.[50] In this thesis, its function as a hydratase was established. As the most poorly characterized protein, Ltp2 (Rv3540c) remains the most mysterious enzyme both structurally and biochemically (Scheme 1-2).



Scheme 1-2. The proposed cholesterol side chain degradation pathway. Identified gene products are labeled corresponding to their catalysis steps. Three β-oxidation cycles are predicted to fully degrade the 8-carbon side chain. The generated acetyl-CoA and propionyl-CoA enter the TCA cycle and methylcitrate cycle, respectively.

As aforementioned, 3β-HSD catalyzes the conversion of cholesterol to cholest-4-en-3-one, thus initiating cholesterol ring degradation. Fully degradation of cholesterol A&B ring requires catalysis of nine enzymes, KstD (Rv3537), KshA/KshB (Rv3526/Rv3571), HsaA (Rv3567c), HsaB (Rv3568c), HsaC (Rv3569c), HsaD (Rv3570c), HsaE (Rv3534c), HsaF (Rv3535c), and HsaG (Rv3536c). Except acetyl-CoA and pyruvate, the final product of these reactions is a C&D ring intact metabolite intermediate, 3aα-H-4α(3'-propanoate)-7aβ-methylhexahydro-1,5-indanedione (HIP).



Scheme 1-3. Typical β-oxidation for fatty acid degradation in mitochondria and bacteria.

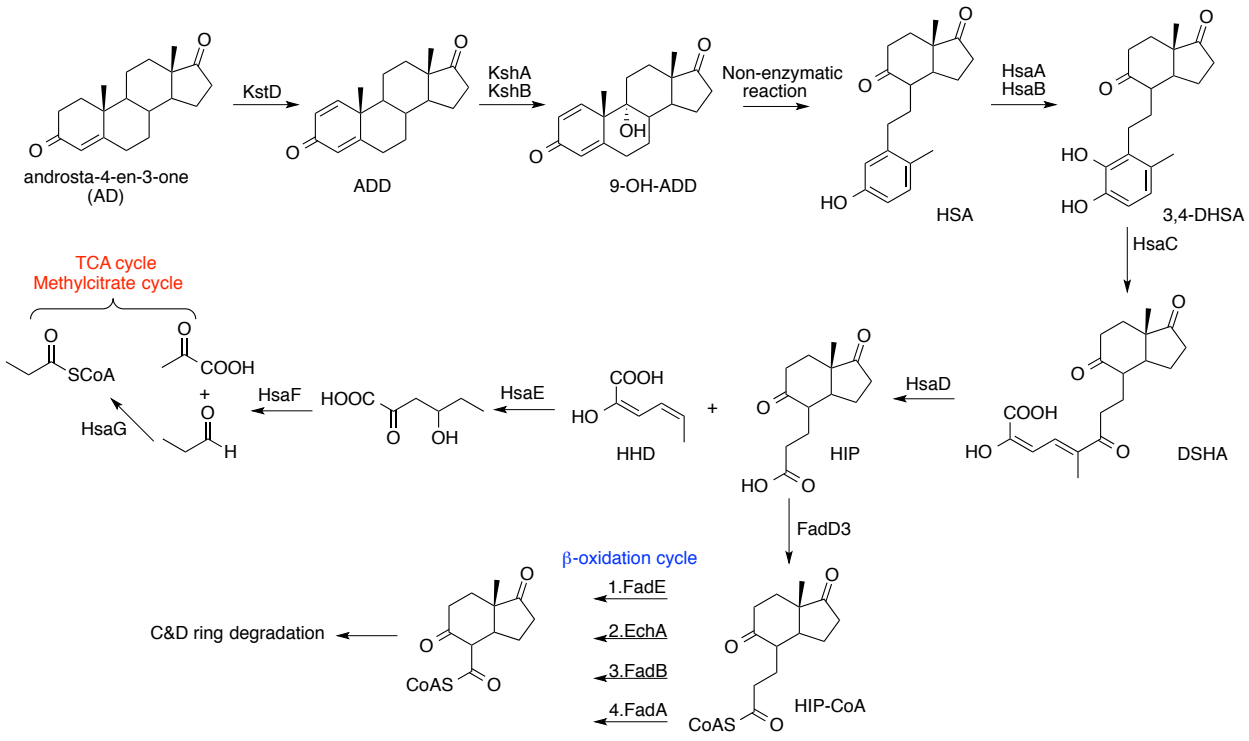
KstD (Rv3537) is the only 3-keto-Δ4-steroid Δ1-dehydrogenase present in the *Mtb* genome, highlighting its specificity and importance in cholesterol metabolism. KstD catalyzes the dehydrogenation of 4-androstene-3,17-dione (4-AD) between C1 and C2 position to form 1,4-androstadiene-3,17-dione (ADD) (Scheme 1-4).[57] The *Mtb*:Δ*kstD* strain has a limited ability to replicate in resting macrophages following infection, which reflect that the *Mtb*:Δ*kstD* strain cannot inhibit the TLR2-dependent bactericidal activity of resting macrophages.[58]

KshA/KshB (3-ketosteroid 9α-hydroxylase) is a two-component Rieske oxygenase (RO) comprising a reductase (Rv3571/KshB) and an oxygenase (Rv3526/KshA). KshAB catalyzes the 9α-hydroxylation of 1,4-androstadiene-3,17-dione (ADD) to 9α-hydroxy-9,10-seconandrost-1,3,5-triene-9,17-dione (HSA),[59] generating a B-ring cleaved cholesterol intermediate (Scheme 1-4). *In vitro* enzymatic analysis of KshAB demonstrated that the catalytic specificities of KshAB towards CoA thioesters was 20-30 times higher than the corresponding 17-keto compounds (ADD), suggesting that side chain degradation and ring degradation occur simultaneously.[60]

HsaA (Rv3570c) and HsaB (Rv3567c) encode an oxygenase and a reductase, respectively. They comprise a monooxygenase with FAD bound. HsaAB catalyzes the conversion of 3-HSA to 3,4-DHSA (Scheme 1-4). RHA017:Δ*hsaA* mutant strain lost the ability to grow on cholesterol, which is consistent with the essential role of HsaA in cholesterol metabolism in Rhodococcus. When grown in pyruvate and supplemented with cholesterol, the RHA017:Δ*hsaA* mutant strain accumulated 3-HSA and other related metabolites.[61]

HsaC (Rv3568c) is an iron-dependent extradiol dioxygenase. *In vitro* enzymatic tests showed that HsaC preferred 3, 4-DHSA as the substrate relative to other non-steroid catechols, meaning that HsaC functioned specifically in cholesterol metabolism. In the presence of oxygen, HsaC catalyzes extradiol ring-cleavege of DHSA to DSHA (3,4-dihydroxy-9,10-seconandrost-1,3,5(1 0)-triene-9,17-dione) (Scheme 1-4). Immuno-compromised mice infected with a H37Rv:Δ*hsaC* mutant displayed a longer survival time than mice infected with the wild-type strain. This experiment demonstrated the importance of *hsaC* in *Mtb* pathogenesis.[62]

Subsequent to HsaC catalysis, HsaD (Rv3569c; 2-hydroxy-6-oxo-6-phenylhexa-2, 4-dienoate hydrolase) catalyzes the hydrolysis of DSHA (Scheme 1-4).[63] The products are 3aα-H-4α(3'-propanoate)-7aβ-methylhexahydro-1,5-indanedione (HIP), a C&D ring intact cholesterol intermediate, and 2-hydroxy-hexa-2,4-dienoic acid (HHD). The crystal structure of HsaD revealed that the size of the substrate binding cleft is more than twice as large as its homologs involved in the catabolism of aromatic compounds. This pocket size suggested that HsaD functions specifically in cholesterol degradation pathway.[64]



Scheme 1-4. The proposed cholesterol A &B ring degradation pathway. Identified gene products are labeled corresponding to their catalysis steps. The genes that are involved in C and D ring degradation have not been reported yet.

HHD can be further degraded. Based on the protein similarities and phylogenetic relations between *Mtb* and *C.testosteroni*, HsaE/HsaF/HsaG (Rv3534c/Rv3535c/Rv3636c) are homologous to TesE/TesF/TesG, which have been demonstrated to catalyze the degradation of HHD in *C. testosterone*.[65] HsaE, similar to tesE, catalyzes the hydration of HHD, forming 4-hydroxy-2-oxo-hexanoic acid. HsaF catalyzes the cleavage of 4-hydroxy-2-oxo-hexanoic acid, generating pyruvate and propionaldehyde. However, HsaF is inactive by itself, and the formation of a complex with HsaG appears to be required in order to induce a catalycially active conformation. Without being exported to the solvent, pyruvate and propionaldehyde were channelled directly to the dehydrogenase, HsaG, where propionaldehyde was transformed into propionyl-CoA in an NAD^+ and CoA dependent manner (Scheme 1-4).[66] HsaF-HsaG folds as a

heterotetramer with each HsaG chain bound to the periphery of the HsaF dimer, forming a tunnel running from the active site of HsaF to the active site of HsaG.[66]

The RHA1:Δ*fadD3* (*Rv3561*) mutant displayed attenuated growth when grown on cholesterol. Metabolite analysis of this mutant revealed the accumulation of HIP. *In vitro* biochemical characterizations confirmed that FadD3 initiates the catabolism of cholesterol C &D ring by ligating CoA onto HIP. The formation of HIP-CoA suggests that β-oxidation is needed to degrade the propionate moiety of HIP.

The catabolism of the C&D rings still remains the least characterized. Transcriptional analysis revealed that KstR2 regulated genes are upregulated when *Mtb* was grown on HIP, whereas the transcription level of KstR1 regulated genes did not change much.[47] Genes involved in cholesterol C&D ring degradation are predicted to be regulated by the KstR2 regulon, which is consistent with the regulation of *fadD3*. FadE30 (Rv3560c) and EchA20 (Rv3550) are the β-oxidation enzymes under this regulon and are probably responsible for degrading the side chain of HIP. *Rhodococcus equi* RE1:*ΔfadE30* was impaired during the growth on 4-OH-HIP and displayed an attenuated phenotype in macrophage infection assays. The results showed *fadE30* played an essential role for HIP degradation.[67] Further characterizations of the biochemical functions of FadE30 and EchA20 are needed.

The three *fadE* genes, *fadE31 (Rv3562), fadE32 (Rv3563), and fadE33* (*Rv3564*), are also under the regulation of KstR2.[39] Their gene products have been proposed to be involved in C&D ring degradation. Similar to the novel $\alpha_2\beta_2$ heterotetrameric ChsE1-ChsE2,[56] FadE31-FadE32 and FadE31-FadE33 have also been characterized as $\alpha_2\beta_2$ heterotetrameric ACADs,[68] further

suggesting that they possess the ability to take steroid substrates. However, their functions require further biochemical characterizations.

1.6 Cholesterol catabolism and other central carbon catabolic pathways in *Mtb*

Instead of relying on cholesterol as its sole carbon source *in vivo*,[53] *Mtb* assimilates multiple carbon sources simultaneously.[69] The metabolized products from cholesterol are propionyl-CoA, acetyl-CoA and pyruvate (Figure 1-2). Propionyl-CoA is toxic to *Mtb* and as the level of propionyl-CoA increases, it can enter the methylmalonyl pathway (MMP) and methylcitrate cycle (MCC). Importantly, as the common precursor of two *Mtb* virulent lipid factors, phthiocerol dimycocerosate (PDIM) and sulfolipid-1 (SL-1), the availability of methymalonyl-CoA controls their chain length and abundance.[70]

Isocitrate lyase (ICL)-deficient *Mtb* showed attenuated growth on propionate due to the defective methylcitrate cycle (MCC), which resulted in an unavailability of succinate and accumulation of other toxic metabolic intermediates. Vitamin B12 can rescue the growth defect by redirecting propionyl-CoA into the MMP pathway to anaplerotically supply succinate to the TCA cycle.[71] Therefore, MMP and MCC are two of the routes that *Mtb* utilizes to release the propionyl-CoA stress generated upon cholesterol metabolism (Figure 1-2).

The pyruvate and acetyl-CoA pool are also enlarged by cholesterol degradation and are fed into the TCA cycle or glyoxylate shunt in *Mtb* (Figure 1-2). As the only precursor for TCA cycle and the major substrate for mycolic acids biosynthesis,[72] the availability of acetyl-CoA directly influences *Mtb*'s energy status and the availability of building blocks. The MCC and MMP have the potential to generate more acetyl-CoA by merging with the TCA cycle and the later gluconeogenesis pathway. Increased acetyl-CoA can feed back into the TCA or glyoxylate

shunt to generate more oxaloacetate, which will be used in the MCC, thereby influencing the propionyl level. Interplay between lipid metabolism, MCC, MMP, and TCA constitutes a highly coordinated metabolic network that is essential for *Mtb* survival and virulence.

1.7 The uniqueness of cholesterol utilization in *Mtb* relative to the host

Cholesterol has indispensible structural and regulatory roles in humans. As a major component of animal plasma membranes, the presence of cholesterol influences membrane fluidity. In addition, cholesterol is abundant in blood plasma in the form of lipoproteins, where the majority of it is esterified to form cholesteryl esters. It has been recognized that dysfunction of lipoproteins relates to atherosclerosis.[73] In addition, cholesterol acts as the precursor of steroid hormones, regulating sexual development and individual metabolism. However, no evidence suggests that mammals use cholesterol as an energy source and enzymes that degrade the core structure (gonane) of steroids are not present in humans. The ability of *Mtb* to metabolize steroids like cholesterol, and the role that these compounds play in the virulence and pathogenesis of this organism makes the enzymes that perform these functions attractive areas of study for future drug development.

Although humans do not catabolize cholesterol, they have the ability to use fatty acids as energy. The β-oxidation of fatty acids provides a major energy source when blood glucose levels are low. β-oxidation in humans has received increasing attention because it is associated with numerous genetic deficiencies.[74] β-oxidation happens either in mitochondria or peroxisomes, where the fatty acid chains are too long to be handled by the mitochondria. In mitochondria, a group of acyl-CoA dehydrogenases (ACADs) are responsible for the dehydrogenation of acyl-CoAs with different but overlapping substrate specificities. Therefore, all the human ACADs

were named as short chain, medium chain, long chain and very long chain acyl-CoA dehydrogenases based on the chain lengths of their substrates specificities. Crotonase is the enoyl-CoA hydratase that catalyzes the double bond hydration, forming only the 3β-(*S*)-hydroxyl-acyl-CoA in mitochondrial β-oxidation.[75] In peroxisomes, the CoA esters of α-methyl-branched acyl-CoAs and the bile acid intermediates, which also possess a α-methyl substitution in their side chain, are shortened via branched chain acyl-CoA oxidase, multifunctional protein-2, and SCPx.[76] Since cholesterol is an oddly shaped lipid and its degradation requires at least 4 cycles of β-oxidations (Scheme 1-2 and Scheme 1-4), it is interesting to compare the β-oxidation enzymes involved in cholesterol metabolism in *Mtb* to their mammalian homologous enzymes to identify their unique features.

1.8 Specific aims

Knowing the functions of genes and their corresponding products involved in the cholesterol metabolism pathway in *Mtb* is critical for understanding *Mtb* infection biology, and may provide new targets to develop anti-TB drugs. Even though the *Mtb* genome sequence is completed, homology-based comparisons have failed to suggest functions for nearly 40 % of the annotated *Mtb* genes. Up to 30% of detected enzymatic activities have not been assigned to a known gene.[77, 78] In the cholesterol metabolism pathway, genes involved in cholesterol side chain degradation and C&D ring degradation have not been identified yet. Fully understanding chemical steps in cholesterol metabolism pathway in *Mtb* is necessary to interpret *Mtb*'s virulence and pathogenesis as well as develop new therapeutics to treat TB. Inspiringly, inhibitors have been discovered to target HsaAB enzymes, which function in cholesterol A/B ring degradation. The inhibition of HsaAB limited *Mtb* replication in macrophages,[79] which

further confirmed the essentiality of cholesterol metabolism for *Mtb* and validated the possibility to develop novel therapeutics based on cholesterol metabolism.

The goal of this thesis is to elucidate cholesterol metabolism in *Mtb*. Molecular biology techniques were used to clone, express and purify potential enzymes involved in cholesterol degradation pathways. Biochemical techniques were applied to characterize enzyme activities and other biophysical features. X-ray crystallography was used to characterize enzymes structures at the molecular level. Western blot analysis was used to detect the post-translational modifications of the enzymes.

The specific aims include:

1. **Identify and characterize cholesterol regulated heteromeric FadEs.** ChsE1-ChsE2 (formerly, FadE28-Fad29) has been reported as a heterotetrameric ACAD to specifically function in cholesterol side chain metabolism. Our hypothesis is that this novel heteromeric architecture has evolved to take steroid substrates. Bioinformatic analysis of the FadE proteome from *Mtb* identified 5 other FadE pairs that form heteromeric ACADs. In addition, all of those FadE pairs are under the regulation of cholesterol and their homologs only exist in bacteria strains that can metabolize steroids. This discovery provided a key step to assign ACADs that function in cholesterol degradation in *Mtb*.
2. **Elucidate ACADs that function in cholesterol side chain degradation structurally and biochemically.** *chsE4* (*fadE26*), *chsE5* (*fadE27*), *chsE3* (*fadE34*) , *chsE1* (*fadE28*) and *chsE2* (*fadE29*) are the five *acads* regulated by KstR1. Genes regulated by KstR1 are suggested to be involved in cholesterol side chain and A&B ring degradation. We demonstrated ChE4-ChsE5, ChsE3, and ChsE1-ChsE2 are the three ACADs that function

in the three β-oxidaiton cycles of cholesterol side chain degradation structurally and biochemically.

3. **Investigate the gene products (Rv3540c, Rv3541c and Rv3542c) from the *igr* operon structurally and biochemically.** Both the structures and functions of Ltp2 (Rv3540c), ChsH1 (Rv3541c), and ChsH2 (Rv3542c) were unknown. Earlier experiments suggested that ChsH1 and ChsH2 form a protein complex. ChsH1 and ChsH2 are hypothesized to function as an enoyl-CoA hydratase in the last cycle of cholesterol side chain β-oxidation. In this thesis, we identified and characterized ChsH1-ChsH2 structurally and biochemically. ChsH1-ChsH2, as the first obligate $\alpha_2\beta_2$ heterotetrameric enoyl-CoA hydratase, functions in the last cycle of β-oxidation in cholesterol side chain degradation. In addition, we reaveled the function of Ltp2 and established conditions for its activity.
4. **Investigate the succinylation modification on ChsE4-ChsE5 and its effects.** ChsE4 has been identified as a succinylatd protein. In this thesis, we showed that succinylated ChsE4-ChsE5 had decreased enzyme acitivity. In addition, we identified Nat (Rv3566c) as a succinylase to catalyze the succinylation of ChsE4-ChsE5. We propose that post-translation modification is another strategy *Mtb* adopts to regulate its metabolism.

Chapter II. Identification and characterization of $\alpha_2\beta_2$ heterotetrameric ACADs from *Mtb*

This chapter is adapted from a paper published in Journal of Bacteriology[68].

1. Introduction

Mtb belongs to the phylum Actinobacteria, the members of which are known for the diverse array of natural products they produce as well as the many diseases they cause. The environmental competition of these bacteria for nutritional resources has resulted in their unique and intricate ability to metabolize complex carbon compounds. Interestingly, the human host provides *Mtb*, a pathogenic relative of these bacteria, an environment within the granuloma that is rich in cholesterol.[31, 80]

The ability of Actinobacteria to metabolize sterols has been of interest for the better part of the last century and several catabolite intermediates have been characterized. However, the relationship between gene products and metabolites remains poorly understood. The increase in availability of genome sequences[81] and the application of transcriptional profiling experiments[39, 51] has led to the tentative assignment of genes encoding cholesterol degrading enzymes. Recombinant expression of cholesterol-regulated genes in combination with biochemical activity assays has provided successful mapping of validated enzymatic activities to specific substrates. Phenotypic profiling of genes required for *Mtb* growth on cholesterol has also been used to establish which genes are involved in sterol metabolism.[82] Many (but not all) of the genes encoding cholesterol metabolism enzymes are regulated by cholesterol. Upon mutation of these genes, phenotypes in intracellular growth and/or persistence in the chronic phase of infection are observed.

More problematic are assignments of cholesterol-regulated β-oxidation enzymes to specific functions. The genome of *Mtb* encodes multiple copies of the genes classically involved in β-oxidation, a case of apparent functional redundancy. In the case of the acyl-Coenzyme A dehydrogenases (acyl-CoA dehydrogenases, ACADs or *fadE*s), flavoproteins that catalyze the

α,β-unsaturation of acyl-CoA thioesters in β-oxidation, there are 35 genes computationally annotated as encoding this activity in the *Mtb* genome (Table 1).

Table 2-1. 35 annotated acyl-CoA dehydrogenase genes (*fadEs*) in *Mtb* genome

fadE genes	Rv number	Relative gene length	Cholesterol-up/down
fadE1	*Rv0130*		n.a.[1]
fadE2	*Rv0154c*		n.a.
fadE3	*Rv0215c*		n.a.
fadE4	*Rv0231*		n.a.
fadE5	*Rv0244c*		Up[2]
fadE6	*Rv0271c*		n.a.
fadE7	*Rv0400c*		n.a.
fadE8	*Rv0672*		n.a.
fadE9	*Rv0752c*		n.a.
fadE10	*Rv0873*		n.a.
fadE12	*Rv0972c*		n.a.
fadE13	*Rv0975c*		n.a.
fadE14	*Rv1346*		Up
fadE15	*Rv1467c*		n.a.
fadE16	*Rv1679*		Down[3]
fadE17	*Rv1934c*		Up
fadE18	*Rv1933c*		Up
fadE19	*Rv2500c*		n.a.
fadE20	*Rv2724c*		n.a.
fadE21	*Rv2789c*		n.a.
fadE22	*Rv3061c*		n.a.
fadE23	*Rv3140*		Down
fadE24	*Rv3139*		Down
fadE25	*Rv3274c*		n.a.
fadE26	*Rv3504*		Up
fadE27	*Rv3505*		Up
fadE28	*Rv3544c*		Up
fadE29	*Rv3543c*		Up
fadE30	*Rv3560c*		Up
fadE31	*Rv3562*		Up
fadE32	*Rv3563*		Up
fadE33	*Rv3564*		Up
fadE34	*Rv3573c*		Up
fadE35	*Rv3797*		n.a.
fadE36	*Rv3761c*		n.a.

[1] means the corresponding mRNA transcription level change is about one or not observed when *Mtb* was grown on cholesterol reative to glycerol. [2] means the corresponding mRNA transcription level change was at least two folds higher when *Mtb* was grown on cholesterol reative to glycerol. [3] means the corresponding mRNA transcription level change was at least two folder lower when *Mtb* was grown on cholesterol reative to glycerol. [4] orange means cholesterol regulated.

In mammals, all the acyl-CoA dehydrogenases (ACADs) reported so far function in either fatty acid β-oxidation cycle or amino acid catabolism.[83] Glutaryl-CoA dehydrogenase (GD) plays

a role in the metabolism of lysine, hydroxylysine and tryptophan by acting upon glutaryl-CoA. Isobutyryl-CoA dehydrogenase (iBD) is responsible for the catabolism of valine. Iso(3)valeryl-CoA dehydrogenase (i3VD) and iso(2)valeryl-CoA dehydrogenase catalyze oxidation in leucine and isoleucine metabolism, respectively.[84] The ACADs responsible for fatty acids degradation, are categorized based on the chain lengths of their preferred substrates. The five ACAD subtypes (SCAD, MCAD, LCAD, VLCAD1, VLCAD2) function in fatty acid β-oxidation by catalyzing the first dehydrogenation step of short, medium, long, and very long chain acyl-CoAs. In general, ACADs have the lowest catalytic activity compared to other enzymes in a β-oxidation cycle. Therefore, their enzymatic activities typically control the flux of the whole β-oxidation cycle.[83]

Interestingly, nearly all ACADs share the same tetrahedral tetrameric assembly in which a single chain has one flavin adenine dinucleotide (FAD) binding site and one acyl-CoA binding site (Figure 2-1a and 2-1b). The exception is the very long chain ACAD (VLCAD), which has a homodimeric architecture. Secondary structure alignments between all the characterized ACAD chains except for VLCAD revealed that they share high sequence similarities and similar secondary structure elements, constituting the representative, ACAD N-terminal domain, ACAD C-terminal domain, and ACAD middle β-barrel domain (Figure 2-1b). In addition to a redox cofactor, a catalytic base to deprotonate the acyl-CoA substrate is required for dehydrogenation. This catalytic base is typically a glutamate, and sometimes an aspartate. Glutamate is found at one of two positions within the tertiary structures of ACADs: either in helix G, exemplified by *iso*-valeryl-CoA dehydrogenase (i3VD),[85] or at the junction of helices J and K, for example as found in long chain acyl-CoA dehydrogenase (LCAD).[86] (Figure 2-2)

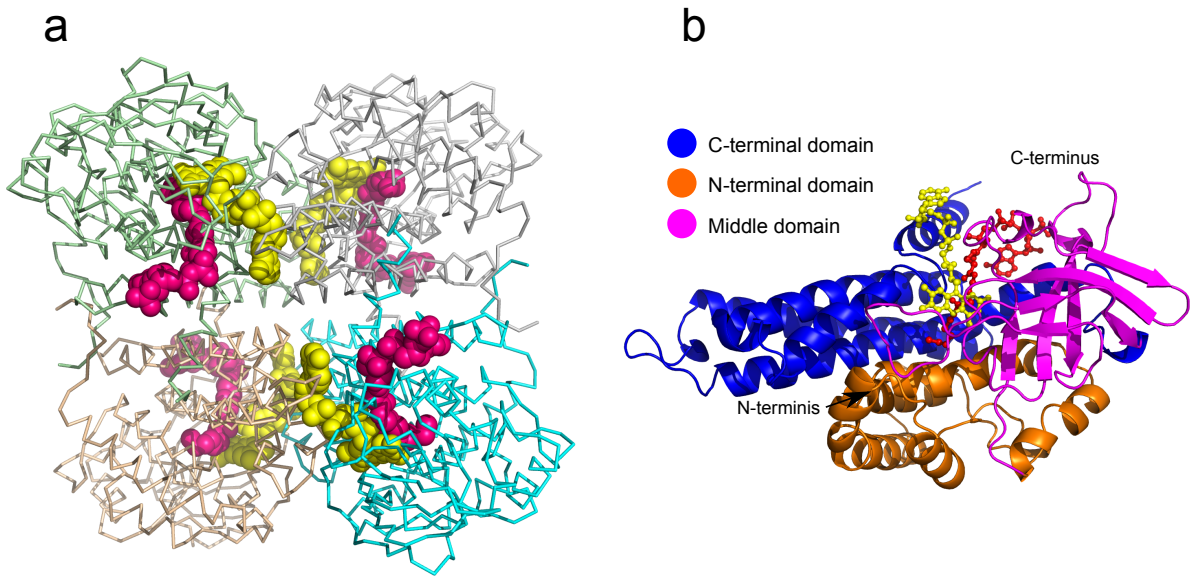


Figure 2-1. Typical human homotetrameric ACAD structure. (a) Homotetrameric assembly of human MCAD (PDB code: 1EGC) with four FAD molecules (yellow spheres) and four octanoyl-CoA molecules (hot pink spheres); (b) Crystal structure of one single α chain from human MCAD with one FAD molecule (yellow) and one octanoyl-CoA molecule (red). The three ACAD domains are colored as described in the figure.

In addition to their highly conserved 3D architecture, all the ACADs characterized so far also share the same catalytic mechanism of α, β-dehydrogenation. The dehydrogenation occurs with a concerted pro R hydrogen abstraction by the active site, Glu, at the α position; the pro R hydrogen at the β position is transferred as a hydride to N5 of FAD forming $FADH^-$. An electron transferr flavoprotein (ETF) oxidizes the reduced $ACAD_{FADH^-}$ to $ACAD_{FAD}$, which facilitates the release of dehydrogenated acyl-CoA products.[83] (Figure 2-3)

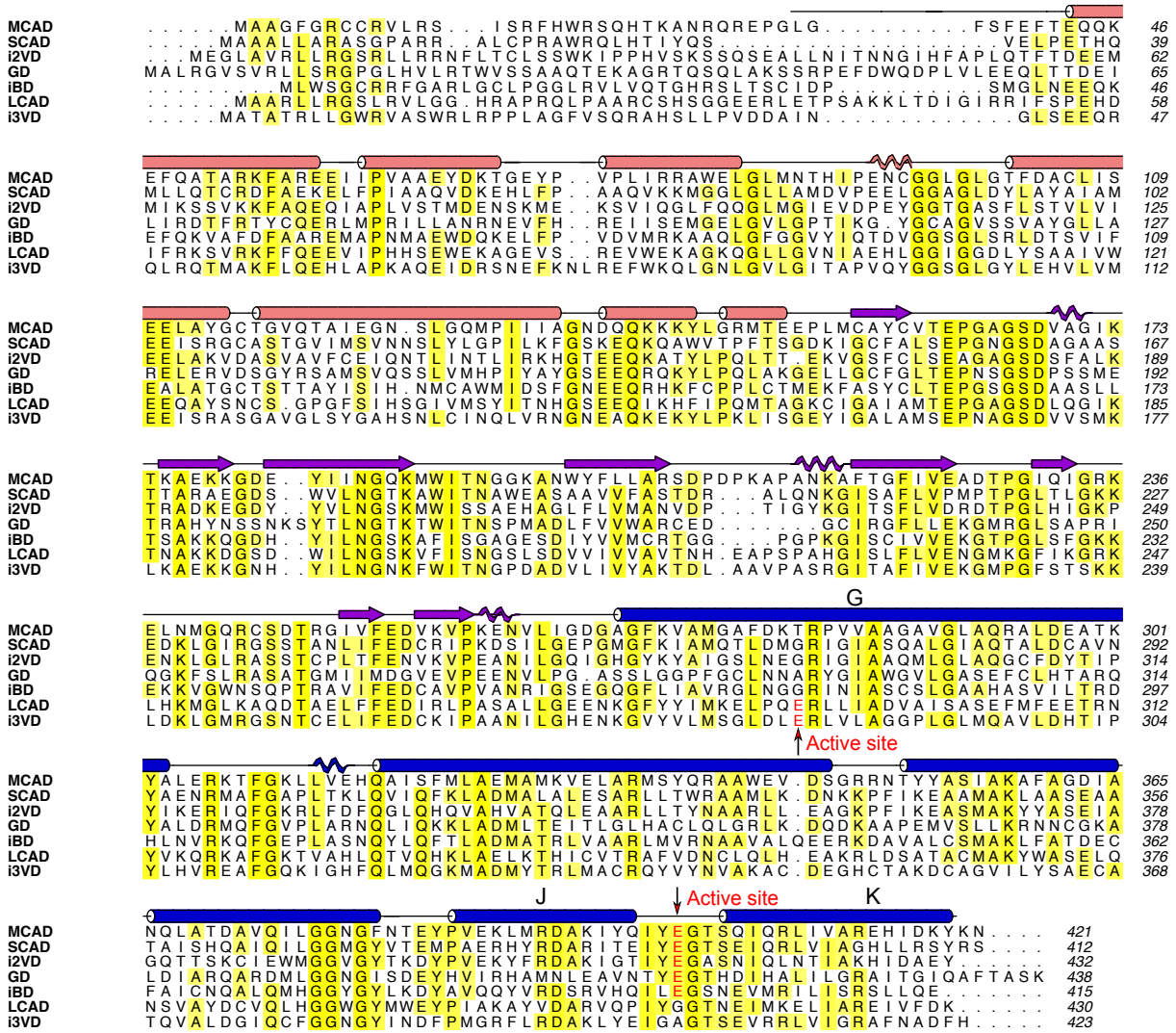
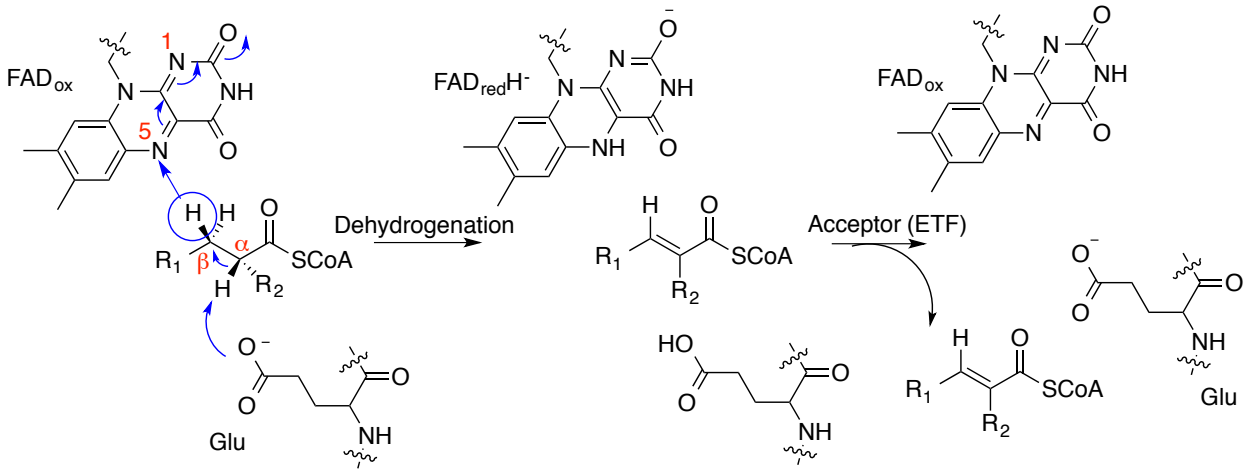


Figure 2-2. Structure based protein sequence alignment among all the human ACADs except VLCADs. Structure based protein sequence alignment of human ACADs including medium chain acyl-CoA dehydrogenase (MCAD), short chain acyl-CoA dehydrogenase (SCAD), long chain acyl-CoA dehydrogenase (LCAD), glutaryl-CoA dehydrogenase (GB), isobutyryl-CoA dehydrogenase (iBD), *iso*-2-valeryl-CoA dehydrogenase (i2VD) and *iso*-3-valeryl-CoA dehydrogenase (i3VD). The secondary structure elements were added based on the crystal structure of MCAD (PDB code: 1EGC) and colored coded similarly as Figure 2-1b.

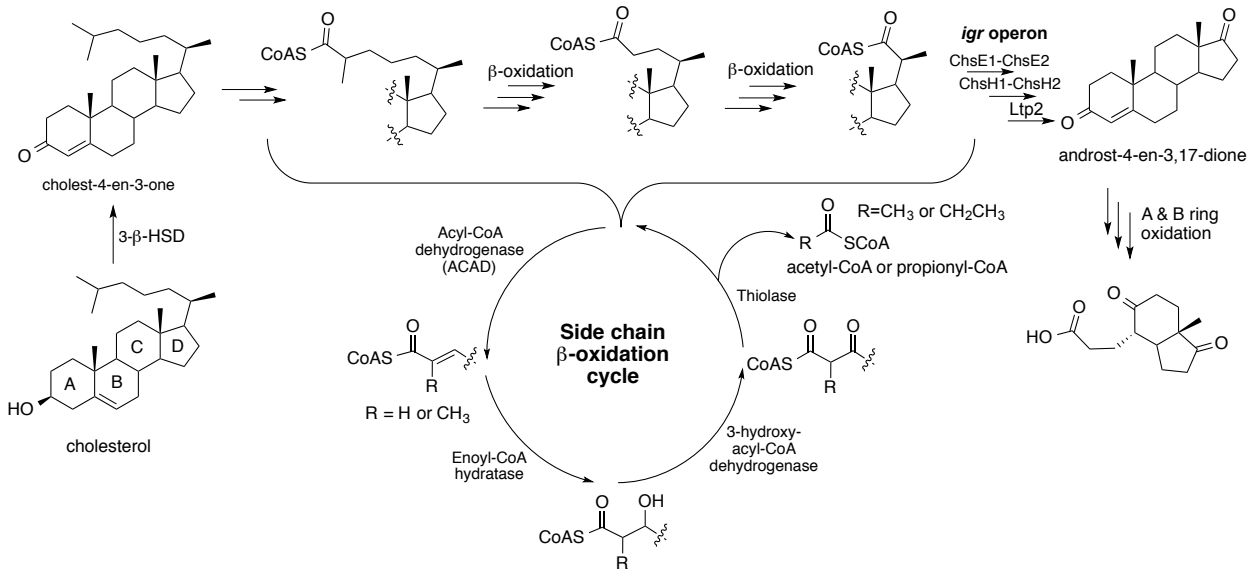
Elucidating structures of cholesterol-related ACADs in *Mtb* and understanding their functions in cholesterol metabolism pathways is valuable for unraveling the whole cholesterol catabolism picture in *Mtb*. The cholesterol-regulated *intracellular growth* (*igr*) operon contains six genes in total, two of which are annotated as acyl-CoA dehydrogenases that reside in adjacent open

reading frames. Interestingly, recent work from our laboratory has demonstrated that *fadE28* and *fadE29* code for two separate proteins that form a functional $\alpha_2\beta_2$ heterotetrameric enzyme complex.[56] ACAD FadE28-FadE29 (now renamed ChsE1-ChsE2) catalyzes the unsaturation of 3-oxo-4-pregnene-20-carboxyl CoA (3-OPC-CoA), an intermediate in the cholesterol metabolism pathway (Scheme 2-2).[56] This work represents the first definitive assignment of catalytic function to FadE enzymes in the cholesterol pathway. Previous studies were unable to resolve the ambiguities in potential function through sequence homology studies.[82] The enzyme activity data of ChsE1-ChsE2 in combination with metabolic knockout studies of the *igr* operon defined the encoded activity of five of the six genes in the *igr* operon to be removal of the C20-C22 propionate moiety of the cholesterol side-chain.[50, 56] (Scheme 2-2) The sixth gene encodes Cyp125 that catalyzes oxidation of C26 of cholest-4-en-3-one.[87]



Scheme 2-1. Mechanism of dehydrogenation process catalyzed by ACADs towards acyl-CoAs.

ChsE1 and ChsE2 form an obligate $\alpha_2\beta_2$ heterotetramer, and either protein expressed individually does not bind flavin adenine dinucleotide (FAD) cofactor.[56] To our knowledge, this was the first example of a heteromeric ACAD in any kingdom of life.



Scheme 2-2. *Mtb* cholesterol side chain degradation pathway. Not all individual steps are shown.

Based on insights gained from this unusual quaternary structure of ChsE1-ChsE2, we searched the *Mtb* genome for additional clusters of *fadE* genes that might form protein complexes. Of the 35 annotated *fadEs*, we identified five additional groups of *fadEs* encoded in operons, all of which are regulated by cholesterol, as well as a sixth not regulated by cholesterol (Figure 2-3). Here, we establish that the $\alpha_2\beta_2$ heterotetrameric ACAD motif is repeated within the cholesterol-regulated ACAD proteome. We conclude *fadE* genes that 1) are regulated by cholesterol, 2) are proximal to another *fadE* gene, and 3) retain only half of the expected cofactor binding residues form heteromeric $\alpha_2\beta_2$ tetramers with two active sites. Furthermore, we identify additional bacteria that utilize this genetic architecture. Some of these bacteria are distantly related to *Mtb*, but all contain additional genes associated with steroid metabolism. These findings suggest that this structural motif has evolved to accommodate polycyclic-CoA substrates such as steroid-CoA thioesters.

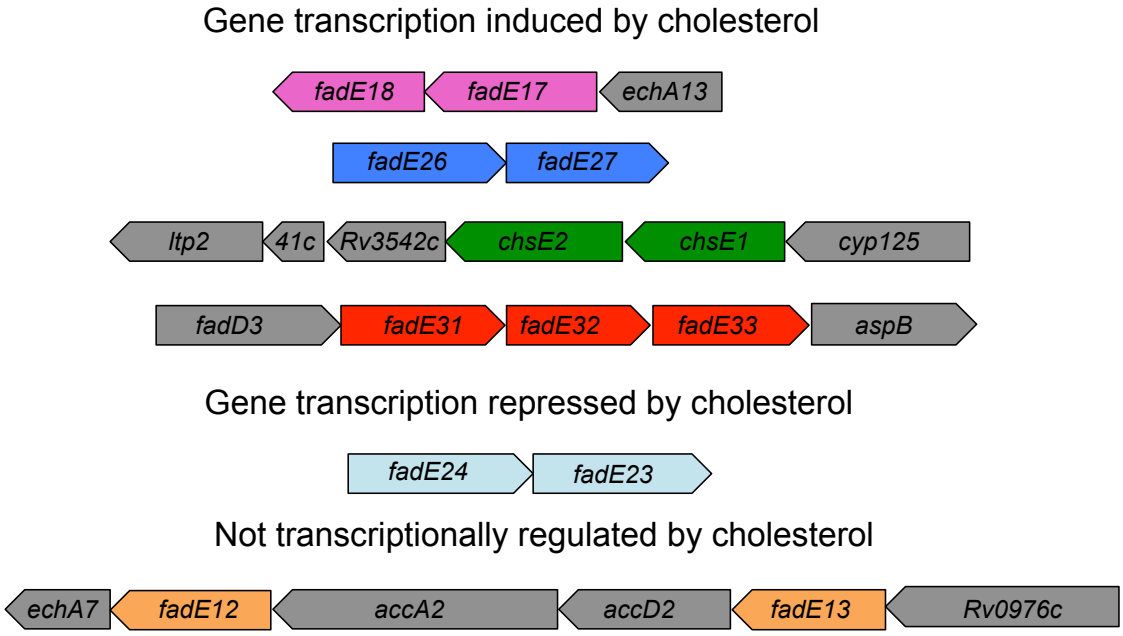


Figure 2-3. Operonic organization of *Mtb fadE* genes studied in this work. In the *Mtb* genome, there are six operons containing multiple genes annotated as *fadE* genes, all of which are regulated by cholesterol except the operon containing *fadE12* and *fadE13*.

2. Experimental methods

Materials and General Methods. Total genomic DNA from *Mtb* H37Rv was obtained from the Tuberculosis Research Materials Facility at Colorado State University (Fort Collins, CO). DNA primers were ordered from Eurofins (Huntsville, AL). iProof High-Fidelity DNA Polymerase, used for gene amplification from genomic *Mtb* H37Rv DNA, was purchased from Bio-Rad Laboratories (Melville, NY). The pET vector system from Novagen was used for cloning (Madison, WI). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). *E. coli* BL21(DE3) cells were obtained from Bio-Rad. The chaperone plasmid set, pG-KJE8, was from Takara Bio Inc (Japan). Tryptone and ampicillin were purchased from Fisher Scientific (Pittsburgh, PA). Yeast extract was purchased from Research Products International Co. (Mount Prospect, IL). DNA sequencing was performed at the Stony Brook Sequencing Facility using an Applied Biosystems ABI 3730 (48 cm capillary array) to confirm recombinant gene sequences. Cell disruption was performed using a Constant Systems, Inc. TS Series Benchtop instrument (Kennesaw, Georgia). The following buffers were used: Buffer A (binding): 20 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazole (pH 8.0); Buffer B (elution): 20 mM Tris-HCl, 0.5 M NaCl, 0.5 M imidazole (pH 8.0); Buffer C: 50 mM Tris-HCl, 200 mM NaCl (pH 8.0). Electroporation was performed using a BTX Electroporation system (Holliston, MA). Isolation of hexahistidine tagged recombinant proteins was achieved via immobilized-metal affinity chromatography (IMAC) using Ni-NTA His•Bind Resin (Novagen). Proteins were identified by trypsin digest peptide mass fingerprinting using a Bruker Autoflex II TOF/TOF instrument operating in positive ion mode (Bruker Daltonics Inc., Billerica, MA). FPLC gel filtration chromatography was performed on a dextran Superdex 200 16/60 column from GE Biosciences (GE Healthcare, Piscataway, NJ). Analytical ultracentrifugation (AUC)

was performed on a Beckman Coulter L-90k AUC instrument (Brea, CA). Ultra-performance liquid chromatography (UPLC) mass spectrometry (LC-MS) was performed on an Acquity Ultra Performance LC instrument equipped with a diode array and SQD detectors from Waters (Milford, MA). The column used was an XBridge BEH C4 Column, 300 Å, 3.5 μm, and 2.1 mm x 100 mm. The following solvents were used: Solvent E: 95:5/H_2O:2-propanol, 0.1% (v/v) trifluoroacetic acid (TFA); Solvent F: 5:95/H_2O:2-propanol, 0.1% (v/v) TFA.

Gene Cloning and Plasmid Construction. All genes were initially amplified from *Mtb* H37Rv total genomic DNA by PCR. PCR products were doubly digested with NdeI and HindIII restriction enzymes and ligated into a similarly digested pET vector for *E. coli*. Ligated plasmids containing the gene of interest were transformed into competent XL1-Blue *E. coli* cells for amplification and isolation of plasmid DNA. DNA sequencing with T7 promoter and terminator primers confirmed that the full gene was present without mutations. For each *fadE* operonic set, individual gene constructs for *in trans* expression, as well as polycistronic operonic constructs were prepared for protein expression (Table 2-2).

Table 2-2. Gene constructs used in this study.

Construct[1]	*Rv* Number	Gene name	Vector[2]	His_6-tag
p24N/23	*Rv3139/3140*	*fadE24-fadE23*	pET28b(+)	N-terminal
p24N	*Rv3139*	*fadE24*	pET28b(+)	N-terminal
p23	*Rv3140*	*fadE23*	pET20b(+)	none
p23N	Rv3140	*fadE23*	pET28b(+)	N-terminal
p26N/27	*Rv3504/3505*	*fadE26-fadE27*	pET-28b(+)	N-terminal
p26N	*Rv3504*	*fadE26*	pET-28b(+)	N-terminal
p27	*Rv3505*	*fadE27*	pET-20b(+)	none
p27N	*Rv3505*	*fadE27*	pET-28b(+)	N-terminal
pigr5[4]	*Rv3544c-Rv3540c*	*fadE28-fadE29*	pET-28b(+)	N-terminal
p12N	*Rv0972c*	*fadE12*	pET-28b(+)	N-terminal
p13N	*Rv0975c*	*fadE13*	pET-28b(+)	N-terminal
p12	*Rv0972c*	*fadE12*	pET-20b(+)	none
p13	*Rv0975c*	*fadE13*	pET-20b(+)	none

[1]Vector names were assigned N or C following the *fadE* number that contained either an N- or C- terminal His_6-tag. The 5'-terminus and 3'-terminus cloning sites were NdeI and HindIII, respectively. [2]Antibiotic resistance markers were as follows: pET20b(+): amp^R; pET28b(+).[3]

Protein Expression in *E. coli*. Competent BL21(DE3) *E. coli* cells were transformed with recombinant constructs, with or without chaperone plasmid pG-KJE8. Single colonies were selected on LB/agar plates supplemented with the appropriate antibiotics(s); 30 μg/mL kanamycin for pET28b(+), 100 μg/mL ampicillin for pET20b(+), and 20 μg/mL chloramphenicol for pG-KJE8. *E. coli* was cultured at 37°C in 2xYT media and when necessary chaperone expression was induced by inoculation with L-arabinose (2 mg/mL) and tetracycline (10 ng/mL). Cultures were grown until the OD_{600} reached 0.6 – 0.8, the temperature was lowered to either 25 °C or 16 °C, cellular cultures were allowed to equilibrate, and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. Cultures were grown for about 12 h at 25 °C or 16 °C.

IMAC Protein Purification. Cells were harvested by centrifugation at 5000 g for 15 min at 4 °C and all subsequent steps were conducted at 4 °C. The cells were suspended in Buffer A, and

either lysed by sonication (5 times, 1 min), French press (3 times at 10,000 psi), or cell disruption (2 times, 27,000 psi). The method of cell lysis did not affect protein complex formation or solubility. Cell disruption was used for large-scale experiments. Cellular debris was removed by centrifugation at 125,000 g for 1 h. The supernatant was loaded onto IMAC His•Bind resin, washed with 10 column volumes of Buffer A, and eluted with Buffer B. Protein solutions were immediately desalted by dialysis with Buffer C or with a G25 desalting column equilibrated and developed in Buffer C.

Size-Exclusion Chromatography Purification of Soluble ACAD Proteins. Isolated ACAD proteins in Buffer C were concentrated by ultrafiltration (MWCO 10 kDa) to less than 3 mL and purified by size-exclusion chromatography (SEC) on a Superdex 200 column equilibrated in Buffer C. The column was equilibrated in Buffer C at a flow rate of 1.0 mL/min and monitored at 220 nm, 280 nm, and 440 nm. Fractions with absorbance at 440 nm were collected and concentrated.

Analysis of Recombinant Proteins. Purified proteins were analyzed by reducing SDS-PAGE and protein band identities were confirmed by in-gel tryptic digestion and subsequent MALDI-TOF mass spectrometry. Protein bands from SDS-PAGE were excised, washed with H_2O, and dried with CH_3CN. The protein cysteine residues were reduced using 45 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate (NH_4HCO_3) for 45 min at 56 °C, and then alkylated using 55 mM iodoacetamide in 100 mM NH_4HCO_3 for 30 min in the dark. Samples were dried completely by lyophilization, and rehydrated with trypsin solution (200 ng/μL) in 50 mM NH_4HCO_3 for about 45 min on ice. Excess trypsin was washed away, the gel pieces were redissolved in 50 mM NH_4HCO_3, and allowed to digest for at least 12 h at 37 °C. This solution was combined with peptide fragments, which were extracted thrice using 50 μL of 60% (v/v)

CH_3CN in aqueous 0.1% (w/v) TFA. The combined extracts were dried completely and dissolved in 10 – 15 μL of aqueous 0.1% TFA. This solution was mixed with an equal amount (v/v) of α-cyano-4-hydroxycinnamic acid matrix, and MALDI-TOF data were acquired and analyzed using Flex-Analysis software (Bruker).

Protein Complex Molecular Weight Determination. Samples were subjected to analytical ultracentrifugation with a Beckman Optima XL-A centrifuge equipped with an absorption optical scanner set to detect at 280 nm. Sedimentation equilibrium (SE) experiments were performed at 20 °C for all ACAD complexes except FadE23-FadE24, which was performed at 4 °C. Following SEC purification, protein samples were analyzed at three concentrations (ranging from absorbances at 280 nm of ~ 0.2 – 1) and three rotor speeds (20 k, 25 k, and 30 k). After centrifugation for 18h and 20 h at each speed, a concentration distribution curve was acquired, monitoring at 280 nm. Approximately 80 curves generated from replicate scans were fit globally in Heteroanalysis (University of Connecticut, Storrs, CT) to the ideal single-species model. The sample buffer density (ρ) and the partial specific volume ($\bar{\nu}$) of each protein were calculated using SEDNTERP (University of New Hampshire, Durham, NH).

Oligomer Stoichiometry Determination. Heteromeric ACAD complexes were analyzed on a Waters LC/UV/MS to determine the complex stoichiometry. Proteins were separated on an XBridge BEH 300 C4 3.5 μm column (2.1 x 100 mm) at 40 °C with a linear gradient from 100% Solvent E to 100% Solvent F over 15 min at a flow rate of 0.5 mL/min. MS data was collected in ESI+ mode with a cone voltage of 40 V, a capillary voltage of 4.5 kV, and source temperature of 150 °C. Peaks in the LC/UV chromatogram were assigned by the deconvolution of the multi-charged states in the corresponding mass spectra. The integrated peak areas at 280 nm of each protein were divided by the corresponding molar extinction coefficient for the protein to yield

the molar concentrations. Protein stoichiometry was determined from the ratio of the molar concentrations.

Determination of the Protein Complex to FAD Ratio. A ~5 μM (~0.8 mg/mL) solution of protein in Buffer C was denatured by boiling for 10 min. The sample was chilled on ice and centrifuged to pellet precipitated protein. The quantity of FAD obtained in the supernatant was determined using the absorbance and extinction coefficient of FAD at 260 nm, ε_{260} = 37,000 $M^{-1}cm^{-1}$[156]. The protein pellet was dissolved in 6 M guanidine-HCl and the concentration determined using the calculated extinction coefficients at 280 nm based on the primary amino acid sequence (determined using ExPasy).[88] The experiments were performed in triplicate for each protein complex, and the averaged concentration for protein and concentration for FAD were obtained for each ACAD.

Dehydrogenase assay. Dehydrogenase activity was assayed using 3β-hydroxy-chol-5-en-24-oyl-CoA. The identity of the CoA thioester was confirmed by LC-MS. Each ACAD (100 nM) was assayed using ferrocenium hexafluorophosphate (250 μM) as an artificial electron acceptor and 1 mM substrate in 100 mM HEPES, pH 7.4.[89] The formation of product was followed spectrophotometrically at 300 nm and 25 ºC with the extinction coefficient of ferrocenium hexafluorophosphate, 4.3 mM^{-1} cm^{-1}. Product formation was confirmed by MALDI-TOF MS. Negative controls without substrate or without enzyme were conducted to check for background reduction of ferrocenium hexafluorophosphate.

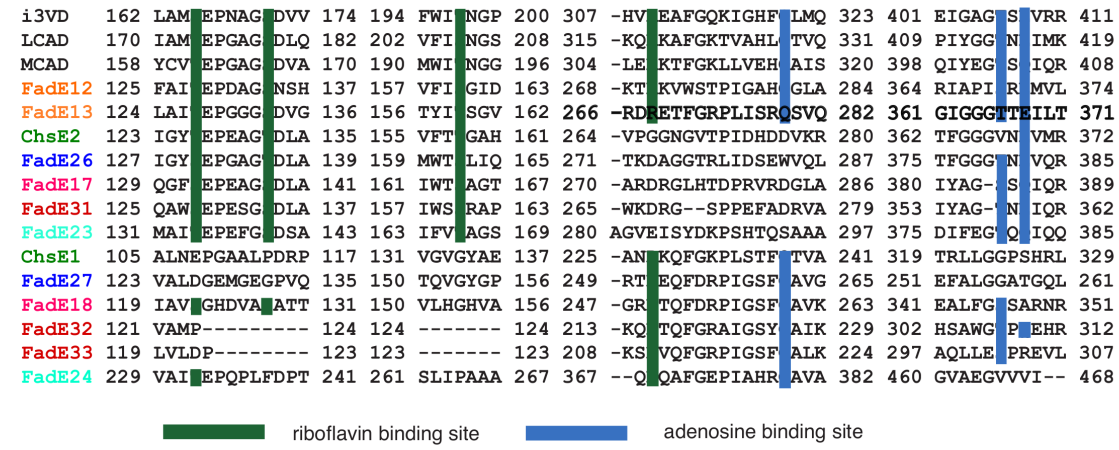
3. Results

Bioinformatic analysis of the annotated FadE proteome of *Mtb*. Based on the discovery of a novel heteromeric acyl-CoA dehydrogenase encoded in the *igr* operon, we undertook a full bioinformatic analysis of the *Mtb* acyl-CoA dehydrogenase proteome. Analysis of the genomic context of the 35 *Mtb fadE* genes found that six operons contain more than one *fadE* gene, and in five of the six cases, these genes were adjacent and regulated by cholesterol.[39, 51] In one case, three *fadE* genes were adjacent (*fadE31*, *fadE32*, *fadE33*). The *fadE*12 and *fadE13* genes are not adjacent in their operon and are not regulated by cholesterol (Table 2-1 and Figure 2-3).

Typically, ACADs are homotetramers that contain four active sites and four FAD cofactors. The homotetramer assembles as a dimer of dimers with an S_2 axis of symmetry. Within a dimer, there are two FAD cofactor-binding sites, which are related by a C_2 axis at the monomer-monomer interface, and each binding site contains residues from both the A and the B chains (Figure 2-1).

We further analyzed the alignments of these *fadE* sequences from *Mtb* with structurally characterized ACADs. We found that for the adjacent *Mtb fadE* gene pairs, many key FAD binding site residues are not conserved in each of the polypeptides (Figure 2-4a and 2-4b). Moreover, if a single polypeptide sequence is threaded onto a homotetramer structure, only half of each of the four FAD binding sites contains conserved amino acids required to form hydrogen bonds with the cofactor. In contrast, when two polypeptides encoded by adjacent genes are threaded onto chains A and B, conserved amino acids from both polypeptides comprise an intact FAD binding site; the residues that align with the second binding site are not suitable for binding FAD. In contrast, for FadE12 and FadE13, which are encoded by non-adjacent genes, the FAD binding residues are fully conserved.

a



b

	i3VD	i3VD
	Chain A	Chain B
Catalytic Base	Glu283	–
FAD binding site 1	Ser165A Ser171A Thr197A Thr406A Glu408A	Gln320B Arg309B
FAD binding site 2	Arg309A Gln320A	Glu408B Thr406B Thr197B Ser171B Ser165B
Catalytic Base	–	Glu283

	FadE12	FadE12
	Chain A	Chain B
	Glu244	–
	Thr128α Ser134α Ser160α Ser369α Glu371α	Gln281β Arg270β
	Arg270α Gln281α	Glu371β Ser369β Ser160β Ser134β Thr128β
	–	Glu244

	FadE13	FadE13
	Chain A	Chain B
	Glu242	–
	Thr127α Ser133α Thr159α Thr366α Glu368α	Gln279β Arg268β
	Arg268α Gln279α	Glu368β Thr366β Thr159β Ser133β Thr127β
	–	Glu242

	ChsE2	ChsE1
	α	β
	Glu241	–
	Thr126α Thr132α Thr158α Val367α Glu369α	Gln238β Arg227β
	Gly266α Asp277α	Ser326β Gly324β Gly134β Pro114β Glu108β
	–	Ala201

	FadE26	FadE27
	α	β
	Glu247	–
	Ser130α Thr136α Ser162α Thr380α Glu382α	Gln262β Arg251β
	Asp273α Trp284α	Thr258β Gly256β Gly153β Gly132β Asp127β
	–	Ser225

	Heterotetramer	
	Fad17	FadE18
	α	β
Catalytic Base	Glu247	–
FAD binding site 1	Ser132α Ser138α Ser164α Ser384α Gln386α	Gln260β Arg249β
FAD binding site 2	Asp272α Asp283α	Ala348β Ser346β Gly153β Ser128β Ser122β
Catalytic Base	–	Ala223

	Heterotetramer	
	FadE31	FadE33
	α	β
	Glu243	–
	Ser128α Ser134α Ser160α Thr357α Glu359α	Gln221β Arg300β
	Asp267α Asp276α	Arg304β Ser302β – – Asp122β
	–	Ala184

	Heterotetramer	
	FadE31	FadE32
	α	β
	Glu243	–
	Ser128α Ser134α Ser160α Thr357α Glu359α	Gln226β Arg215β
	Asp267α Asp276α	Gln309β Thr307β – – Pro124β
	–	Gly189

	Heterotetramer	
	MCAD	MCAD
	Chain A	Chain B
	Glu401	–
	Thr161A Ser167A Thr193A Thr403A Gln405A	Gln317B Arg306B
	Arg306A Gln317A	Gln405B Thr403B Thr193B Ser167B Thr161B
	–	Glu401

	Heterotetramer	
	FadE23	FadE24
	α	β
	Glu378	–
	Thr134α Ser140α Thr166α Thr380α Gln382α	Gln379β Arg368β
	Glu283α Ser294α	Val467β Val465β Pro264β Phe238β Thr232β
	–	Glu463

Figure 2-4. Protein sequence alignment between operonic FadEs from *Mtb* with characterized human ACADs. (a) Sequences of *Mtb* FadEs are aligned with human i3VD (P26440), LCAD (P28330), and MCAD (P11310). Residues highlighted in green aligned with those that interact with the isoalloxazine and diphosphate moiety of FAD (riboflavin) and residues highlighted in blue align with those that interact with the adenosine and nucleotide moiety of FAD. (b) Active site and FAD binding sites in *Mtb* homotetrameric and heterotetrameric ACADs compared to human i3VD and MCAD. Conserved residues for riboflavin binding and adenosine binding are shown in green and blue, respectively, and nonconserved residues are shown in gray.

More interestingly, we found that only one gene from each adjacent *fadE* gene pair contained a glutamate in the sequence alignments. The other partner did not have a potential general base residue, and frequently gaps were present in the alignments. One exception was the case of

FadE23 and FadE24, for which a glutamate residue from both proteins aligned with the medium chain acyl-CoA dehydrogenase (MCAD) glutamate despite the missing FAD binding site. Moreover, both the FadE12 and FadE13 sequences contain a glutamate residue that aligns with i3VD (Figure 2-4b).

In the case of ChsE1-ChsE2, only two FAD cofactors are bound per $\alpha_2\beta_2$ heterotetramer. Furthermore, site-directed mutagenesis experiments demonstrated that the conserved Glu241 in ChsE2 is required for catalytic activity. These observations are consistent with the hypothesis that only two active sites are formed in the ChsE1-ChsE2 $\alpha_2\beta_2$ heterotetramer.[56] By analogy, we hypothesized that the adjacent or proximal *fadE* gene pairs, *fadE17-fadE18, fadE26-fadE27, fadE31-fadE32, fadE31-fadE33,* and *fadE23-fadE24* encode heterotetrameric ACADs with two active sites, and that *fadE12* and *fadE13* each encode homotetrameric ACADs with four active sites.

Protein expression of individual *fadEs*. In order to test our hypothesis, we first examined the ability of each FadE protein to form a soluble, holoprotein independently. Each *fadE* gene was expressed individually with a His_6-tag without its operonic partner, and subsequently purified by IMAC. Individual *fadE* gene expressions were performed with coexpression of chaperones DnaK, DnaJ, GrpE, GroES, and GroEL. All FadE proteins, with the exception of FadE17, FadE27, and ChsE1 (FadE28), were expressed in soluble form. However analysis by UV/visible spectroscopy revealed that only FadE12 and FadE13 were isolated as the holoprotein with flavin cofactor bound. FadE18, FadE23, FadE24, FadE26, ChsE2 (FadE29), FadE31, FaE32, and FadE33 were obtained as the apoprotein without FAD bound as indicated by UV-visible spectroscopy (Table 2-3).

Table 2-3: Individual gene expression of operonic *Mtb* FadE proteins[1]

FadE number	18	23	24	26	27	28[2]	29[3]	31	32	33	12	13
Observed	S[4]	S	S	US[5]	UF[6]	UF	S	S	S	S	S	S
Isolated with	N	N	N	N	N	N	N	N	N	N	Y	Y
His_6-tag Location[8]	C	N	N	N	N	N	N	N	N	C	N	N

[1]Proteins were isolated by immobilized metal ion affinity chromatography (IMAC). [2]ChsE1. [3]ChsE2. [4]S: soluble and stable at 4 °C. [5]US: soluble but precipitated after IMAC. [6]UF: located in inclusion bodies and unfolded. [7]Flavin adenine dinucleotide (FAD) binding was determined by UV-visible spectroscopy. [8]The His_6-tag was either placed on the N- or C-terminus as indicated.

Protein expression of operonic *fadE* partners. Next we evaluated complex formation of operonic *fadE* gene products. We prepared expression constructs in which only one *fadE* gene of each proposed pair contained a His_6-tag for purification. For all *fadE* genes, we prepared constructs for expression *in cis* on a single plasmid and for expression *in trans,* in which each gene was cloned into a separate expression vector. For the gene expressions *in cis*, adjacent genes were cloned into pET28b with an N-terminal His_6-tag. For the gene expressions *in trans*, one gene was cloned into pET28b with an N-terminal His_6-tag and the adjacent gene was cloned into pET20b without any tag.

After *in cis* expression, soluble protein complexes with bound FAD cofactor were isolated by IMAC purification for FadE23-FadE24 and FadE26-FadE27 (Table 2-4). Using the *in cis* expression constructs, p31N/32/33, in which FadE31 contained an N-terminal His_6-tag, and p31/32/33C, for which FadE33 contained a C-terminal His_6-tag, we isolated FadE31-FadE32 and FadE31-FadE33, respectively (Table 2-4). In the case of FadE17 and FadE18, soluble protein was not obtained.

We further tested to see if pairs of *fadEs* could be expressed *in trans*. FadE23-FadE24 and FadE26-FadE27 again formed soluble complexes. For *fadE* gene triplet *fadE31-fadE32-fadE33,* we coexpressed every pair combination *in trans*. However, *fadE31* expressed with either *fadE32*

or *fadE33* did not yield soluble protein. Likewise, expression of *fadE32* and *fadE33 in trans* did not provide soluble, cofactor bound protein.

Table 2-4: Biophysical characterization of ACAD protein complexes from *Mtb*

ACAD complexes	H-ChsE2-ChsE1[56]	H-FadE26-FadE27	H-FadE24-FadE23	H-FadE31-FadE32	H-FadE33-FadE31-	H-FadE18-FadE17-	H-FadE12	H-FadE13
Predicted	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	α_4	α_4
Observed	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	NS[1]	α_4	α_4
Expected MW (kDa)	161	168	188	151	152	166	166	163
AUC MW (kDa)	156±1	158±1	211±2	163±0.5	164±0.5	NS[1]	139[2]	179±2
FAD:complex	2:1	2:1	2:1	2:1	2:1	ND[3]	ND	4:1
Cholesterol regulation[4]	Up	Up	Down	Up	Up	Up	NR[5]	NR
Transcriptionl Repressor	KstR1[6]	KstR1	SigE[7]	KstR2[8]	KstR2	Mce3R[9]	SenX[10]	SenX

[1]NS: not soluble. [2]FadE12 was not stable for the duration of the AUC experiment. Therefore, the molecular weight of FadE12 was obtained using analytical size exclusion. [3]ND: not determined. [4]Based on the transcriptional profiling experiments of Nesbitt et al[51]. [5]NR: not regulated by cholesterol. [6]Regulated by KstR1[46]. [7]Regulated by SigE [90]. [8]Regulated by KstR2[46]. [9]Regulated by Mce3R [91]. [10]Regulated by SenX [92].

Although individual expression of *fadE12* or *fadE13* resulted in soluble, holo-FadEs, we tested to see if FadE12 and FadE13 might form a complex. The genes were expressed *in trans*; one construct encoded a His_6-tag, whereas the second construct did not. Regardless of His_6-tag location, FadE12 and FadE13 were each isolated as homomeric complexes with bound flavin.

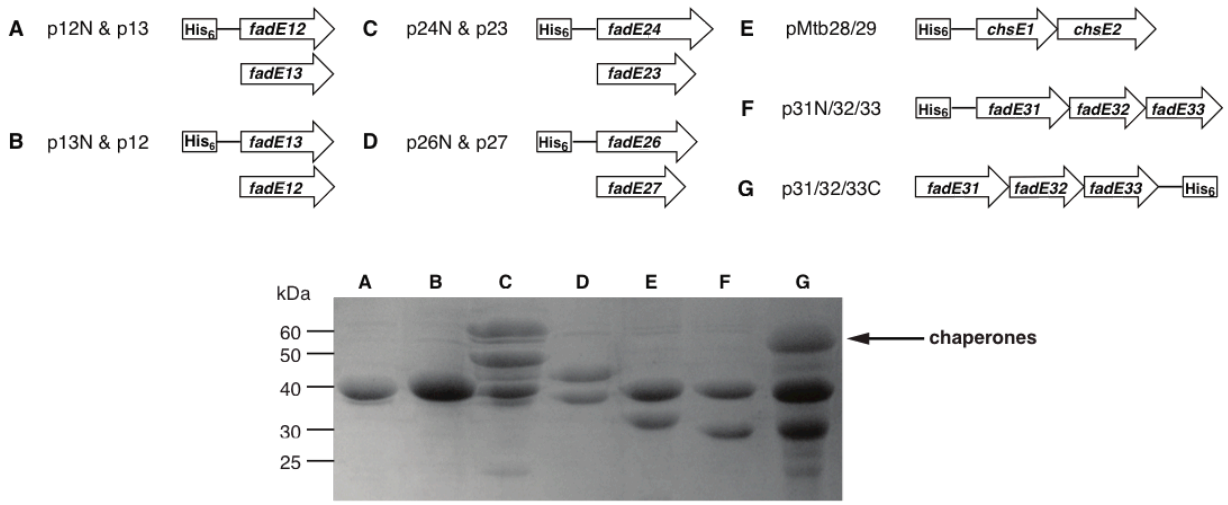


Figure 2-5. SDS-PAGE gel revealed *Mtb* ACAD complex formation. Each pair or triplet of *fadE* genes was expressed *in cis* or *in trans* as indicated. In each case, only one protein contained a His_6-tag. Protein complexes expressed with chaperones in lane C (FadE23-FadE24) and lane G (FadE31-FadE33) co-purify with overexpressed chaperone GroEL (denoted by arrow).

With the exception of FadE17-FadE18, we were able to demonstrate that all cholesterol-regulated *fadE*s that share an operon form heteromeric complexes with FAD cofactor bound (Figures 2-5 and Figure 2-6). The *fadE*s that are not regulated by cholesterol, not adjacent in their operon, and which contain fully conserved FAD binding sites, i.e., *fadE12* and *fadE13*, form the usual homomeric complexes.

Protein expression of non-operonic *fadEs*: FadE crossover experiments. In order to test the promiscuity of protein complex formation, we coexpressed *fadE26*/*chsE1* and *fadE27*/*chsE2*. FadE26 and ChsE2 share 46% amino acid identity; the highest identity between any of the cholesterol-regulated *fadEs*. Therefore, they are most likely to substitute for one another. Likewise, FadE27 and ChsE1 share 30% amino acid identity. Coexpression of $fadE26_{His6\text{-}tag}$/*chsE1 in trans* resulted in purification of apo-His_6-tagged FadE26. Similarly, coexpression of *fadE27*/$chsE2_{his6\text{-}tag}$ resulted in purification of apo-His_6-tagged ChsE2. No heteromeric ACAD complexes were isolated (Figure 2-7).

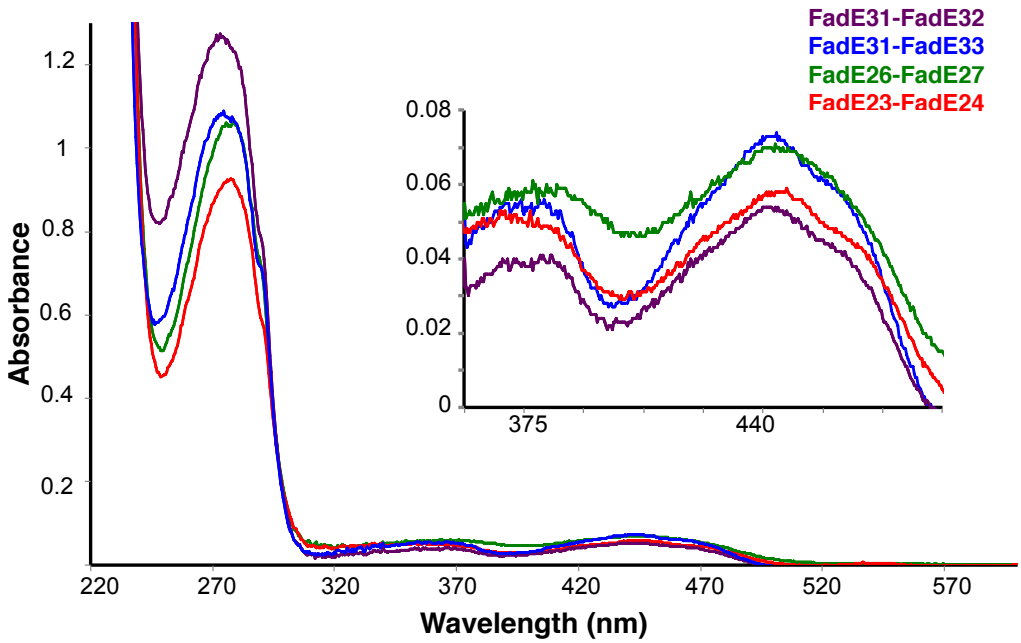


Figure 2-6. UV-visible absorbance spectra for each FadE heteromeric complex. All the concentrations were adjusted to 10 μM. Inset: the zoomed in absorbances from 340 nm to 500nm for FAD cofactor.

Characterization of isolated FadE oligomeric complexes. After expression, proteins were purified by IMAC. Soluble complexes were characteristically yellow by eye, and after further purification by size-exclusion chromatography, UV-visible spectra were obtained. The spectra of all the complexes contained typical absorbance maxima around 370 nm and 440 nm consistent with bound FAD (Figure 2-6).

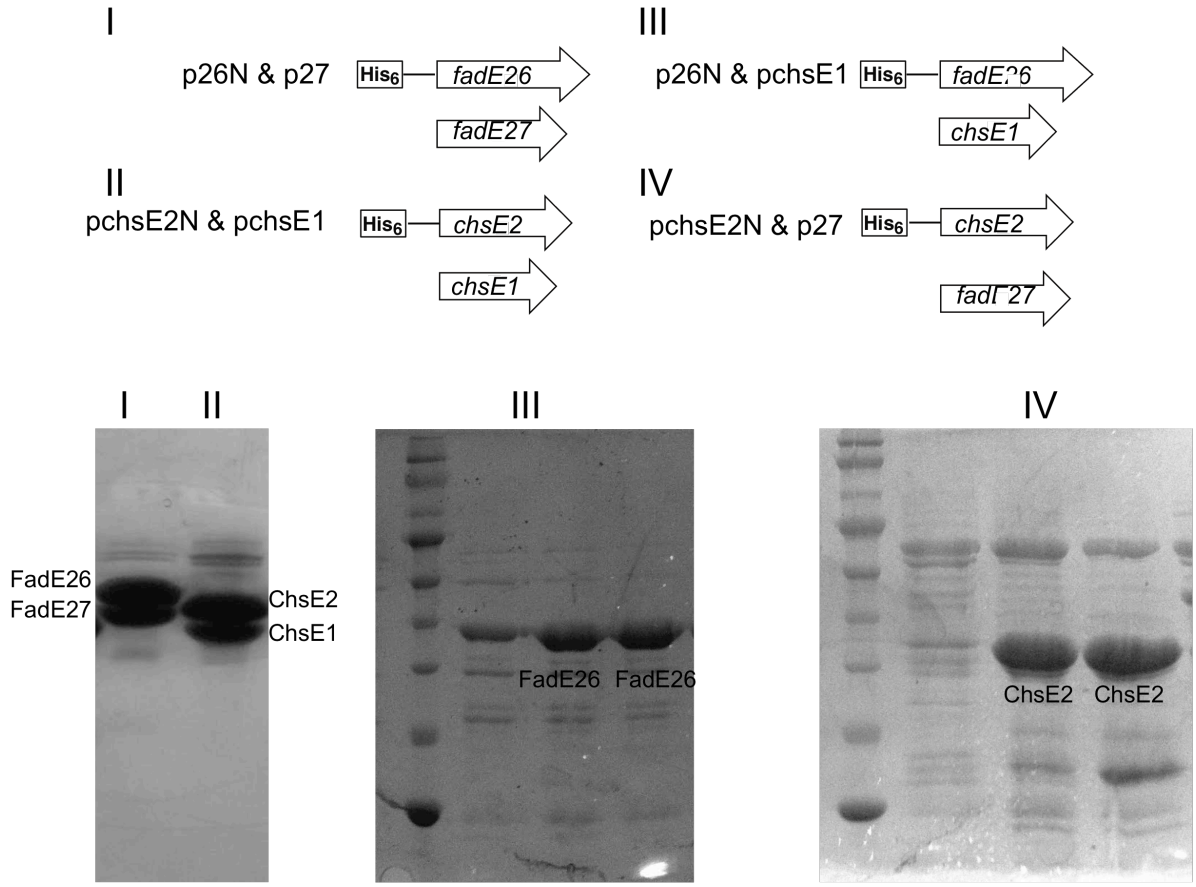
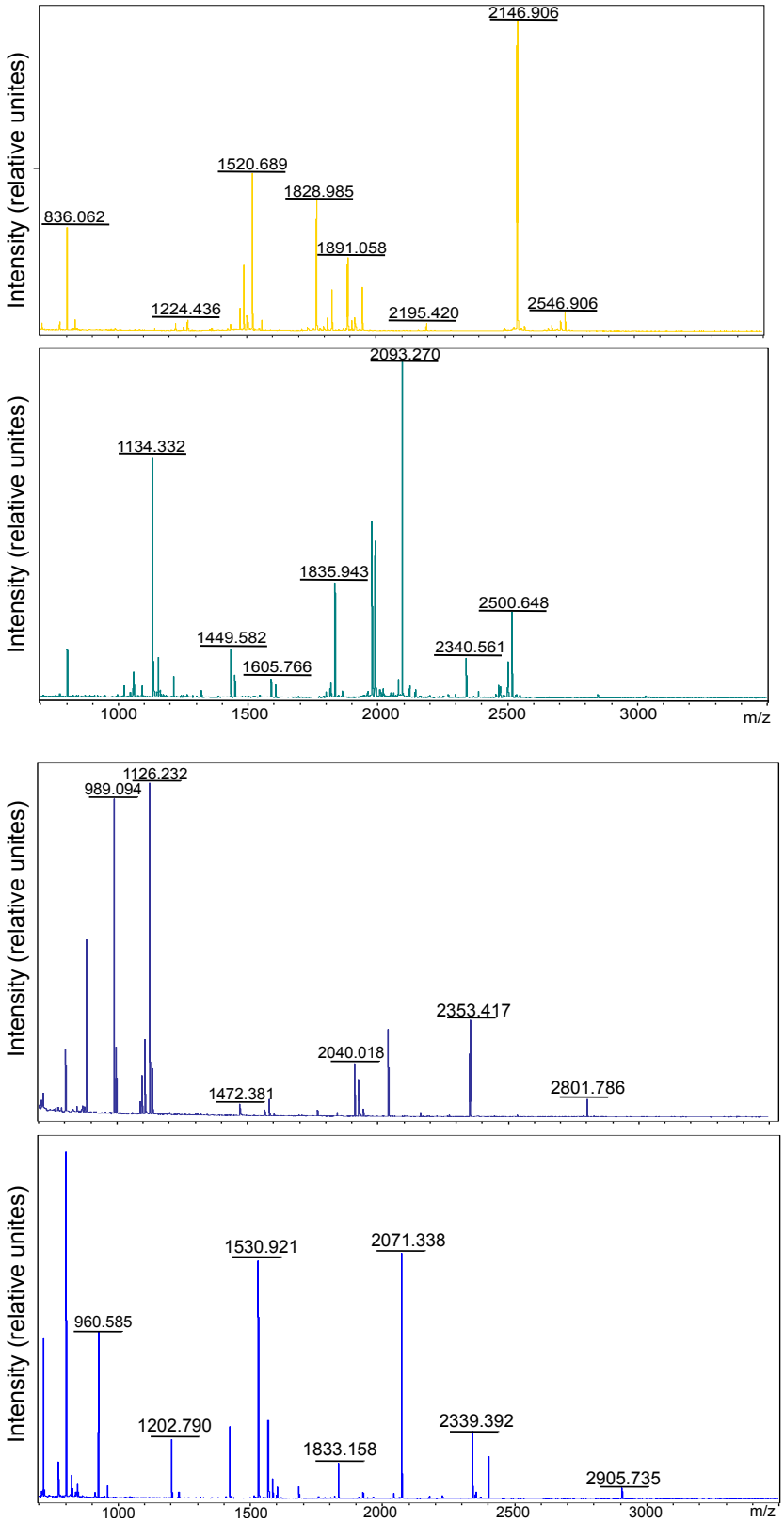


Figure 2-7. Protein expression of non-operonic *fadEs*: FadE crossover experiments. I and II are conducted as described in Figure 2-5. III. Co-expression of $fadE26_{His6\text{-}tag}$/*chsE1* in trans resulted in purification of apo-His_6-tagged FadE26; co-expression of *fadE27*/$chsE2_{His6\text{-}tag}$ resulted in purification of apo-His_6-tagged ChsE2.

For each ACAD complex isolated, we performed in-gel tryptic digest followed by MALDI-TOF mass spectrometry to confirm the identity of the polypeptides in the complexes. In all cases, the peptide fingerprint obtained by experiment matched the expected theoretical fingerprint with greater than 40% sequence coverage. The MALDI-TOF mass spectra of FadE26-FadE27 and FadE23-FadE24 after trypsin digest are shown (Figure 2-8).



FadE26

Observed	Predicted
2546.906	2545.268
2195.420	2194.099
1891.058	1889.929
1828.985	1827.887
1557.777	1556.843
1520.689	1519.786
1224.436	1223.649
836.062	835.494

FadE27

Observed	Predicted
2500.648	2500.363
2340.561	2339.212
2093.270	2092.068
1989.191	1988.092
1835.943	1834.919
1605.766	1604.850
1449.582	1448.749
1134.332	1133.631

FadE24

Observed	Predicted
2801.786	2801.592
2353.417	2352.183
2040.018	2038.919
1944.987	1944.001
1472.381	1471.622
1137.319	1136.653
1126.232	1125.579
1098.169	1097.541
989.094	988.521

FadE23

Observed	Predicted
2905.735	2905.469
2401.537	2401.309
2339.392	2339.188
2227.441	2227.256
2071.338	2071.155
1833.158	1832.998
1530.921	1530.802
1202.790	1202.710
960.585	960.526

Figure 2-8. The MALDI-TOF mass spectra of FadE26, FadE27, FadE23 and FadE24 after in gel trypsin digestion.

To determine the quaternary structure of the purified heteromeric ACAD complexes, we performed AUC sedimentation equilibrium and LC/UV/MS analysis, as well as gel filtration chromatography. The molecular weights determined by AUC were consistent with each ACAD complex forming a tetramer (Table 2-4, Figure 2-9). To confirm the stoichiometry of the complexes, they were subjected to denaturing C18-HPLC separation. For each heteromeric complex, equimolar concentrations of each polypeptide were present confirming the formation of an $\alpha_2\beta_2$ quaternary structure (Figure 2-10). All the results mentioned above concurrently demonstrated that the ACAD complexes from *Mtb* are heterotetramers. Here, only the results of FadE26-FadE27 and FadE23-FadE24 are shown.

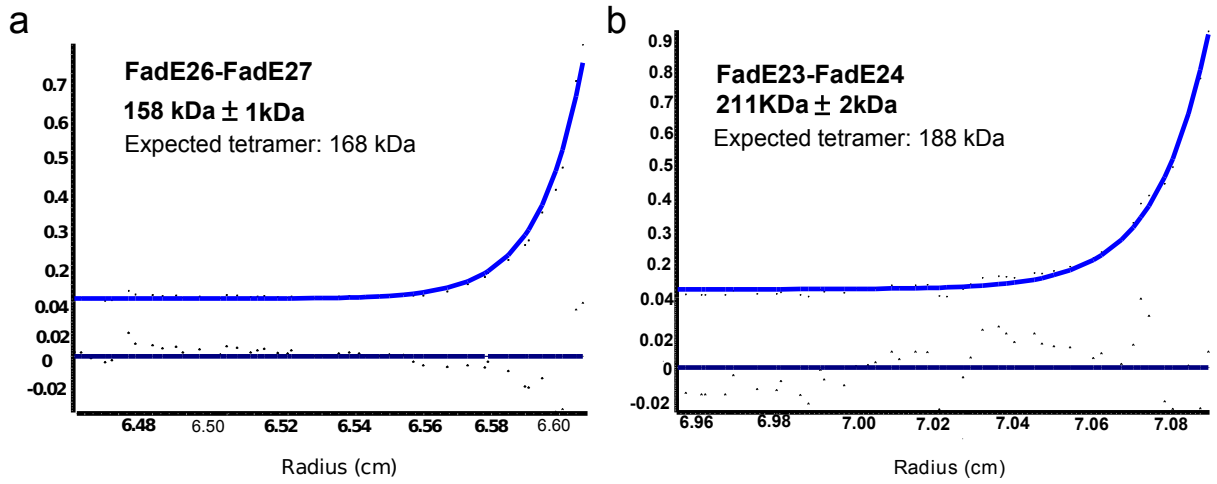


Figure 2-9. Analytical ultracentrifugation sedimentation equilibrium data for FadE26-FadE27 complex and FadE23-FadE24 complex, respectively. (a) FadE26-FadE27 (5.18 μM, 2.28 μM, and 1.08 μM) were centrifuged at speeds of 20k, 25k, and 30k rpm at 20° C in an analytical ultracentrifuge. (b) FadE23-FadE24 (5.75 μM, 3.14μM, 1.61 μM) were centrifuged at speeds of 20k, 25k, and 30k rpm at 4° C in an analytical ultracentrifuge. A representative fit for each sample is shown. The solid line (light blue) shows the fit of the data to the ideal species model and the residuals of the fit are graphed below the fit (dark blue). The best global fit for each protein is shown. Experimentally determined molecular weights are shown for each sample.

The FAD stoichiometry for each holoprotein complex was determined. Protein complexes were denatured and protein was separated from FAD by centrifugation. The concentrations of protein and FAD were determined spectroscopically at 280 nm and 260 nm, respectively (Table

2-4). The results indicate that for all of the heteromeric *Mtb* cholesterol-regulated ACAD enzyme complexes, there are two FAD cofactors bound per $\alpha_2\beta_2$ tetramer.

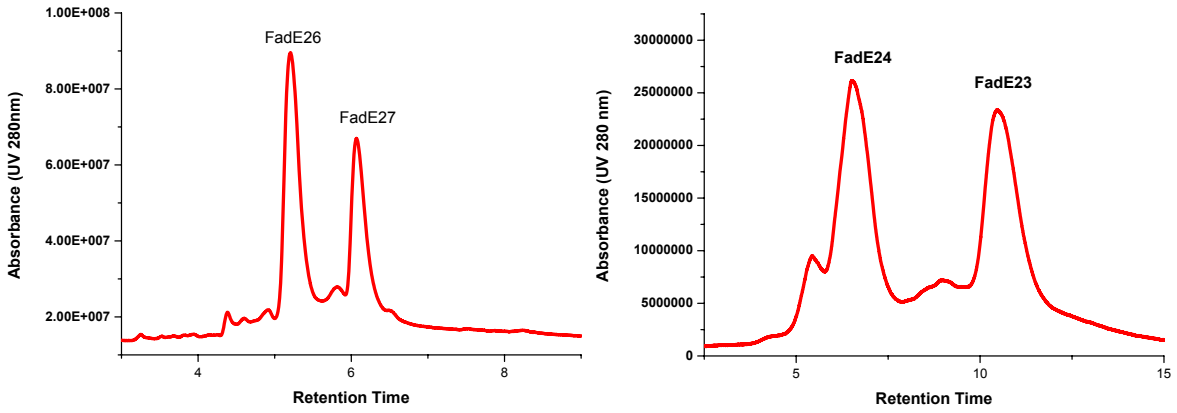


Figure 2-10. Reverse phase LC–UV chromatogram of FadE26–FadE27 and FadE23-FadE24. Peaks were identified by deconvolution of multiple charged states in the corresponding ESI+ MS spectra. The absorbance peaks were integrated and relative concentrations determined from the calculated extinction coefficients of FadE26, FadE27, FadE23, and FadE24. [ε_{280nm}(FadE26) = 68995 M^{-1} cm^{-1}, ε_{280nm}(FadE27) = 36565 M^{-1} cm^{-1}, ε_{280nm}(FadE23) = 43430 M^{-1} cm^{-1} and ε_{280nm}(FadE24) = 40068 M^{-1} cm^{-1}].

Purified FadE23-FadE24, FadE26-FadE27, ChsE1-ChsE2, FadE31-FadE32, FadE31-FadE33, and FadE13 complexes were assayed for acyl-CoA dehydrogenase activity with 3β-hydroxy-chol-5-en-24-oyl-CoA. This substrate was used as an analog of the 5-carbon side chain degradation intermediate. FadE26-FadE27 catalyzed the dehydrogenation of 3β-hydroxy-chol-5-en-24-oyl-CoA, whereas FadE23-FadE24, FadE31-FadE32, FadE31-FadE33, and FadE13 and did not (Figure 2-11).

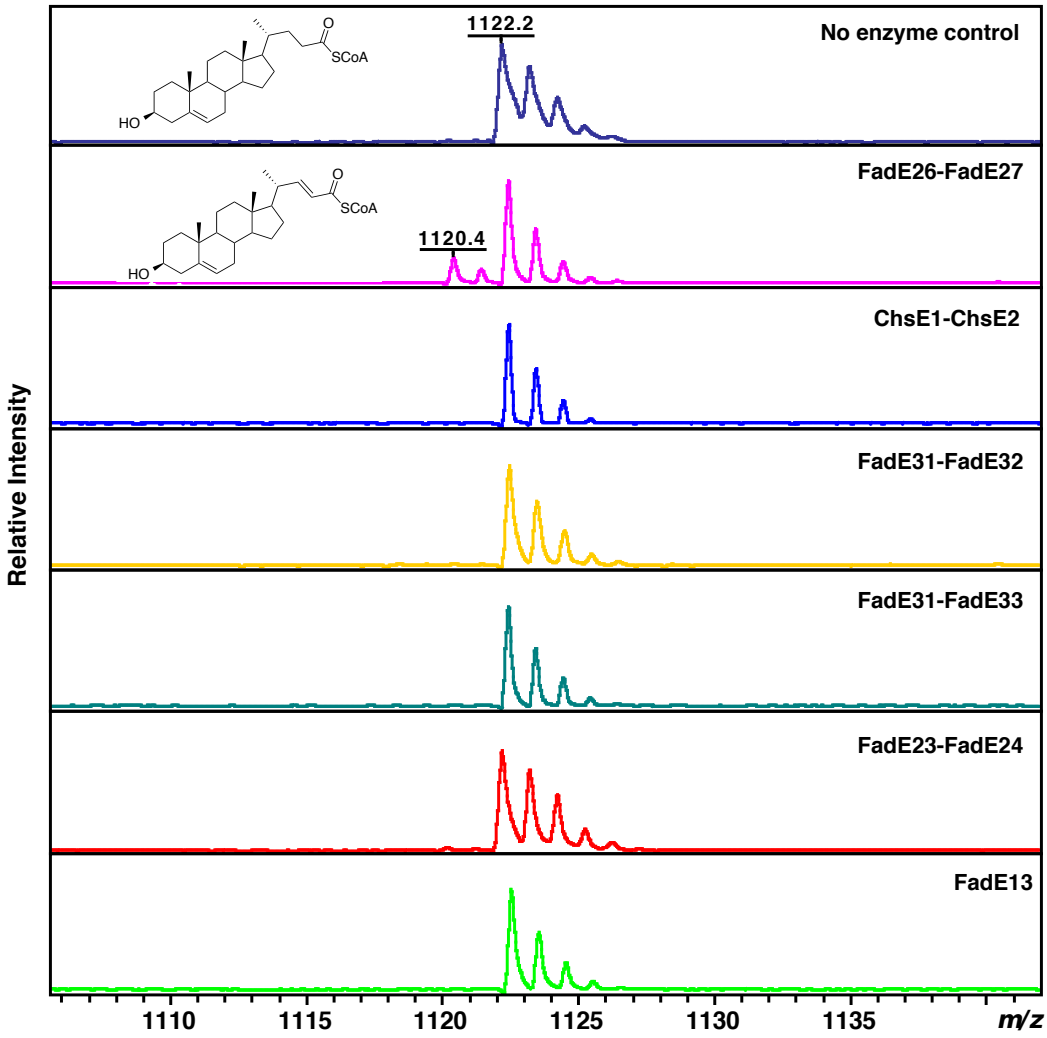


Figure 2-11. MALDI-TOF MS spectra of activity assay data for ACADs. ACAD complexes FadE26-FadE27, FadE23-FadE24, ChsE1-ChsE2, FadE31-FadE32, FadE31-FadE33, and FadE13 (100 nM) were assayed for dehydrogenase activity with ferrocenium hexafluorophosphate (250 μM) as an artificial electron acceptor and 100 μM 3-hydroxyl-chol-5-en-24-oyl-CoA $(M-H)^-$ = 1122.4) in 100 mM HEPES buffer, pH 7.4. The formation of the product was followed spectrophotometrically at 300 nm at 25 °C. Product formation ($(M-H)^-$ = 1120.4) was confirmed by MALDI-TOF mass spectrometry.

Heterotetrameric ACAD structures are present in other Actinobacteria and Proteobacteria. Within the bacterial kingdom, we found that several phyla contain homologs of acyl-CoA dehydrogenases that also are paired with a second *fadE* gene in the same operon, most notably Actinobacteria, Proteobacteria (Figure 2-12). We found that nearly all genomes

contained a cholesterol oxidase or 3β-HSD ortholog and in most cases contained a KshA ortholog as well, which suggests that the presence of $\alpha_2\beta_2$ ACADs correlates with the catabolism of sterols. Two consecutive acyl-CoA dehydrogenase genes were not identified in plants, fungi, or animals consistent with the different genomic structures present in bacteria versus other kingdoms.

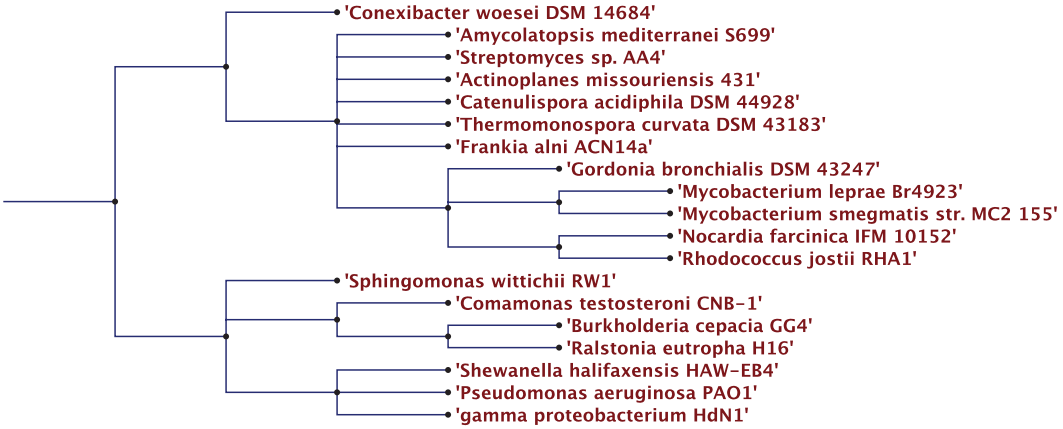


Figure 2-12. Rooted phylogenetic tree of taxonomically representative organisms with acyl-CoA dehydrogenase pairs. The organisms on this tree were used to generate a network of acyl-CoA dehydrogenases. This tree was generated using the NCBI Taxonomy Browser, and edited in CLC Sequence Viewer 6.

In order to visualize the operon relationships of these genes, a network analysis was performed (Figures 2-13), in addition to a traditional evolutionary tree (Figure 2-12). The *fadEs* cluster into four sets of similar genes (Figure 2-13a and 2-13b, groups I-IV). The proteins encoded by group I genes (group I: *fadE17, fadE26, chsE2, fadE31*) all contain the active site base, glutamate, in the sequence alignments. The polypeptides that pair with group I form two distinct clusters, group II: *fadE27, chsE1,* and group III: *fadE18, fadE32, fadE33*. The *fadE23* and *fadE24* genes and their homologs cluster independently (group IV). From this representation, it is clear that each set of the operonic pairs is conserved and that pairs do not swap between operons in different species. This work was performed by Matthew Wipperman.

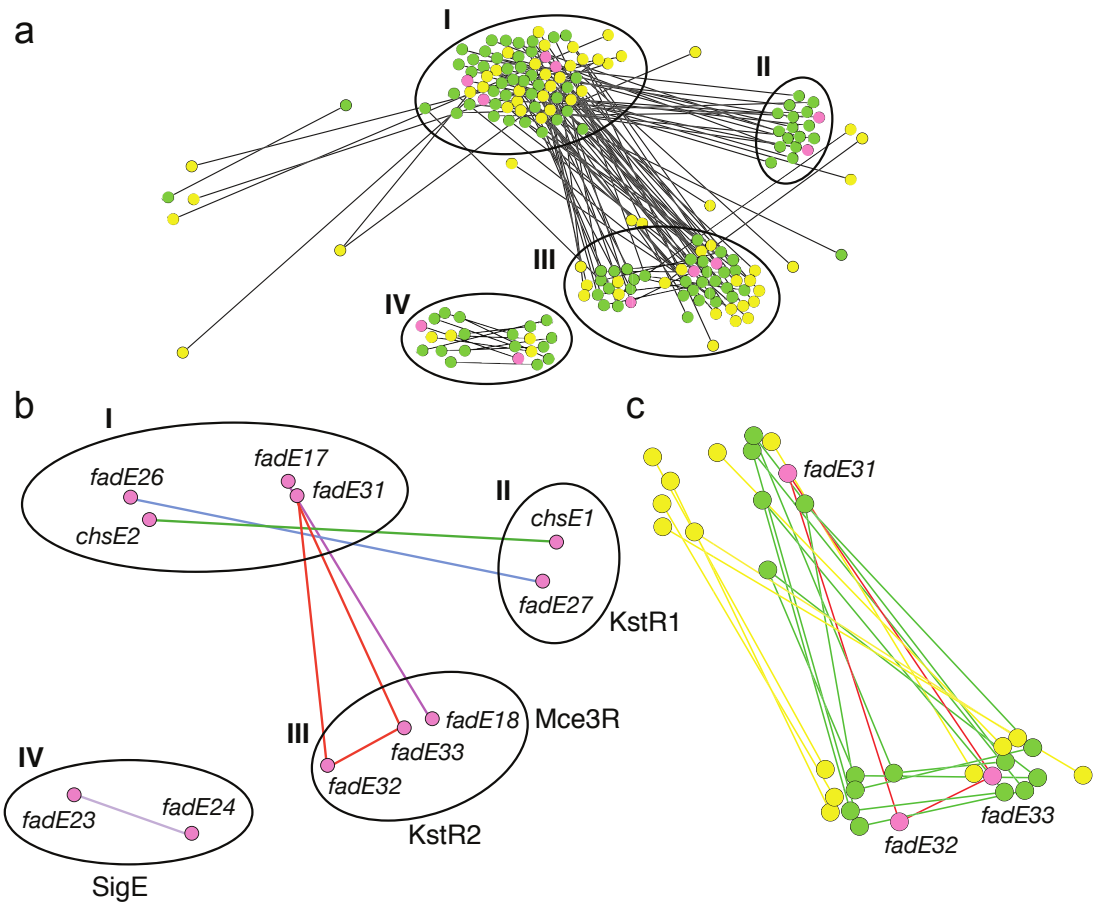


Figure 2-13. Phylogenetic network for bacterial ACAD genes homologous to cholesterol-regulated and proximal operonic *Mtb fadE* genes. Each node (circle) represents a gene. Distances between nodes represent relative evolutionary distance between two genes. The pink nodes are the 11 *Mtb fadE* genes that are regulated by cholesterol. The green nodes are non-*Mtb* Actinobacteria. The yellow nodes are Proteobacteria. Gene clusters I to IV referenced in the text are indicated as I to IV, respectively, in the figure. (A) Full phylogenetic network. (B) The 11 *Mtb fadE* genes present in operons are highlighted. (C) Magnification of the *fadE31-fadE32-fadE33* gene triplet.

4. Discussion

Almost all typical ACADs characterized to date have an α_4 homotetrameric quaternary structure with four FAD cofactors bound per tetramer (the exception being very long chain ACADs, which are homodimers). In earlier work, an unusual heterotetrameric ACAD that dehydrogenates the 2'-propionyl-CoA side chain of a sterol degradation intermediate was discovered.[56] Herein, we have identified five additional $\alpha_2\beta_2$ heterotetramers from *Mtb*, and characterized four of them. All four of these ACADs bind two molecules of FAD noncovalently and are expected to have two active sites per tetramer. Interestingly, the expression of all six $\alpha_2\beta_2$ *Mtb* ACADs identified within the *Mtb* genome are regulated by cholesterol.

We found that the adjacent *fadE* gene pair/ heterotetrameric ACAD motif is present in several different species of bacteria from the Actinobacteria and Proteobacteria phyla. These genera include Pseudomonas,[93] Burkholderia,[94] and Comamonas.[95] The network analysis demonstrates that operonic pairs encoding an $\alpha_2\beta_2$ protein architecture in *Mtb* are conserved in these phyla. Moreover, different pairings from those observed in *Mtb* were not identified. The conservation of operons suggests the $\alpha\beta$ protein interface is unique for each heterotetrameric complex and that α FadEs have a single β FadE partner. This specificity is further confirmed by the failure of new heteromeric pairings to form in an expression crossover experiment.

We used the presence of *choA* (also *choD* or *choE*), 3β-*hsd*, and/or *kshA* genes as markers for an organism's ability to metabolize steroids. We found these markers in the genomes of almost all of the organisms encoding putative $\alpha_2\beta_2$ ACADs. The presence of the $\alpha_2\beta_2$ ACAD motif in bacteria distantly related to *Mtb* (i.e., separate phyla) that also metabolize steroids supports the importance of the heterotetramer structure for binding steroids. Taken together, these data

suggest a modified ACAD architecture evolved for dehydrogenation of cycloalkane acyl-CoA substrates.

The ability of the Actinobacteria to adapt to a diverse set of environments is in part due to the complex regulation of a large number of genes, a subset of which have functions necessary for survival in each of those environments. In the *Mtb* genome, there are over 250 genes identified as involved in lipid, sterol, and fatty acid metabolism and there are many variants of TetR-like transcriptional repressors throughout the *Mtb* genome. A large subset of the genes (~80–90) in the *Mtb* cholesterol-regulated transcriptome are under control of two TetR-like transcriptional repressors,[96] KstR1 (also known as KstR)[46] and KstR2.[96]

The cholesterol-dependent KstR1 and KstR2 regulons regulate the expression of four of the FadE protein complex pairs: FadE26-FadE27, ChsE1-ChsE2, FadE31-FadE32, and FadE31-FadE33. Although up-regulated by cholesterol, FadE17-FadE18 expression is under control of the TetR-like Mce3R repressor.[51] In contrast, FadE23-FadE24 expression is down-regulated by cholesterol and requires SigE, a transcriptional activator, for expression.[90]

ChsE1-ChsE2 catalyzes dehydrogenation of the partially oxidized sterol side chain.[56] FadE26-FadE27 most likely catalyzes an analogous step (or steps) earlier in the β-oxidation pathway of sterol side-chain degradation based on its homology with ChsE1-ChsE2 and similar transcriptional regulation by KstR1.

The *fadE31-fadE32-fadE33* gene triplet from clusters I and III are highlighted in Figure 2-13c (pink edges). Although the triplet organization is conserved in Actinobacteria (green edges), the Proteobacterial homologs are organized as duplet gene assemblies. That is, each genome has two distinct orthologs of *fadE31*. Each of these orthologs is paired with either a *fadE32* or *fadE33*

gene in separate operons (yellow edges). These pairings are consistent with the oligomeric complexes isolated upon expression of the *Mtb fadE31-fadE32-fadE33* gene triplet.

The *fadE31-fadE32-fadE33* operon is regulated by KstR2. FadD3 from the KstR2 regulon catalyzes the formation of a CoA ester of the C/D ring intermediate of cholesterol degradation.[97] Thus the role of the proteins encoded in the KstR2 regulon is proposed to be catabolism of the C and D-rings of the degradation intermediate. Further assignment of function in this regulon awaits biochemical confirmation.

Based on our bioinformatics analysis, we hypothesize that *in vivo*, FadE17-FadE18 is an $\alpha_2\beta_2$ heterotetramer, although our attempts to isolate soluble FadE17-FadE18 have thus far been unsuccessful. *Mycobacterium bovis* (*M. bovis*) Type 17, in which the *echA13-fadE17-fadE18* operon is missing, is the predominate strain in the United Kingdom. In this strain, propionate flux shifts away from complex lipid synthesis toward pyruvate synthesis.[98] This shift is believed to provide an evolutionary advantage to *M. bovis,* since this strain naturally lacks two key genes required for pyruvate synthesis. The *M. bovis* Type 17 phenotype suggests that FadE17-FadE18 ties propionate production from cholesterol to complex lipid biosynthesis. Furthermore, sequences of the multi- and totally drug resistant strains of *Mtb*, R1207 and CTRI-4^{XDR} respectively, both contain single nucleotide polymorphisms (SNPs) in their *fadE17* gene sequences. How the SNPs alter FadE17-FadE18 function is not known.[99] Importantly, of the four mammalian cell entry (*mce*) regulons that have been shown to be important for virulence in *Mtb*, *fadE17-fadE18* belongs to *mce3* and is regulated by the Mce3R transcriptional repressor.[100] Homologs of *fadE17-fadE18* do not exist in many other closely related non-pathogenic Actinobacteria like *M. smegmatis,* suggesting a role in pathogenesis for this enzyme.

The *fadE23* and *fadE24* genes are most distantly related to the other ACAD homolog pairs. FadE23 and FadE24 are mysterious since they both contain a predicted active site glutamic acid from sequence alignment, yet only two FAD cofactors bind per tetramer. A recent activity-based protein profiling experiment identified both FadE23 and FadE24 as possibly having the ability to bind ATP,[101] hinting that the protein has a catalytic function distinct from acyl-CoA dehydrogenation. These ACAD genes are up-regulated in the presence of the first-line antimycobacterial drug isoniazid (INH). Genetic polymorphisms are associated with first- and second-line drug resistance in clinical *Mtb* isolates throughout the world. *fadE24* is one of the 22 identified genes that were associated with INH resistance.[102] INH causes increased cellular pools of fatty acids. The fact that both of the two genes are down-regulated by cholesterol is consistent with the proposition that *fadE23* and *fadE24* are involved in the recycling of fatty acids.[103] Moreover, SigE, which is required for survival in the host in order to counteract environmental stress from the host immune system, regulates the expression of the *fadE23-fadE24* operon.[90]

Although there are 35 *fadE* genes that are currently annotated in the *Mtb* genome,[77] this work provides evidence that there are not 35 functional ACAD enzymes. Rather, two adjacent ACAD genes code for a single protein complex, and only this complex is able to bind the FAD cofactor necessary for catalysis. From the 11 cholesterol-regulated *fadE* genes described in this work, we have demonstrated the existence of 5 holoprotein complexes, and provide strong evidence for the existence of a 6th (FadE17-FadE18). Although there are 35 *fadE* ORFs in the *Mtb* genome, there are at most 30 functional ACAD protein complexes, thus shrinking the number of functional enzymes available for catalysis of acyl-CoA dehydrogenation.

In conclusion, we have identified a subfamily of genes that encode $\alpha_2\beta_2$ heterotetrameric ACADs and linked this unusual ACAD protein architecture to the dehydrogenation of steroid-

CoA thioester substrates. The transcriptional regulation of $\alpha_2\beta_2$ ACAD gene expression suggests that these ACADs function in four different sub-pathways of cholesterol metabolism. Two of these sub-pathways are cholesterol side chain catabolism, and cholesterol C/D ring catabolism as demonstrated by previous biochemical experiments. The transcriptional regulation of the $\alpha_2\beta_2$ ACAD genes in the remaining two sub-pathways suggests that elaboration of cholesterol metabolites plays a role in modulating *Mtb* propionate flux and *Mtb* response to the host immune system.

5. Acknowledgements

The constructs of *fadE17*, *fadE18*, *fadE31*, *fadE32*, *fadE33* and *fadE13* were made by Matthew Wipperman. Protein purifications and characterizations of FadE17-FadE18, FadE31-FadE32, FadE31-FadE33, and FadE13 were accomplished by Matthew Wipperman. Bioinformatic analysis of all the $\alpha_2\beta_2$ ACADs was accomplished by Matthew Wipperman.

Chapter III. Elucidating the ACADs evolved in cholesterol side chain degradation in *Mtb*

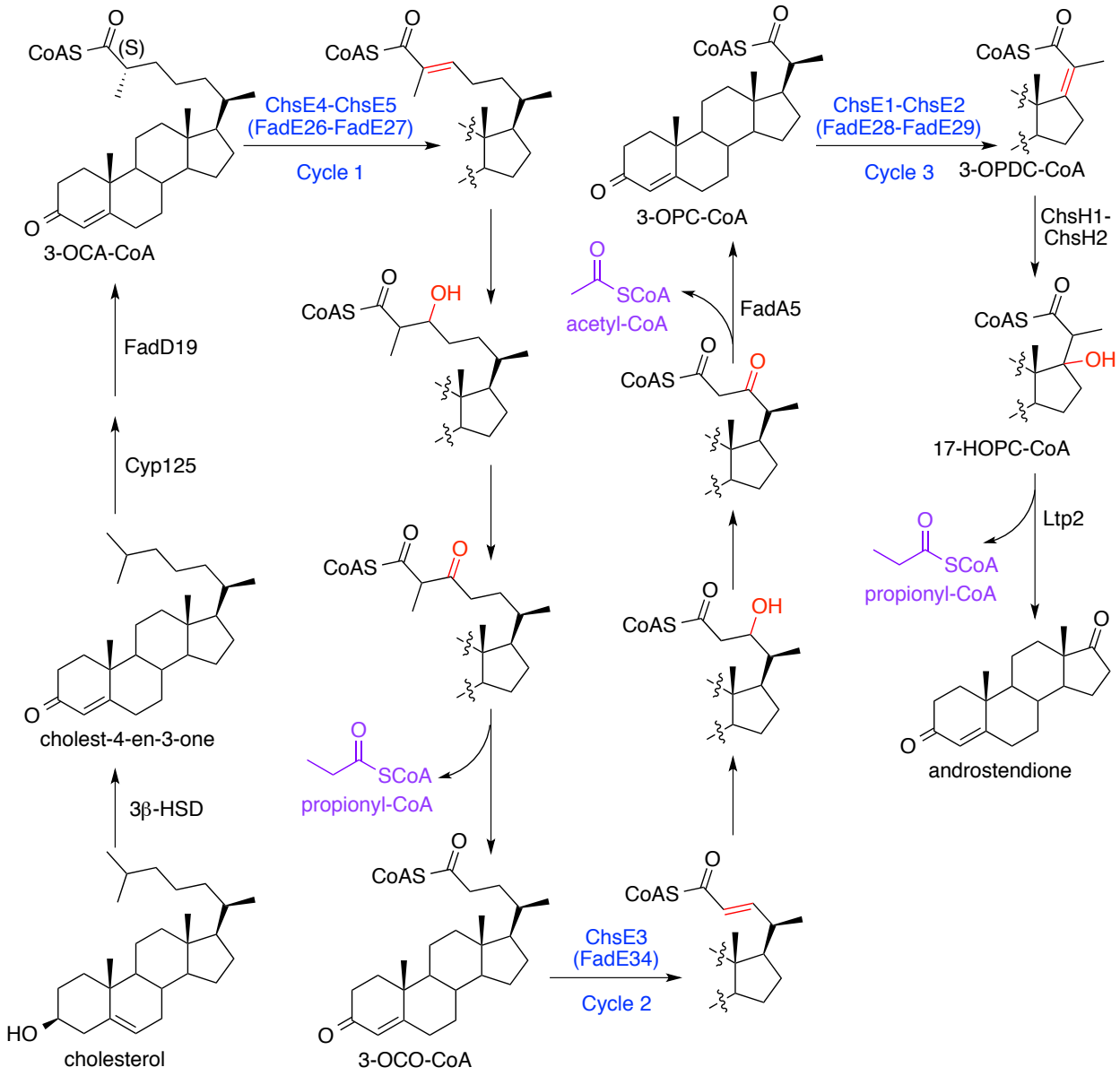
This chapter is adapted from a paper that was published in *ACS Infectious Diseases*[104]

1. Introduction

Throughout the course of *Mycobacterium tuberculosis* (*Mtb)* infection and Tuberculosis (TB) disease, the metabolism of lipids, including steroids like cholesterol, is preferred by the Mycobacteria. *Mtb* can degrade cholesterol *in vitro*, even as a sole carbon source, and *in vivo* the bacteria require cholesterol metabolism for establishing and maintaining chronic infection.[40, 49, 51, 62] Cholesterol metabolism provides *Mtb* with a source of acetyl-CoA and propionyl-CoA, which can be utilized for energy production.[32, 40] In addition, potentially valuable steroid-derived metabolites that might contribute to *Mtb*'s survival and virulence could be biosynthesized from cholesterol.

Degradation of the branched fatty acid-like side chain of cholesterol is proposed to proceed through β-oxidation,[105, 106] in a manner that is analogous to fatty acyl-CoA metabolism in human mitochondria and peroxisomes.[107] Initiation of β-oxidation requires conversion of cholesterol into cholest-4-ene-3-one, which is catalyzed by 3β-HSD,[108] followed by Cyp125-catalyzed oxidation of the C26 terminal methyl[54, 87, 109] to provide the initial fatty acid that is activated to the CoA thioester by FadD19.[110] The complete metabolism of the cholesterol side chain results in a 17-keto steroid intermediate, as well as one acetyl-CoA and two propionyl-CoA molecules (Scheme 3-1).[39] In the fatty acid β-oxidation cycle, ACADs typically have the lowest enzymatic throughput by several orders of magnitude, and therefore control overall reaction flux.[83] Identification of the specific ACAD enzymes that catalyze each step in the three cycles of β-oxidation is essential for determining the best target to block generation of propionyl-CoA and acetyl-CoA, and thus, energy production, from cholesterol. Moreover, inhibition of this activity would also potentially block the production of downstream steroid-derived metabolites with the

caveat that there are no published reports of 3-oxo-cholest-4-en-26-oyl-CoA (3-OCS-CoA) serving as a substrate for ring-degrading enzymes, however, future studies could potentially demonstrate such an activity.



Scheme 3-1. Cholesterol Side Chain Metabolism in *Mycobacterium tuberculosis*. Cyp125 initiates cholesterol side chain degradation. FadD19 acyl-CoA ligase activates the resultant steroid carboxylic acid through esterification with CoA. The steroid side chain is truncated via three cycles of β-oxidation to yield one acetyl-CoA and two propionyl-CoAs (highlighted in purple) and androstendione. The first step in each β-oxidation cycle is labeled with the acyl-CoA

dehydrogenase that catalyzes the oxidation and highlighted in blue text. The bonds undergoing modification are highlighted in red. Additional steps are labeled with specific enzyme names, if known.

Work from our laboratory has demonstrated that the KstR1-regulated *igr* (intracellular growth) operon encodes the enzymes that catalyze the final β-oxidation cycle in cholesterol side-chain degradation that removes the C20 to C22 propionate moiety of the cholesterol side chain (Scheme 3-1).[50, 56, 111] ChsE1-ChsE2 catalyzes the dehydrogenation of 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA) into 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA).[50, 56]

We discovered that ChsE1-ChsE2 is an $\alpha_2\beta_2$ heterotetramer. This $\alpha_2\beta_2$ structural architecture has, thus far, only been found in bacteria known to metabolize sterols. With the insights gained from study of ChsE1-ChsE2, we further identified additional $\alpha_2\beta_2$ heterotetrameric ACADs encoded in the *Mtb* genome, including FadE17-FadE18, FadE23-FadE24, FadE26-FadE27, FadE31-FadE32 and FadE31-FadE33.[68] All of these ACADs are encoded by genes that reside in single operons, and are therefore expressed polycistronically *in vivo*.[68] Furthermore, the expression of all of the $\alpha_2\beta_2$ ACAD enzymes from *Mtb* are regulated by cholesterol.[51]

In addition to being induced by cholesterol in *Mtb,*[51] four of the heterotetramer-encoding ACAD genes, *fadE26, fadE27, chsE1*, and *chsE2*, as well as a fifth *acad* gene, *fadE34,* are in the KstR1 regulon.[46] (Figure 3-2a) In a related Actinobacteria, *G. neofelifaecis*, these five *acad* genes are induced by cholesterol but not androstenedione, the sterol metabolite formed after complete removal of the side chain.[112] The demonstrated catalytic activity of ChsE1-ChsE2[56], the requirement of a sterol side chain for induction,[112] and their repression by KstR1[46, 96] suggests these five *acad* genes encode the three ACADs that catalyze the first step of side chain dehydrogenation in the three β-oxidation cycles of cholesterol side chain catabolism (Scheme 3-

1). Therefore, we investigated the catalytic activities of the *fadE26, fadE27, and fadE34* gene products. We employed biophysical characterization, substrate synthesis, and steady-state kinetics to determine which ACADs dehydrogenate the 5-carbon and 8-carbon cholesterol side chain metabolic intermediates, and the degree to which the substrate specificities overlapped. Our data clearly demonstrate that FadE34 is the ACAD responsible for catalyzing dehydrogenation in the second cycle of cholesterol side chain β-oxidation, and that FadE26-FadE27 is the ACAD in the first cycle of β-oxidation. With an established function in cholesterol side chain degradation, we now refer to FadE34 and FadE26-FadE27 as ChsE3 and ChsE4-ChsE5, respectively, to distinguish them from the FadE (fatty acid degradation E) acyl-CoA dehydrogenase subfamily.

The ChsE4-ChsE5 activity profile provides insight into compensatory activities that may contribute to the *in vivo* phenotype of the *igr* mutant. The X-ray crystal structure of ChsE4-ChsE5 reveals the apparent evolutionary relationship with the canonical homotetrameric ACADs and key differences between them. Binding site features of the $\alpha_2\beta_2$ ACAD distinguish it from the mammalian host homotetrameric structure[84] and will provide guidance for rational inhibitor design.

2. Experimental methods

Materials, strains, media, and general methods. Ferricenium hexafluorophosphate was purchased from Sigma-Aldrich (St. Louis, MO). Stigmasterol and Coenzyme A were purchased from MP Biomedicals (Solon, Ohio). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Denville Scientific (Metuchen, NJ). Tryptone, HEPES, and TRIS were purchased from Fisher Scientific (Pittsburgh, PA). Kanamycin was purchased from IBI Scientific (Peosta, IA). Yeast extract was purchased from Research Products International Co. (Mount Prospect, IL). iProof DNA polymerase was purchased from Bio-Rad (Hercules, CA). Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, and protein ladder were purchased from New England Biolabs (Beverly, MA). HisTrap FF columns and Superdex 200 HiLoad 16/60 and 10/300 GL columns were purchased from GE Healthcare Biosciences Corp. (Piscataway, NJ). Oligonucleotides were purchased from IDT Inc. (Coralville, IA). Total genomic DNA of *M. tuberculosis* H37Rv was obtained from the TB Research Materials Facility at Colorado State University (Fort Collins, CO) (NIAD NO1-AI40091). MALDI mass spectra were acquired on a Bruker Autoflex II TOF/TOF. Big Dye DNA sequencing (Applied Biosystems, Foster City, CA; performed by the Stony Brook University Sequencing Facility) was used to verify the coding sequence of the expression plasmids. BL21 (DE3) *E. coli* was obtained from BioRad. Protein expression 2 × YT media is composed of 16 g tryptone, 10 g yeast extract and 5 g NaCl per liter. Protein expression M9 minimal media is composed of 12.8 g Na_2HPO_4• $7H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 1g NH_4Cl, 2 mM MgSO4, 0.4% glucose (w/v) and 100 μM $CaCl_2$ per liter. Cell disruption was performed using a Constant Systems, Inc. TS Series Benchtop instrument (Kennesaw, Georgia). Buffer A: 20 mM Tris-HCl buffer pH 8.0, supplemented with 300 mM NaCl and 10 mM imidazole. Buffer B: 20 mM Tris-HCl buffer pH

8.0, supplemented with 300 mM NaCl and 500 mM imidazole. Buffer C: 50 mM Tris-HCl buffer pH 8.0, supplemented with 200 mM NaCl. Primary crystallization screens were from Hampton research. LC/MS analysis and ESI-MS were conducted on a Waters UPLC/MS instrument with diode array and SQD detectors. UV-visible spectra were acquired on a Shimadzu UV2550 UV/visible light spectrophotometer. NMR data were acquired on 400-MHz, 500-MHz and 700-MHz Bruker spectrometers. Chemical shifts are denoted in ppm (δ) and calibrated from residual undeuterated solvents. Assigned resonances are reported, and the full spectra are provided in the Supplementary Information. Flash chromatography was carried out on a CombiFlash Rf 200 (Teledyne Isco, Lincoln, NE) system. Mutagenesis was carried out using a QuikChange® II XL Site-Directed Mutagenesis Kit (Agilent Technologies, CA).

Expression plasmid construction. Genes *chsE4* (*Rv3504*), *chsE5* (*Rv3505*), and *chsE3* (*Rv3573c*) 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 pET28b or pET20b vector (Table 3-1). DNA sequencing of the plasmids confirmed that the sequence was correct and that no mutations were introduced during cloning.

ChsE4 Glutamate 247 was mutated to alanine in *p26N/27* using the method of quick change site-directed mutagenesis. Briefly, the following primers, the mutated nucleotide colored in red, are designed to introduce the expected mutation. After thermal cycling amplification, 1 μL of the DpnI restriction enzyme (10 U/μL) was added to digest the parental strands. 2 μL of the DpnI treated DNA product was transferred into XL1-Blue competent cells. The mutations were confirmed by DNA sequencing.

Forward primer: 5'-GCTCAACCACGCGCGGGTCGCCC-3'

Reverse primer: 5'-GGGCGACCCGCGCGTGGTTGAGC-3'

Table 3-1. Gene constructs used in this study.

construct	vector	gene	gene name	*His_6 tag*	*selection*	*source or ref*
p26N/27	*pET28b*	*Rv3504/Rv3505*	*ChsE4/ChsE5*	*N-terminal*	*Kan*	[68]
p26N	*pET28b*	*Rv3504*	*ChsE4*	*N-terminal*	*Kan*	*This work*
p27	*pET20b*	*Rv3505*	*ChsE5*	*no tag*	*Amp*	*This work*
p34N	*pET28b*	*Rv3573c*	*ChsE3*	*N-terminal*	*Kan*	*This work*
$p26N_{E247A}/27$	*pET28b*	*Rv3504/Rv3505*	*ChsE4/ChsE5*	*N-terminal*	*Kan*	*This work*

***chsE4* and *chsE5* gene co-expression and protein purification**. BL21(DE3) *E. coli* cells were co-transformed with constructs *p26N* and *p27* for *in trans* expression. Gene co-expression *in cis* with construct *p26N/27* was performed as reported previously.[68] (Chapter II) Similarly, the $ChsE4_{E247A}$-ChsE5 mutant protein was prepared with construct *$p26N_{E247A}/27$* using the same expression conditions that were used for *p26N/27*. Expression was induced at OD_{600} ~ 0.6 – 0.8 by the addition of 1 mM IPTG and cells were grown 20 h at 25 °C. Cells were lysed by cell disruption in Buffer A and cellular debris was removed by ultracentrifugation at 125,000 × g for 1 h. Both ChsE4-ChsE5 complex and $ChsE4_{E247A}$-ChsE5 were purified by IMAC, with a HisTrap FF column using Buffers A for binding and B for elution. Protein was further purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with Buffer C.

Protein expression and purification. To express ChsE3, BL21(DE3) *E. coli* cells were transformed with construct *p34N*. Single colonies were selected on LB plates containing 30

μg/mL kanamycin and cultured in 2 × YT media at 37 °C. Protein expression was induced at an OD_{600} ~ 0.6 – 0.8 by the addition of 1 mM IPTG, and cells were grown for 20 h at 25 °C. Purified proteins were analyzed by reducing SDS-PAGE. The protein was purified using a similar method as described for ChsE4-ChsE5 purification.

Selenomethionine substituted ChsE4 and ChsE5 co-expression and protein purification. BL21 (DE3) *E.coli* cells were transformed with construct *p26N/27* for *in cis* expression. Single colonies were selected on LB plates containing 30 μg/mL kanamycin and cultured in M9 minimal media supplemented with 30 μg/mL kanamycin. When the OD_{600} value reached ~1.15, 100 mg/L Lys, Phe, Thr; 50 mg/L Ile, Leu, Val; and 60 mg/L of Selenomethionine (SeMet) were added into M9 media, then cells were grown at 18 °C for 20 h. SeMet substituted ChsE4-ChsE5 protein was purified in the same manner as the native protein.

Solution-state biophysical analysis of ChsE3. Molecular weights were determined using analytical ultracentrifugation sedimentation equilibrium (Beckman Optima XL-A). ChsE3 (6.09 μM, 3.05 μM, and 1.52 μM) was centrifuged at speeds of 20k, 25k, and 30k at 20 °C. Scans were acquired after 18 and 20 h of centrifugation at each speed monitoring at 280 nm. Approximately 80 curves generated from replicate scans were fit globally in Heteroanalysis (University of Connecticut, Storrs, CT) to the ideal single-species model. The sample buffer density (ρ) and partial specific volume (ν) of each protein were calculated using SEDNTERP (University of New Hampshire, Durham, NH).

Determination of the protein to FAD ratio. ChsE3 to FAD ratio were determined using the method described previously (Chapter II).[68]

Synthesis of octanoyl-CoA. Octanoyl-CoA were synthesized using the mixed anhydride method.[113] Briefly, the mixed anhydride was prepared by mixing the corresponding acid (0.2 mmol) with ethyl chloroformate (0.4 mmol) in 4 mL dry THF in the presence of TEA (56 μL). After approximately 25 min, the acid was completely converted into the corresponding anhydride according to TLC. The mixed anhydride was filtered though glass wool in a disposable Pasteur pipet into newly prepared lithium CoA solution (55 mg) dissolved in 5 mL of H_2O and THF (3:2 v/v) (pH 8). The reaction was stirred at rt for 2 days. The pH of the reaction mixture was adjusted to 3 and the unreacted acid was removed by extraction with ether. The enoyl-CoAs were purified by HPLC using 20 mM ammonium bicarbonate, pH 8 with a linear gradient from 0 to 90% methanol in 20 mM ammonium bicarbonate.

Synthesis of 3-oxo-4-pregnadiene-20-carboxylic acid. 3-Oxo-4-pregna-diene-20-carboxylic acid was synthesized following a previously reported method starting from stigmasta-4, 22-diene-3-one.[114] 3-Stigmasta-4, 22-diene-3-one (1 g) was dissolved in 200 mL CH_2Cl_2 and cooled to -78 °C. 1% (v/v) pyridine was added to selectively ozonolyze the double bond at the 22 position. O_2 was purged from the solution for 5 min, ozone was sparged until the solution turned blue, then O_2 was sparged until the blue color dissipated. Methyl disulfide was added and reaction was warmed to rt and allowed to stand for 12 h. The resulting aldehyde was purified by silica gel chromatography. Sodium chlorite (10 eq) was added to the aldehyde (dissolved in 10% H_2O/ACN) and stirred for 16 h. The reaction was monitored by TLC and 3-oxo-4-pregnadiene-20-carboxylic acid was purified by silica gel chromatography.

Synthesis of 3-oxo-4-pregnene-20-carboxyl CoA, 3-OPC-CoA. 3-oxo-4-pregnene-20-carboxylic acid (3-OPC) was dissolved to 50 mM in 94% ethanol containing 60 mM NaOH. The thioesterification was performed for 4 h at 22 °C in a 1 mL reaction volume containing 5 mM 3-

OPC, 10 μM CasI (Ro05822), 100 mM pH 7.4 HEPES, 5 mM $MgCl_2$, 2.5 mM ATP, and 1 mM CoASH.[97] The reaction was quenched by adding 1 mL of MeOH to precipitate the enzyme. The supernatant was isolated by centrifugation and filtration. 3-oxo-4-pregnadiene-20-carboxyl CoA was purified by HPLC using 100 mM ammonium acetate, pH 4.5 with a linear gradient from 0 to 90% MeOH in 100 mM ammonium acetate to yield 3-OPC-CoA.

Dehydrogenase assay. The dehydrogenase activities of ChsE1-ChsE2, ChsE3, and ChsE4-ChsE5 (50 mM) were tested individually with 3-OPC-CoA, 3-OCO-CoA and 3-OCS-CoA (1–100 μM) with the artificial electron acceptor ferricenium hexafluorophosphate (250 μM, the extinction coefficient is 4.3 mM^{-1} cm^{-1})[89] in 100 mM TAPS buffer (pH 8.5) as reported previously.[56] Assays were initiated by the addition of enzyme. Product formation was monitored spectroscopically at 300 nm at 25 °C. Initial velocities were obtained for the first 10-15% of the reaction. Controls were run without enzyme or without substrate, and showed negligible decreases in absorbance at 300 nm. The rates of product formation were fit to the Michaelis-Menten equation to determine K_M and k_{cat} for each substrate, except in the case of ChsE4-ChsE5/3-OCS-CoA for which diastereomer inhibition was observed at high concentrations. In that case, data were fit to uncompetitive inhibition equation (1) for which [S] = [I] = the concentration of a single diastereomer:

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

The effects of propionyl-CoA on ChsE4-ChsE5 dehydrogenase activity. In order to test whether propionyl-CoA plays a regulatory role on ChsE4-ChsE5's dehydrogenase activity since it is a commonly observed regulatory molecule, different concentrations (from 12.1 μM to 121 μM) of propionyl-CoA were used to investigate its influence ChsE4-ChsE5's specificity on low concentration of 3-OCS-CoA (5 μM) and high concentration of 3-OCS-CoA (25 μM),

respectively. Briefly, different concentrations of propionyl-CoA were incubated with artificial electron acceptor ferricenium hexafluorophosphate (250 μM, the extinction coefficient is 4.3 mM^{-1} cm^{-1}), a defined concentration of 3-OCS-CoA (5 μM or 25 μM) in 100 mM TAPS buffer (pH 8.5) for 15 min at 25 °C, then assays were initiated by the addition of enzyme. Product formation was monitored spectroscopically at 300 nm at 25 °C. Initial velocities were obtained for the first 10-15% of the reaction.

Protein crystallization. The ChsE4-ChsE5 apoenzyme crystals were obtained by hanging drop vapor diffusion at room temperature. Briefly, 1 μL of 12 mg ml^{-1} protein was mixed 1:1 with a reservoir solution of 25% v/v polyethylene glycol monomethyl ether 550, 0.005 M $MgCl_2$ • $6H_2O$, 0.05 M HEPES sodium pH 7.0 and equilibrated against 500 μL of the reservoir solution. The crystals were then harvested and transferred to a cryoprotectant solution containing 27% v/v polyethylene glycol monomethyl ether 550, 0.005 M $MgCl_2$ •$6H_2O$, 0.05 M HEPES sodium pH 7.0, and 10% glycerol. All crystals were cryocooled in liquid N_2 before data collection.

X-ray data collection and structural determination. Diffraction data were collected on Beamline X25 of the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY) at the selenium absorption edge, using a wavelength of 0.9792 Å at 100 K. Datasets were processed using XDS[115] and Aimless[116] as implemented in the autoPROC pipeline[117]. All twenty-six selenium sites were located with SHELXD.[118] Phases were calculated to 2.0 Å via single-wavelength anomalous dispersion[119] using Sharp.[120] An initial round of density modification and automated model building was carried out with Solomon,[121] Buccaneer,[122] and Parrot.[123] The resulting map and model was further improved in a second round of density modification and automated model building in Solomon and ARP/wARP. This procedure yielded a starting model that was 94% complete (1408 of 1496 residues correctly built). Manual model

adjustments were carried out in Coot[124] followed by refinement with Refmac[125] and Phenix.[126] The geometric quality of the refined model was assessed with MolProbity[127] and the structure validation tools in the Phenix suite. Data collection and refinement statistics are shown in Table 3-2. This work was performed in collaboration with Dr. Kip Guja.

Single-crystal optical spectra collection during the course of X-ray diffraction data collection. Single-crystal optical spectra for ChsE4-ChsE5 crystal were collected on Beamline X26-C of the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY) using X-ray exposure at a wavelength of 1.0 Å. With 72 different optical spectra taken every 5° by rotating the crystal 360°, the optimal spectroscopic angle for the cryoloop and crystal was defined using the software CBASS.[128] X-ray diffraction collection started with 180 × 1℃ rotation to acquire a full data set. During the readout of X-ray diffraction images, correlated electronic absorption spectra were collected with the crystal rotating back to the defined optimal angle. A set of optical spectra were overlayed and compared with the cumulative X-ray exposure time after the full data collection.[129]

Docking of 3-oxo-cholest-4-en-26-oyl-CoA (3-OCS-CoA) into the crystal structure of ChsE4-ChsE5. The mol2 files of 3-OCS-CoA and 3-HCS-CoA were built using MOE. The crystal structure of a medium chain acyl-CoA dehydrogenase (MCAD) with octanoyl-CoA bound (PDB code: 3MDE) was superimposed onto the apo-ChsE4-ChsE5 structure in PyMOL (PyMOLX11Hybrid) using a backbone alignment. Octanoyl-CoA was extracted and merged with the apo-ChsE4-ChsE5 structure. The new complex was further analyzed in Chimera (UCSF Chimera) using default parameters of Dock Prep tool for docking the ligand (25*S*)-3-OCS-CoA and (25*S*)-3-HCS-CoA into the ChsE4-ChsE5 structure. To add charge to the ligand, AMBER ff99SB and gasteiger were used as the charge model and charge method, respectively. SPHGEN

was used to generate receptor spheres. During docking, the conserved position of Coenzyme A was fixed and used as a constraint. Dock6 was used to finish the docking. Consensus scoring and manual inspection were used to select docking poses further analysis.

3. Results and Discussion

The $\alpha_2\beta_2$ heterotetrameric acyl-CoA dehydrogenase ChsE1-ChsE2 functions in the last cycle of β-oxidation in cholesterol side chain degradation.[56] (Scheme 3-1) The *chsE1* and *chsE2* genes are part of the *igr* operon, which is regulated by the KstR1 repressor.[46, 96] We demonstrated that the unusual $\alpha_2\beta_2$ heterotetrameric ACAD assembly of ChsE4-ChsE5 catalyzes the dehydrogenation of 3-hydroxyl-chol-5-en-24-oyl-CoA, an analog of 3-OCO-CoA (chapter II).[68] We reasoned that the three additional *acad* genes regulated by the KstR1 repressor function in a related set of steps during cholesterol metabolism. Two of these genes, *chsE4* and *chsE5,* encode an $\alpha_2\beta_2$ ACAD.[68] The third gene, *chsE3* encodes a homolog of the very long chain ACAD subfamily. The unusual structural organization of two of these ACADs, the enzymatic activity of ChsE1-ChsE2 and their common regulation led to our hypothesis that the ACADs function in the first and second cycles of sterol side-chain β-oxidation. Therefore, we undertook elucidation of the enzymatic functions and structures of ChsE4-ChsE5 and ChsE3, as well as further substrate specificity testing of ChsE1-ChsE2.

Biophysical characterization of ChsE4-ChsE5 and ChsE3. ChsE4 and ChsE5 belong to the ACAD family. As described in chapter II, ChsE4-ChsE5 has been demonstrated as an $\alpha_2\beta_2$ heterotetrameric complex that binds two FAD cofactors, which is different compared to typical α_4 homotetrameric ACADs.[68]

ChsE3 (Rv3573c, formerly FadE34) also belongs to the ACAD family. However, the ChsE3 protein sequence is most similar to the very long chain acyl-CoA dehydrogenase (VLCAD) subfamily that forms homodimers rather than homotetramers.[84] ChsE3 was expressed as an N-terminally His_6-tagged protein in *E. coli* and purified by IMAC, then further purified by size

exclusion chromatography. The UV-vis spectrum of purified ChsE3 showed distinctive absorbance maxima at 370 nm and 440 nm, indicating the presence of bound FAD cofactor. Further analysis by sedimentation equilibrium analytical ultracentrifugation (AUC) revealed that ChsE3 forms an α_2 homodimer in solution (Figure 3-1). There are two FAD cofactors bound per α_2 dimer in ChsE3 as expected for a VLCAD subfamily member. This work was performed by Johnna R. St. Clair.

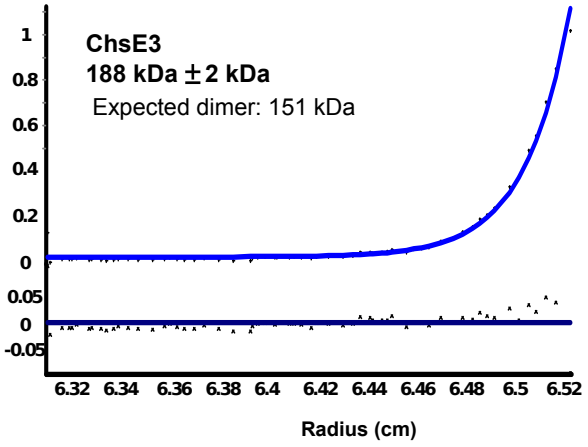


Figure 3-1. Analytical ultracentrifugation (AUC) sedimentation equilibrium analysis. ChsE3 (6.09 μM, 3.05 μM, and 1.52 μM) were centrifuged at speeds of 20k, 25k, and 30k rpm at 20° C in an analytical ultracentrifuge. The solid line (light blue) shows the fit of the data to the ideal species model and the residuals of the fit are graphed below the fit (dark blue). The best global fit for each protein is shown. Experimentally determined molecular weights are shown for each sample.

Preparation of the three acyl-CoA metabolic intermediates of cholesterol side chain b-oxidation. In order to test the enzymatic activity of the KstR1-regulated ACADs, we required their putative substrates.[50] Previous work with ChsE1-ChsE2 demonstrated a five-fold higher substrate specificity for the ring-intact 3-OPC-CoA ester compared to a 2-ring indanone CoA ester metabolite.[50] Therefore, we undertook synthesis of the 3-carbon, 5-carbon, and 8-carbon CoA ester metabolites with the 4-ring skeleton. We prepared (3-OPC-CoA) as previously described,[50] and developed a preparative method for the two additional substrates.

For the preparation of 3-oxo-chol-4-en-24-oyl-CoA (3-OCO-CoA) and 3-oxo-cholest-4-en-26-oyl-CoA (3-OCS-CoA), we developed synthetic methods to obtain these two compounds. This work was performed by Rui Lu.

ChsE4-ChsE5, ChsE3, and ChsE1-ChsE2 catalyze the three sequential acyl-CoA dehydrogenations required for cholesterol side chain β-oxidation. In earlier work, we demonstrated that ChsE1-ChsE2 catalyzes the dehydrogenation of 3-OPC-CoA, and that disruption of the *igr* operon encoding ChsE1-ChsE2 resulted in accumulation of the analogous methyl ester 3-carbon side chain intermediate.[50, 56] However, the identity of the ACADs that dehydrogenate the 5-carbon side chain and 8-carbon side chain cholesterol metabolites were unknown. Based on our hypothesis that the KstR1-regulated ACADs are responsible, we determined the steady-state rate constants for ChsE1-ChsE2, ChsE3, and ChsE4-ChsE5 with 3-OPC-CoA[56], 3-OCO-CoA, and 3-OCS-CoA as substrates (Table 3-3). Because the physiologic electron transfer protein is unknown, we used the ferricenium system[89] for oxidation of reduced ACADs.

As expected from the metabolism study,[50] ChsE1-ChsE2 possessed the highest catalytic activity with 3-OPC-CoA (Figure 3-2b). Although ChsE1-ChsE2 also catalyzed the dehydrogenation of 3-OCO-CoA, the apparent second-order rate constant k_{cat}/K_m was 10-fold lower than with 3-OPC-CoA as the substrate. When 3-OCS-CoA was used as a substrate, no catalytic activity was detected. These substrate specificity data further confirmed that ChsE1-ChsE2 is the ACAD responsible for the last cycle of β-oxidation in *Mtb* cholesterol side chain metabolism.[50, 56]

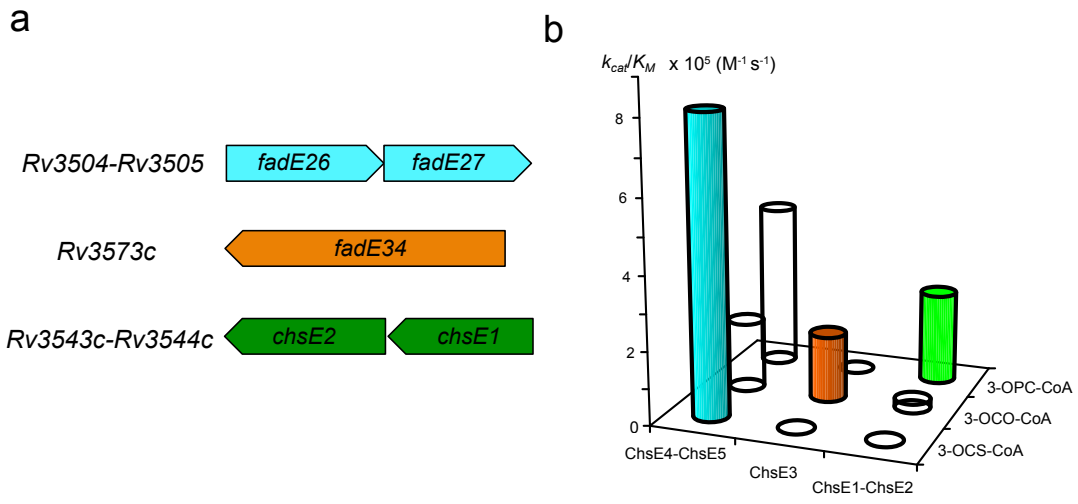


Figure 3-2. KstR1 regulated ACADs function in cholesterol side chain degradation. (a) *fadE* genes in *Mtb* genome that are regulated by KstR1 regulon. (b) Catalytic specificity for the KstR1-regulated acyl-CoA dehydrogenases. Plot of the catalytic specificity (k_{cat}/K_M) of ChsE4-ChsE5, ChsE3, and ChsE1-ChsE2 for the three acyl-CoA metabolic intermediates of cholesterol side chain b-oxidation, 3-OPC-CoA, 3-OCO-CoA and 3-OCS-CoA, respectively.

ChsE3 showed much narrower substrate specificity. The only substrate with which catalysis of dehydrogenation by ChsE3 was observed is 3-OCO-CoA (Table 3-3, Figure 3-2). No dehydrogenase activity was observed with the 3-carbon side chain (3-OPC-CoA) or the 8-carbon side chain (3-OCS-CoA) intermediates as judged by both steady-state spectroscopic assays and MALDI-TOF mass spectrometry of the reaction mixture. The data demonstrate that ChsE3 specifically functions in the second cycle of β-oxidation (Scheme 3-1).

Table 3-3. Steady-state kinetic parameters for ChsE1-ChsE2, ChsE3, and ChsE4-ChsE5 with the acyl-CoA thioesters of the cholesterol side chain degradation intermediates[a]

Substrate	ChsE1-ChsE2		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
3-OPC-CoA	5.3 ± 0.9[b]	1.03 ± 0.016[b]	(1.9 ± 0.32) ×10^5 [b]
3-OCO-CoA	6.5 ± 0.9	0.13 ± 0.004	(2.0 ± 0.28) ×10^4
3-OCS-CoA	na[c]	na[c]	na[c]

Substrate	ChsE3		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
3-OPC-CoA	na[c]	na[c]	na[c]
3-OCO-CoA	28 ± 8	4.0 ± 0.2	(1.4 ± 0.4) ×10^5
3-OCS-CoA	na[c]	na[c]	na[c]

Substrate	ChsE4-ChsE5		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
3-OPC-CoA	3.3 ± 0.7	1.2 ± 0.04	(3.6 ± 0.7) ×10^5
3-OCO-CoA	2.6 ± 0.4	0.38 ± 0.008	(1.5 ± 0.24) ×10^5
3-OCS-CoA	3.4 ± 1.0[d]	2.1 ± 0.3[d]	(6.2 ± 2.0) ×10^5
octanoyl-CoA	*Kiu*=20±2.7 4.1 ± 0.9	0.033 ± 0.0016	(8 ± 1.6) × 10^3

[a] Error bars are the standard deviations of global fits to three independent experiments. [b]Data from Ref 55. [c]No activity. [d]Inhibition was observed with 3-OCS-CoA. The data were fit assuming uncompetitive substrate inhibition.

In contrast, ChsE4-ChsE5 shows the broadest substrate specificity. All product formation was confirmed by MALDI-TOF mass spectrometry (Figure 3-3). Although ChsE4-ChsE5 catalyzes the dehydrogenation of all three acyl-CoA substrates: 3-OPC-CoA, 3-OCO-CoA, and 3-OCS-CoA, it catalyzes the dehydrogenation of 3-OCS-CoA most efficiently with a k_{cat}/K_M that is two to four times greater than for 3-OCO-CoA or 3-OPC-CoA, respectively (Figure 3-2b and Table 3-3). Although, ChsE4-ChsE5 can dehydrogenate octanoyl-CoA, a medium length fatty acyl ester chain, the k_{cat}/K_M is nearly 80-fold lower (Table 3). Notably, ChsE4-ChsE5 is the only ACAD of the three that accepts 3-OCS-CoA as a substrate. This substrate specificity is consistent with an indispensable role in the first cycle of β-oxidation of the 8-carbon side chain.

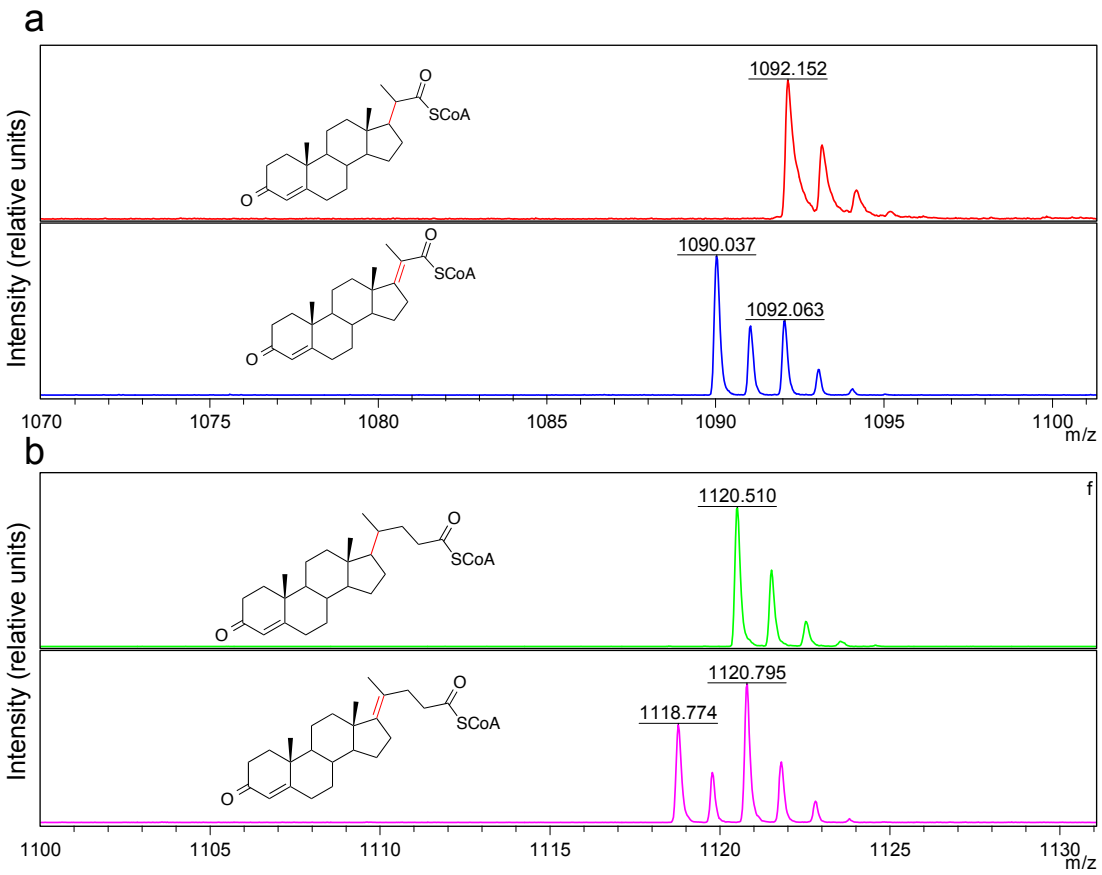


Figure 3-3. Analysis of product formation catalyzed by ChsE4-ChsE5 with 3-OPC-CoA (a) and 3-OCO-CoA (b).

Analysis of product formation by MALDI-TOF mass spectrometry revealed that the percent substrate conversion remained fixed at a 1:1 ratio of product to substrate, even after overnight incubation at 25 °C and with a 6-fold increase in ChsE4-ChsE5 concentration over that required for turnover (Figure 3-4). We reasoned that ChsE4-ChsE5 can dehydrogenate only one diastereomer of 3-OCS-CoA. Most ACADs typically catalyze dehydrogenation of the (*S*) α-methyl, acyl-CoA diastereomer.[130] In the case of ChsE4-ChsE5, the stereochemistry has not been elucidated. However, specific dehydrogenation of the (*25S*)-OCS-CoA would be consistent with the stereospecificity of Cyp125, the P450 enzyme that catalyzes oxidation of the C26 methyl of cholest-4-en-3-one two steps earlier in the pathway.[109]

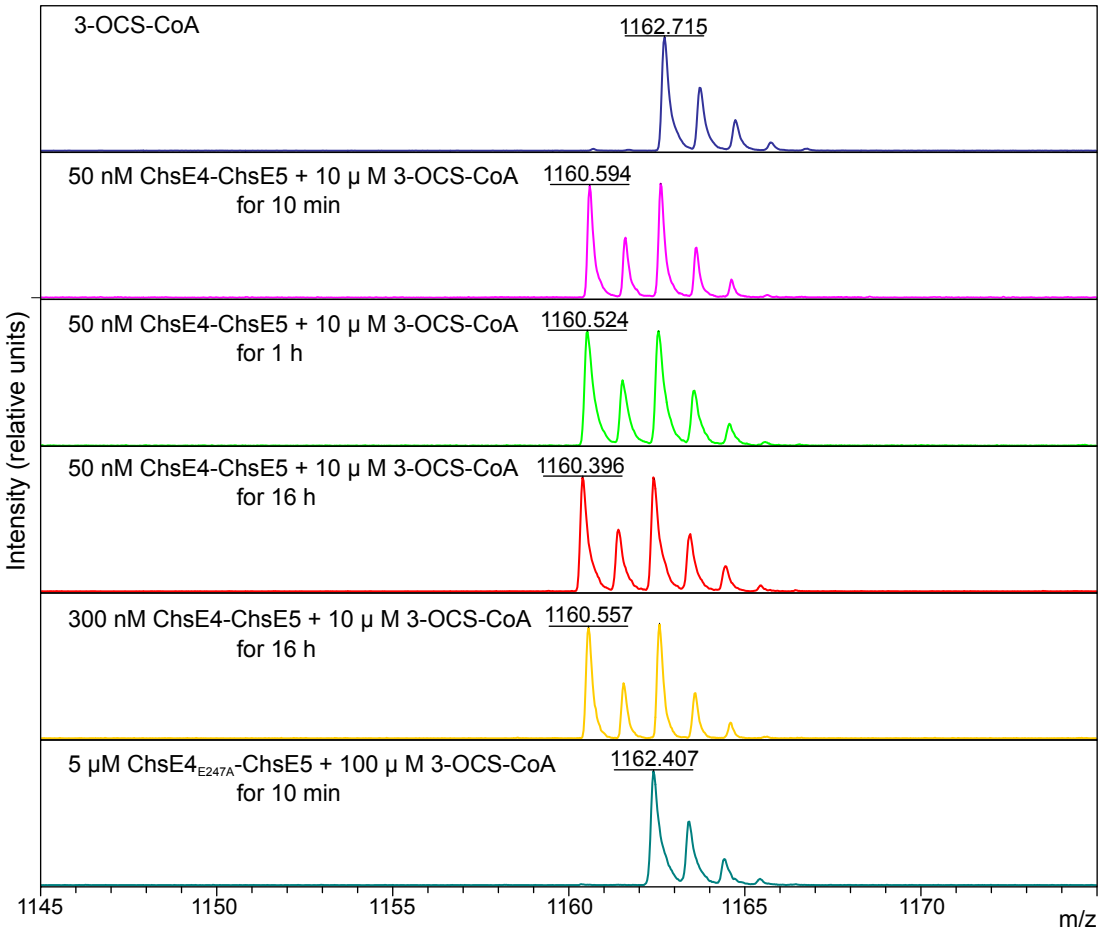


Figure 3-4. Analysis of product formation catalyzed by ChsE4-ChsE5 with 3-OCS-CoA diastereomers as the substrate. Analysis of the product formation of ChsE4-ChsE5 with 3-OCS-CoA diastereomers as a function of time and increased enzyme concentration. The spectra revealed that the percent substrate conversion remained fixed at a 1:1 ratio of product to substrate.

At high concentrations of 3-OCS-CoA, inhibition was observed. However, no inhibition was observed with either OPC-CoA or 3-OCO-CoA. We assume inhibition is due to the presence of unreactive 3-OCS-CoA diastereomer that binds to ChsE4-ChsE5 nonproductively. Therefore, the steady-state kinetic data were fit to a modified Michaelis-Menten equation that took uncompetitive substrate inhibition into account in order to estimate k_{cat} and K_M.

Propionyl-CoA has no influence on the dehydrogenase activity of ChsE4-ChsE5. Initial velocities of ChsE4-ChsE5 were calculated in the presence of different concentrations of

propionyl-CoA at either low concentration (5 μM) or high concentration (25 μM) of 3-OCS-CoA. The results showed that propionyl-CoA has negligible influence on the dehydrogenase activity of ChsE4-ChsE5 after comparing with the initial velocity in the absence of propionyl-CoA, suggesting propionyl-CoA does not regulate the activity of ChsE4-ChsE5 *in vitro* (Table 3-4).

Table 3-4. Investigation the regulation of propionyl-CoA on ChsE4-ChsE5 enzyme activity.

ChsE4-ChsE5 (nM)	3-OCS-CoA (μM)	Propionyl-CoA (μM)	Initial velocity (μM /min)[1]	Initial velocity (μM /min)
50	5	0	3.108	
50	5	12.1	3.433	
50	5	24.2	3.615	3.504[2] ±.0.221[3]
50	5	36.3	3.633	
50	5	121	3.733	
50	25	0	5.531	
50	25	12.1	5.274	5.362 ± 0.118
50	25	24.2	5.410	
50	25	36.3	5.231	

[1] For each concentration of propionyl-CoA, the initial velocity was measured once. [2] The average of initial velocities. [3] The standard deviation of initial velocities at different concentrations of propionyl-CoA

ChsE4-ChsE5 forms (*24E*)-3-oxo-cholest-4,24-dien-26-oyl-CoA. The ChsE4-ChsE5 reaction product was subjected to alkaline hydrolysis and the free acid purified by extraction. Comparison of ^{1}H spectra of 3-oxo-cholest-4-en-26-oic acid **11** and the hydrolyzed ChsE4-ChsE5 reaction product (Figure 3-5) highlights the appearance of a new proton resonance in the product spectrum alkene region at 6.85 ppm consistent with unsaturation at C24-C25 (Figure 3-5). As expected, only half of the reaction mixture was converted to product and the C24 alkene hydrogen resonance integrates to 0.5 hydrogen. Comparison of the H24 chemical shift to

chemical shifts calculated for (*E*) and (*Z*) alkene resonances of α,β-unsaturated carboxylic acids,[131] reveals that the (*E*) alkene is formed. Analogous NMR experiments with the 3-OCO-CoA reaction product revealed that the (22*E*)-3-oxo-chol-4,22-dien-24-oyl-CoA product was formed (Figure 3-6).

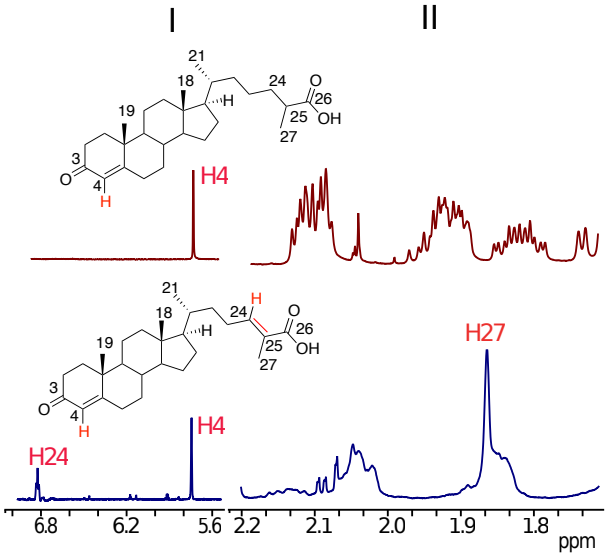


Figure 3-5. The product regio- and stereochemistry for ChsE4-ChsE5 catalyzed dehydrogenation of 3-OCS-CoA. ChsE4-ChsE5 forms (24*E*)-3-oxo-cholest-4,24-dien-26-oyl-CoA as determined by the ^{1}H NMR spectra (700 MHz) of substrate precursor 3-oxo-cholest-4-en-26-oic acid and ChE4-ChsE5 assay product after alkaline hydrolysis of its thioester, top and bottom, respectively. The spectra illustrate the changes in the alkene (I) and methyl (II) regions. The biochemical reaction catalyzed by ChsE4-ChsE5 is shown in Scheme 1.

In addition, a new methyl resonance that corresponds to the C27 methyl of the α,β-unsaturated acid appears as a singlet at 1.85 ppm (Figure 3-5). If C25-C27 dehydrogenation had occurred, two new alkene protons would have replaced the C27 methyl in the product spectrum, and there would be no methyl resonance at 1.85 ppm. We conclude that ChsE4-ChsE5 forms the thermodynamically favored (*E*)-tetrasubstituted alkene product (24*E*)-3-oxo-cholest-4,24-dien-26-oyl-CoA. This work was performed by Rui Lu.

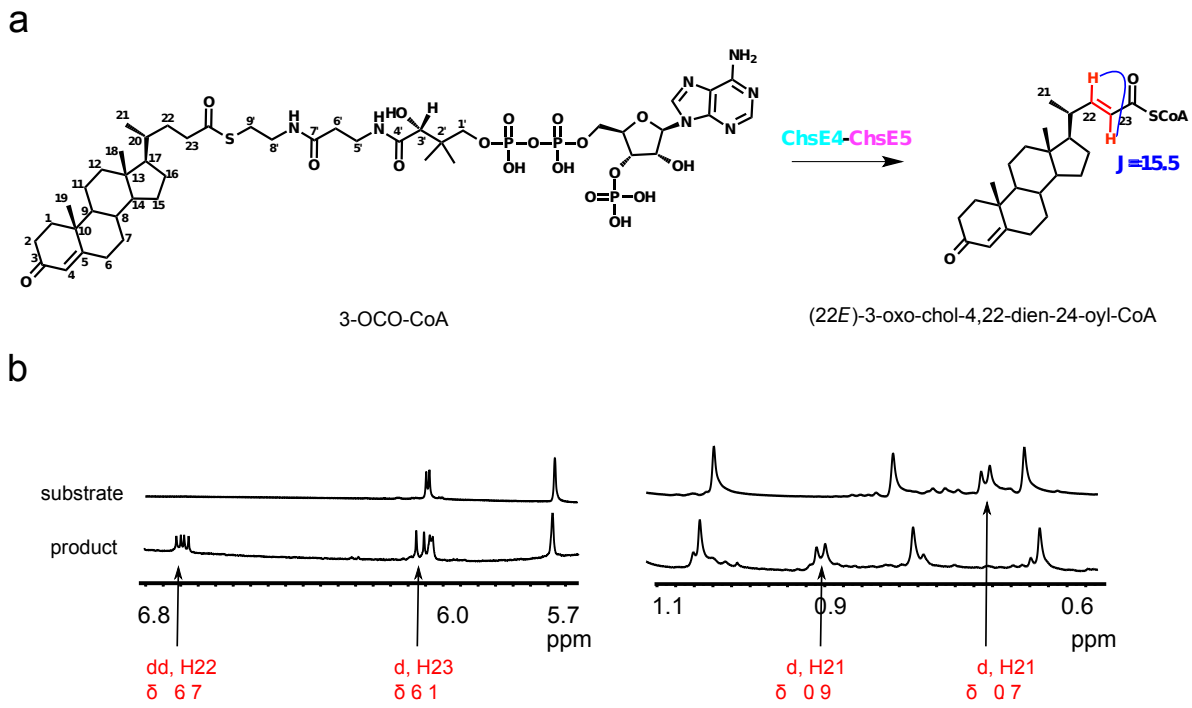


Figure 3-6. ChsE4-ChsE5 catalyzes the dehydrogenation of 3-OCO-CoA. (a) ChsE4-ChsE5 catalyzes the formation of (22*E*)-3-oxo-chol-4,22-dien-24-oyl-CoA, (b) as determined by the observed chemical shift and coupling constant. The C22 and C23 olefinic protons appear at 6.7 and 6.1 ppm as doublets in the product spectrum (panel b, left), and the double bond induces a downfield shift of the C20 proton and C21 methyl proton resonances (panel b, right).

First molecular structure of an $\alpha_2\beta_2$ heterotetrameric acyl-CoA dehydrogenase in any kingdom of life. Because there is no molecular structure of an $\alpha_2\beta_2$ heterotetrameric ACAD, we undertook structure determination of ChsE4-ChsE5 using X-ray crystallography. Crystals were obtained at pH 7.0 using the hanging drop vapor diffusion method. The crystal structure was determined using single-wavelength anomalous dispersion (SAD) phasing of selenomethionine (SeMet) substituted protein. The asymmetric unit contains one $\alpha_2\beta_2$ heterotetramer with two ChsE4 chains and two ChsE5 chains, and the initial model was refined to a resolution of 2.0 Å (Table 3-2).

Table 3-2. Data collection and refinement statistics

Crystal	ChsE4:ChsE5:$FADH_2$[a]
Space group	*C 1 2 1*
Cell dimensions	
a, *b*, *c* (Å)	185.7, 108.1, 82.0
α, β, γ (˚)	90.0, 93.0, 90.0
Data collection[b]	
Resolution (Å)	46.4−2.0 (2.00−1.99)
Wavelength (Å)	0.9792
Reflections	
Observed	739,658 (7,336)
Unique	109,767 (1,104)
R_{merge}[c]	0.076 (0.624)
R_{meas}[d]	0.082 (0.700)
R_{anom}[e]	0.057 (0.588)
R_{pim}[f]	0.031 (0.262)
$CC_{1/2}$[g]	0.999 (0.939)
$I / \sigma I$	19.9 (2.9)
Completeness (%)	99.9 (99.8)
Multiplicity	6.7 (6.6)
Wilson B	28.9
Refinement[b]	
Resolution (Å)	46.4−2.0 (2.05−1.99)
No. reflections	104,247 (7,657)
R_{work} / R_{free}	0.1508 / 0.1862
CC^*[g]	1.000 (0.984)
No. atoms	
Protein	11,396
$FADH_2$	106
Solvent	899
B-factors	
Protein	36.6
$FADH_2$	32.5
Solvent	42.5
R.M.S. deviations	
Bond lengths (Å)	0.017
Bond angles (˚)	0.84
Ramachandran	
Favored (%)	98
Allowed (%)	2
Outliers (%)	0
PDB ID	4X28

[a]See Figure 3-9.

[b]Values in parenthesis are for the highest resolution shell.

[c]The merging R-factor describes the deviation of an individual intensity measurement from the mean value of all its symmetry-equivalent reflections.

[d]The redundancy-independent merging R-factor R_{rim} or [c]CC_1R_{meas} indicates the precision of an individual intensity measurement independent of the multiplicity of that measurement.

[e]The precision-indicating merging R-factor, R_{pim} describes the precision of the averaged intensity measurement.

[f]The anomalous R-factor quantifies the relative differences of Friedel-related reflections.

[g]$CC_{1/2}$ and CC^* are statistics for assessing the effective resolution limits and quality of diffraction data in the context of a refined model.[132]

The complete tetramer is a dimer of heterodimers in a tetrahedral arrangement (Figure 3-7). Each heterodimer contains one single FAD binding site. The binding site is at the heterodimer interface and is composed of residues from both ChsE4 and ChsE5, as previously predicted.[56, 68] (Figure 3-7a, 3-7b) The molecular envelope of ChsE4-ChsE5 in solution, determined from small-angle X-ray scattering, agrees well with the shape of the tetrameric crystal structure with a χ value of 2.8 fitting from CRYSOL, thus confirming that the observed tetrameric architecture is the native solution state, and not a result of crystal packing (Figure 3-8).

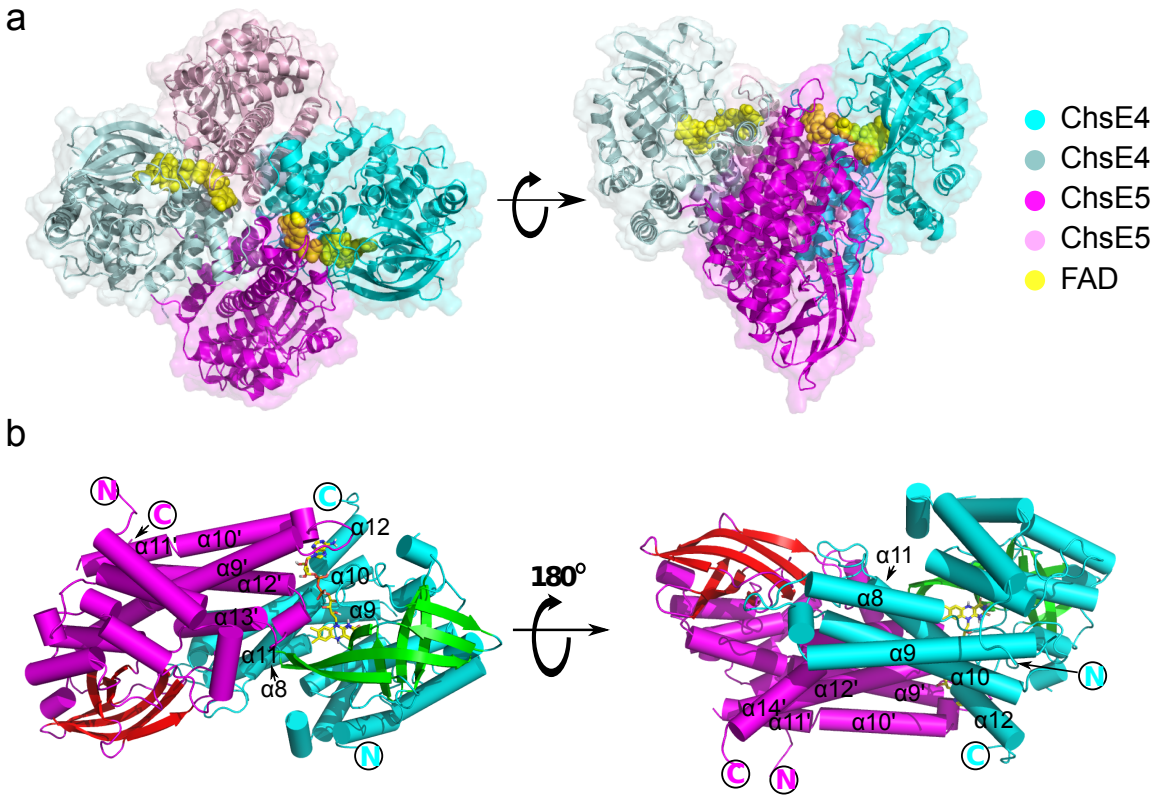


Figure 3-7. Overall atomic picture of ChsE4-ChsE5. (a) The biologically functional unit is an $\alpha_2\beta_2$ heterotetramer with two ChsE4 chains and two ChsE5 chains. The tetramer has two flavin adenine dinucleotide (FAD) binding sites and two FAD are bound. Surface representation of the structure corresponds to chain color. (b) A ChsE4-ChsE5 αβ heterodimer is shown as a cylindrical cartoon and colored by secondary structure. FAD binds to the interface of ChsE4 and ChsE5. The structure on the right is a 180° rotation around the X-axis of the structure on the left.

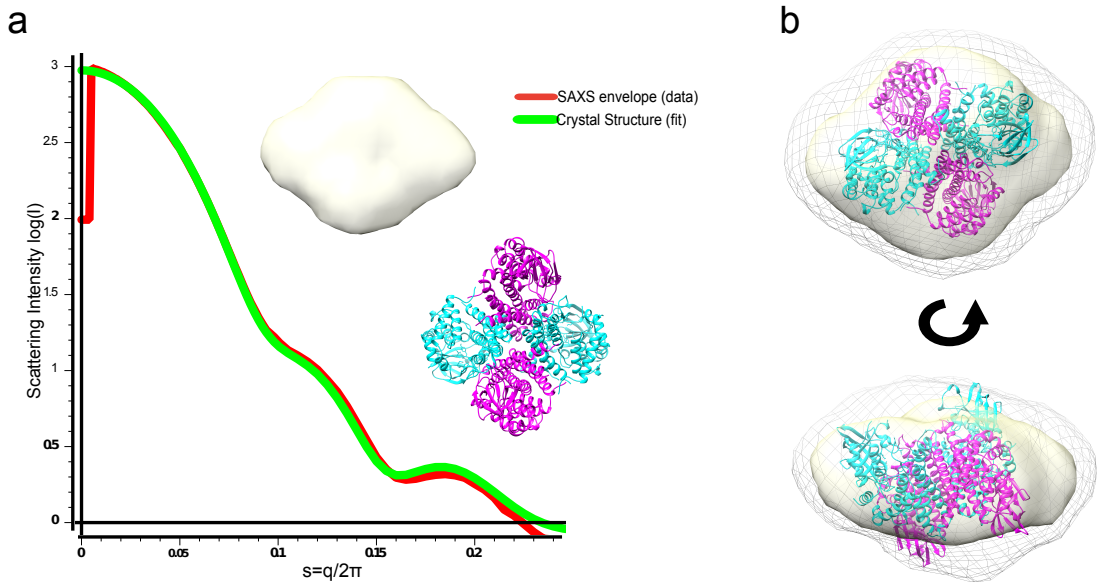


Figure 3-8. Small Angle X-ray Scattering (SAXS) generates a tetrameric envelope for solution state ChsE4-ChsE5. Small-angle X-ray scattering was used to determine the envelope of the solution state ChsE4-ChsE5 protein complex. (a) The theoretical scattering profile of the ChsE4-ChsE5 crystal structure, defined as scattering intensity (I), versus scattering vector (s) was calculated with CRYSOL and fit to the experimentally determined scattering profile for ChsE4-ChsE5 in solution, and the Chi-value (the square root of Chi squared) is 2.8. (b) The molecular envelope of solution state ChsE4-ChsE5 was determined from the experimental scattering profile using the ASTAS package. Then the high-resolution crystal structure was subsequently docked into the molecular envelope in Chimera, and the correlation coefficient is 0.90.

Without any substrate bound to the ChsE4-ChsE5 structure and therefore, no electron acceptor present, reduced FAD (a mixture of $FAD^{\bullet-}$ and $FADH_2$) was observed in the ChsE4-ChsE5 structure. A simulated annealing *Fo-Fc* omit electron density map of the bound FAD molecules in ChsE4-ChsE5 was calculated and showed an obvious bend of 154 – 156 ° in the electron density for the isoalloxazine moiety, instead of the conjugated flat electron density map that would be expected for the oxidized form (Figure 3-9). Using correlated optical spectroscopy

and X-ray diffraction data collection, we found that in the process of diffraction data collection, the FAD absorbance from 300 nm to 550 nm decreased as the X-ray exposure time increased, which determined that the FAD is reduced by X-ray irradiation (Figure 3-10a) and after the full data set collection, the FAD absorbance approached to almost fully reduced $FADH_2$. In addition, the portion of the crystal exposed to X-ray lost its original yellow color, whereas other portions of the crystal without exposure to X-ray retained yellow (Figure 3-10b), which further confirmed that X-ray exposure could reduce FAD. Therefore, we refined the ChsE4-ChsE5 structure with $FADH_2$ instead of FAD.

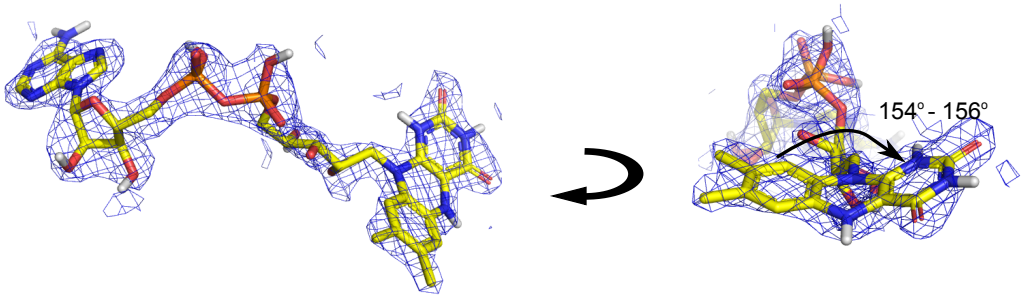


Figure 3-9. A simulated annealing *Fo-Fc* omit map for the bound FAD cofactor was calculated to reduce the effects of model bias. The electron density was contoured at 2.5 σ calculated in the absence of the ligand and displayed as blue mesh. The polar hydrogens are shown to clearly represent the oxidation state, but they are not present in the deposited coordinates.

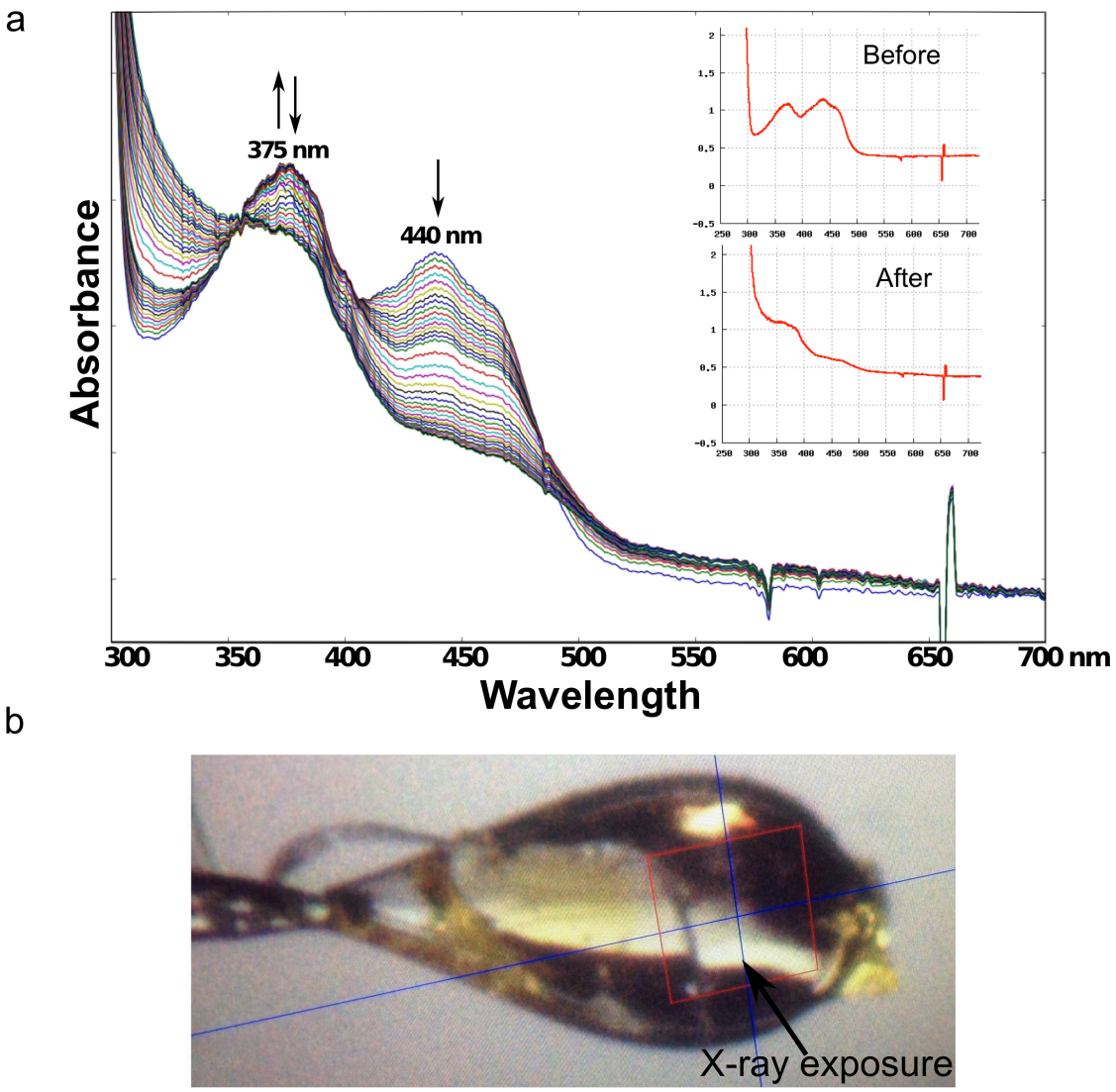


Figure 3-10. The FAD cofactor is reduced upon X-ray radiation. To investigate the potential reduction of FAD cofactor during our X-ray diffraction data collection, we collected a full dataset of X-ray diffraction with simultaneous optical spectroscopy as a function of X-ray exposure on beamline X26-C of the NSLS. The optical absorption spectra were collected during the readout of each X-ray diffraction image and an overlay of spectra during the diffraction data collection was constructed. (a) Correlated absorption spectroscopy data for the ChsE4-ChsE5 crystal were collected as a function of X-ray exposure. The spectra were overlayed and the changes at 440 nm and 375 nm are shown. Inset: Two spectra, one at the beginning and one at the end of X-ray exposure, were selected to highlight the change in absorbance. (b) The crystal after spectra coupled X-ray data collection.

The overall fold of both ChsE4 and ChsE5 is similar to that of typical ACADs, the quintessential example being medium chain acyl-CoA dehydrogenase (MCAD).[84] ChsE4 and ChsE5 both contain a typical N-terminal ACAD domain, typical C-terminal ACAD domain, and

a middle β-barrel domain, which are comparable in size (Figure 3-11). The N-terminal domains, consisting of mainly α-helices, lie on the surface of the tetramer (Figure 3-12b). The middle β-barrel domains of ChsE4 and ChsE5 are composed of two orthogonal β-sheets. The middle domain of ChsE4 is partially buried inside the tetramer and interacts with the FAD molecules. In contrast, the middle domains of ChsE5 are located completely on the surface (Figure 3-12a, 3-12b). The C-terminal domains are primarily composed of α-helices and form a four-helix up and down bundle. The C-terminal domains pack against each other to form the central core of the whole tetrahedral assembly.

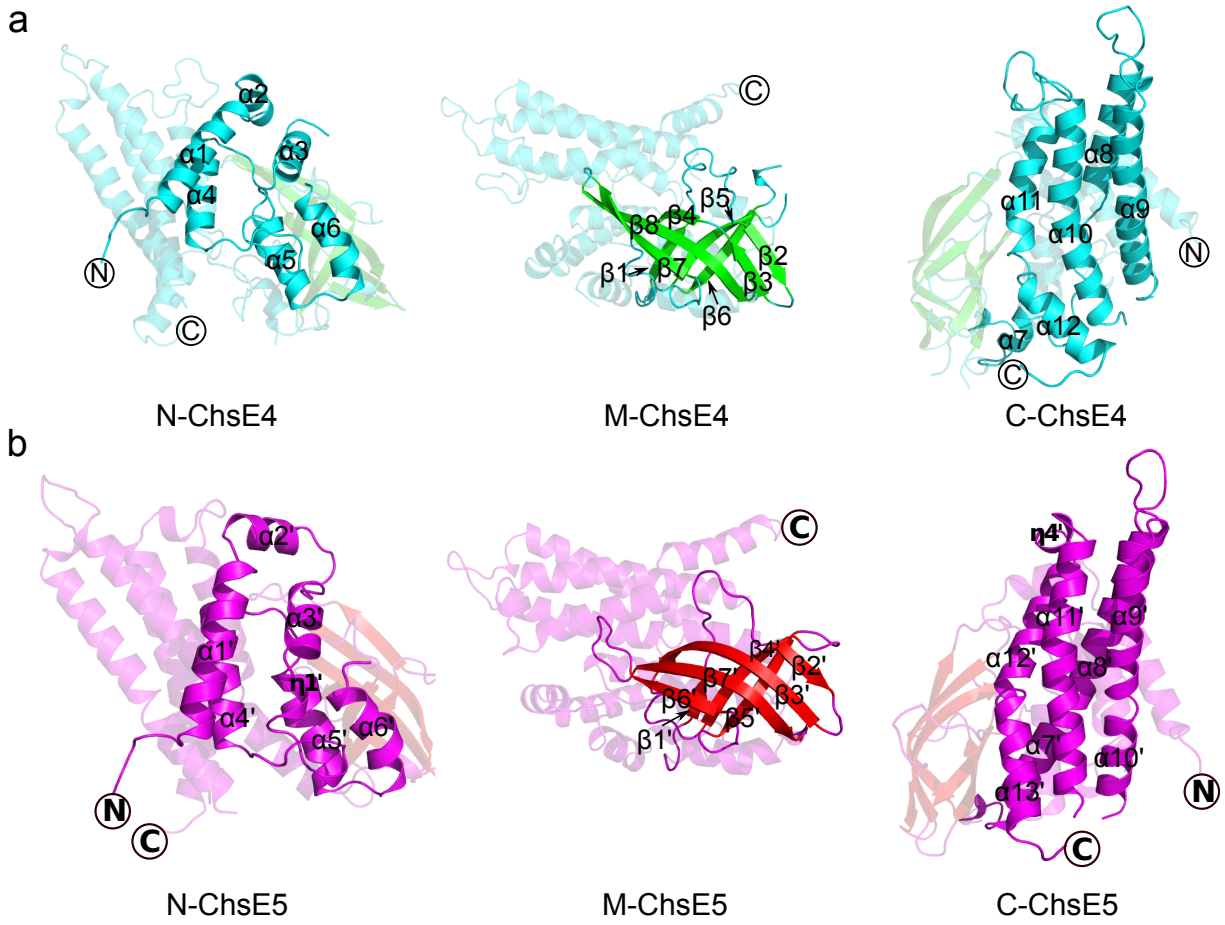


Figure 3-11. The three acyl-CoA dehydrogenase (ACAD) domains of ChsE4 and ChsE5. (a) ChsE4 and (b) ChsE5 both possess three ACAD domains: N-terminal domain (N-ChsE4; N-ChsE5), Middle domain (M-ChsE4; M-ChsE5), and C-terminal domain (C-ChsE4; C-ChsE5).

The ChsE4-ChsE5 dimer interface. Because of the divergent sequences and resultant asymmetry of ChsE4 and ChsE5, two different monomer-monomer interfaces are present in the heterotetrameric assembly. Interactions involving α10, α11, and α12 from ChsE4 with α11′, α12′, and α13′ from ChsE5 comprise a functional ChE4-ChsE5 heterodimer with one FAD cofactor bound (Figure 13a). Meanwhile, α8 and α9 from the same ChsE4 interact with α9′, α10′, α11′, and α13′ from the other ChsE5 chain in the tetramer to form a pseudo-heterodimer without a FAD cofactor bound (Figure 13b).

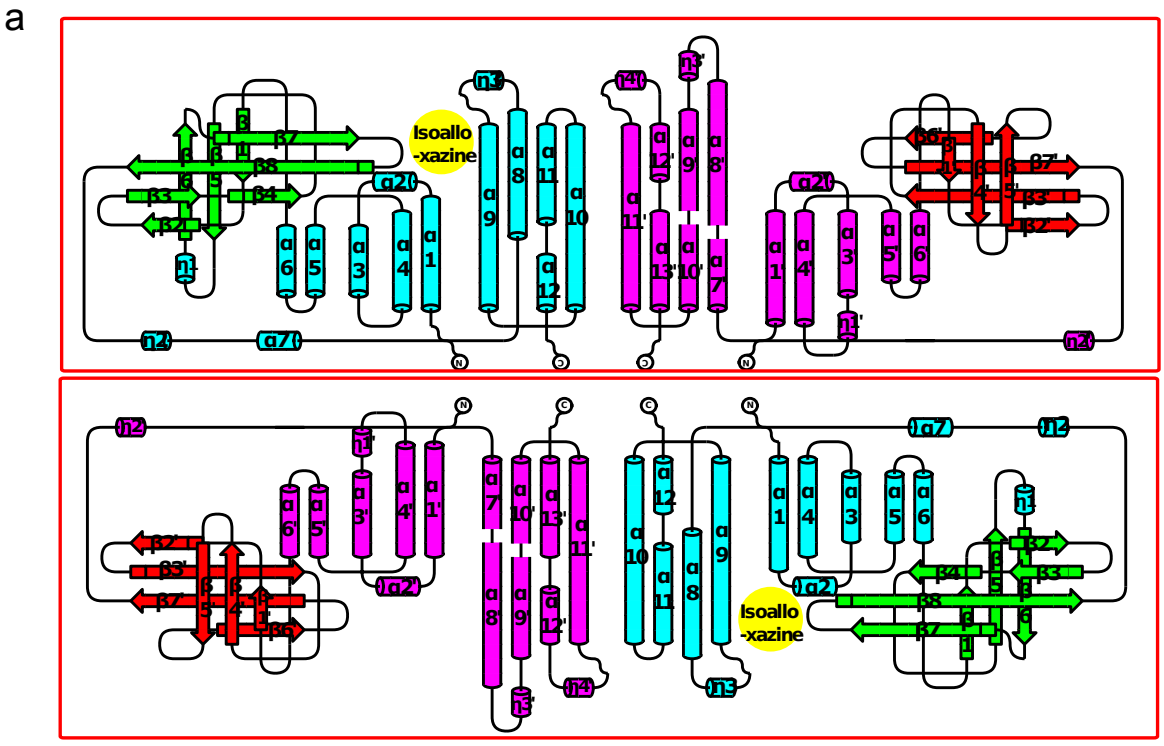


Figure 3-12. Topology of the ChsE4-ChsE5 $\alpha_2\beta_2$ heterotetramer and organization of ACAD domains relative to the tetramer and dimer interfaces. (a) Topology scheme for the ChsE4-ChsE5 heterotetramer. The functional heterodimers are circled in red boxes with FAD bound at the interface of ChsE4 and ChsE5; the isoalloxazine ring of FAD is highlighted in the yellow circle (b) N-terminal domains of both ChsE4 and ChsE5 are on the surface of the tetramer; C-terminal domains of both ChsE4 and ChsE5 are at the monomer-monomer interfaces and form the core of the tetramer. Middle β-barrel domains from ChsE5 are on the surface of the tetramer, whereas middle β-barrel domains from ChsE4 are partially exposed to the solvent and the other face contacts FAD.

ChsE4 and ChsE5 are structurally similar and complementary. As previously discussed, one ChsE4 chain and one ChsE5 chain are in intimate contact to form an FAD binding site (Figure 3-7b). The majority of the contacts are between the C-terminal domain helices (Figure 3-13a, 3-14a, and 3-14b). Specifically, the isoalloxazine ring lies in the cavity composed of the hydrophobic faces of the ChsE4 C-terminal domain and middle β-barrel domain (Figure 3-13a, 3-14b). The diphosphate and ribose sugar are bonded to residues from helices of the C-terminal domain of ChsE5. Just as in a canonical homotetramer,[84] half of the FAD binding residues are contributed by the ChsE4 protomer while the other half are from the ChsE5 protomer, together forming an intact FAD binding site (Figure 3-14a, and 3-14b).

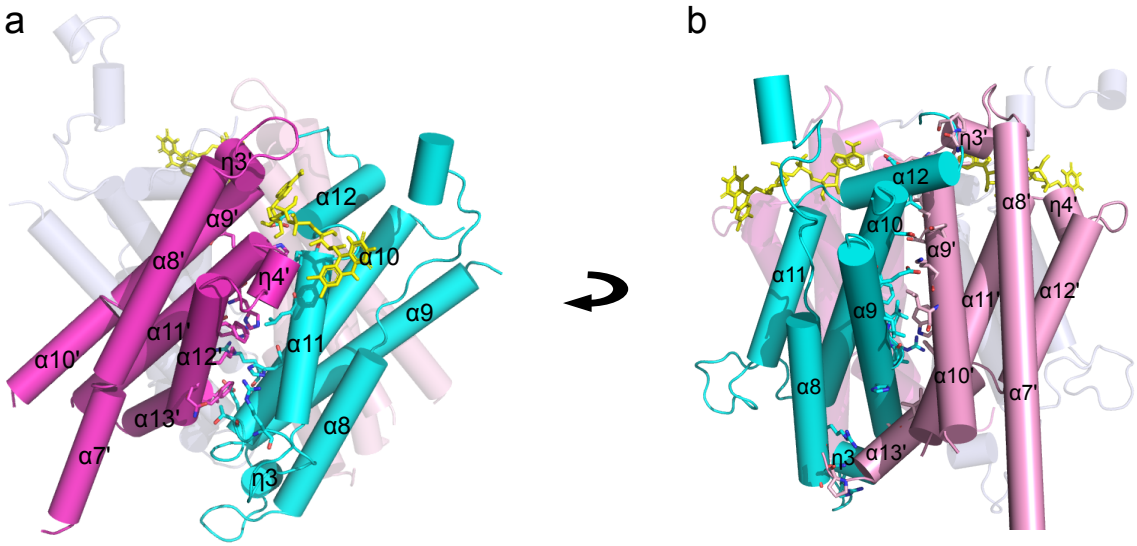


Figure 3-13. ChsE4 and ChsE5 interface. (a) The ChsE4 and ChsE5 interface forms the functional heterodimer with an FAD bound at the interface. (b) The ChsE4 and ChsE5 interface forms a pseudo heterodimer without an FAD bound. (b) is a 90° rotation around the Y-axis of (a).

In distinct contrast to the ChsE4 C-terminal domain, the ChsE5 C-terminal domain does not bind an isoalloxazine ring. The lack of bound FAD in the crystal structure is consistent with the stoichiometry of FAD:ChsE4-ChsE5 binding in solution. Superposition of the ChsE4 domain with the ChsE5 domain revealed that the analogous region of ChsE5 is unsuitable for binding an isoalloxazine moiety (Figure 3-14c, 3-14d). The side chains of ChsE5 residues Arg362, Met129, Asp126, and Gln151 fill the FAD site, and thus block FAD binding (Figure 3-14c), which is consistent with the surface representation that ChsE5 conflicts with the conserved FAD binding (Figure 3-14e). In ChsE4, the homologous residues Gln262, Thr136, Ser130 and Ser162 line a larger cavity and are positioned to form hydrogen bonds with FAD (Figure 3-14b).

Two, not four, substrate binding pockets are present in the ChsE4-ChsE5 tetramer. The acyl-CoA substrate-binding pocket is necessarily proximal to the bound FAD cofactor. In typical ACADs (for example, the well characterized MCAD), there are four acyl-CoA binding pockets adjacent to four FAD cofactors.[84, 133] Therefore, we analyzed whether the presence of only two FAD binding pockets in the ChsE4-ChsE5 tetramer was correlated with a commensurate reduction in the number of substrate binding pockets in the heterotetramer.

We superimposed the homodimer of MCAD (PDB: 3MDE) with the ChsE4-ChsE5 heterodimer and examined the position of FAD:octanoyl-CoA from MCAD in the ChsE4-ChsE5 heterodimer (Figure 3-15a). As expected, one MCAD FAD perfectly overlays with the FAD from ChsE4-ChsE5 (Figure 3-14b), and the other MCAD FAD exhibits steric clashes with residues from ChsE5 as seen in the ChsE4-ChsE5 superposition (Figure 3-14c). This superposition further confirms that ChsE5 cannot accommodate the isoalloxazine moiety.

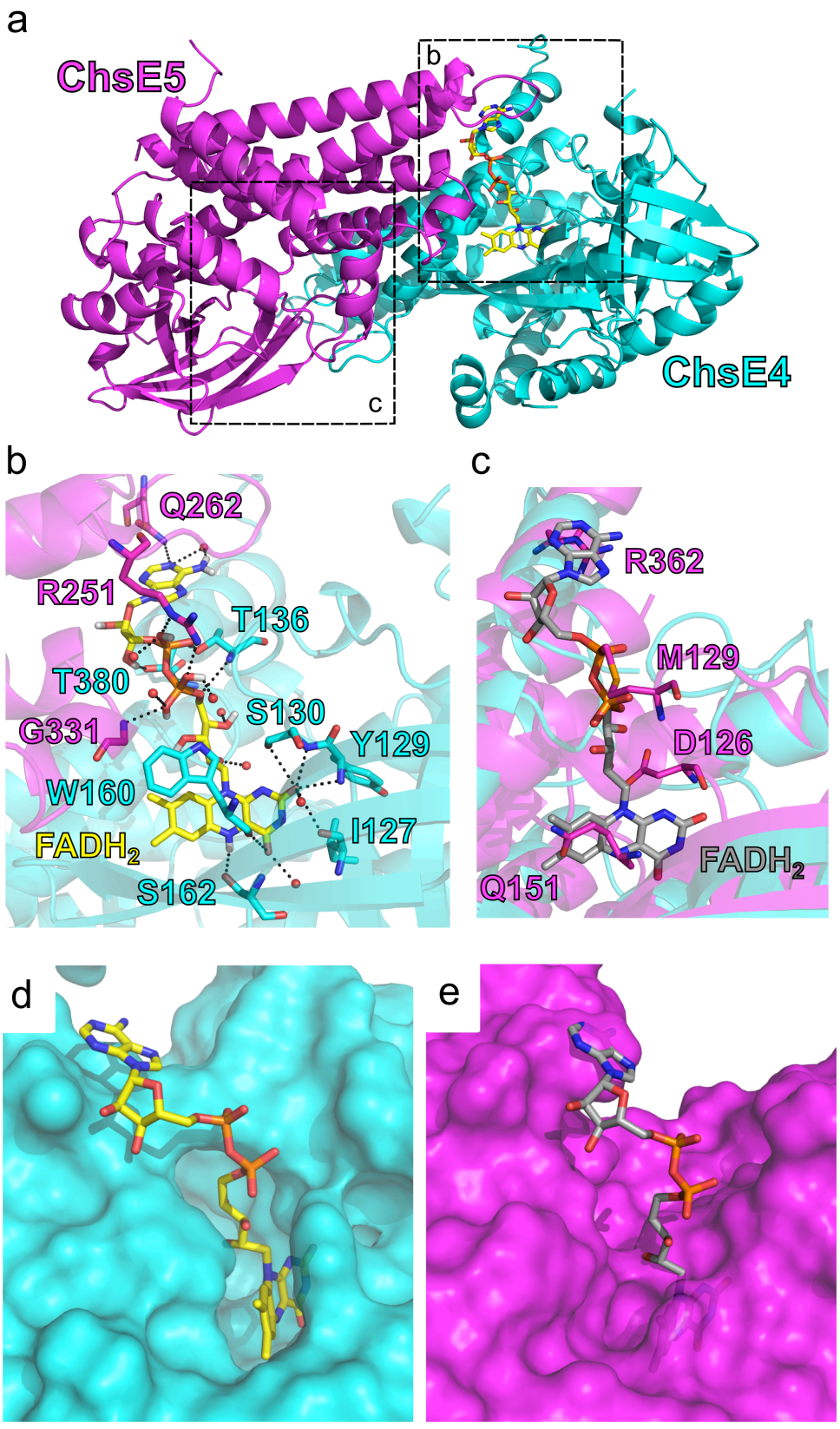

a
ChsE5
b
c
ChsE4
b
Q262
R251
T136
T380
S130
G331
Y129
W160
I127
FADH2
S162
c
R362
M129
D126
FADH2
Q151
d
e

Figure 3-14. The FAD binding sites. (a) There is only one FAD cofactor per ChsE4-ChsE5 dimer. The FAD cofactor binds at the interface of the ChsE4 and ChsE5 chains. The actual FAD binding site is in box b and the non-functional FAD binding region is in box c. (b) The detailed hydrogen bonding network that stabilizes bound FAD is shown. FAD is colored yellow. Water molecules are shown as red balls. Hydrogen bonding interactions are shown as black dashes. Residues that interact with FAD are labeled. The isoalloxazine ring is located inside ChsE4 and both ChsE4 and ChsE5 stabilize adenine. (c) The equivalent region in ChsE5 to the FAD binding site in ChsE4 is shown after superimposing ChsE5 onto ChsE4:FAD. The FAD molecule bound to ChsE4 is shown in gray and residues from ChsE5 that would clash with FAD are shown and labeled. (d) Surface representation of the actual FAD binding site in ChsE4. Protein chains are shown in surface and the FAD molecule is colored in yellow and shown in sticks. (e) Surface representation of the non-functional FAD binding region in ChsE5. The FAD molecule is colored in gray.

In all the holo-ACAD structures with acyl-CoA ligand bound, the coenzyme A moiety is exposed to solvent. The bound conformation is stabilized by hydrogen bonding interactions with the protein and the aliphatic chains extend deep into a hydrophobic protein pocket from which solvent is excluded. Octanoyl-CoA from the MCAD structure superimposes directly into the ChsE4 hydrophobic pocket without any steric clashes. The CoA moiety is stabilized by hydrogen bonds with ChsE4 residues Thr136, Arg187, and Gln243. These residues are highly conserved among ACAD CoA binding sites (Figure 3-15b).[56]

The second MCAD octanoyl-CoA does not fit into ChsE5. Specifically, residues Leu88, Asp126, and Arg229 project into the proposed binding tunnel and sterically clash with the docked aliphatic chain of octanoyl-CoA (Figure 3-15b). Acidic and basic side chains further disfavor binding hydrophobic substrates. Moreover, the highly conserved residues that would ordinarily stabilize CoA binding are absent in ChsE5. Thus, the ChsE5 sequence and therefore structure, diverges from the typical ACAD sequence and structure such that it does not possess isoalloxazine or acyl-CoA binding pockets, and does not have a catalytic site or an obvious metabolite binding site.

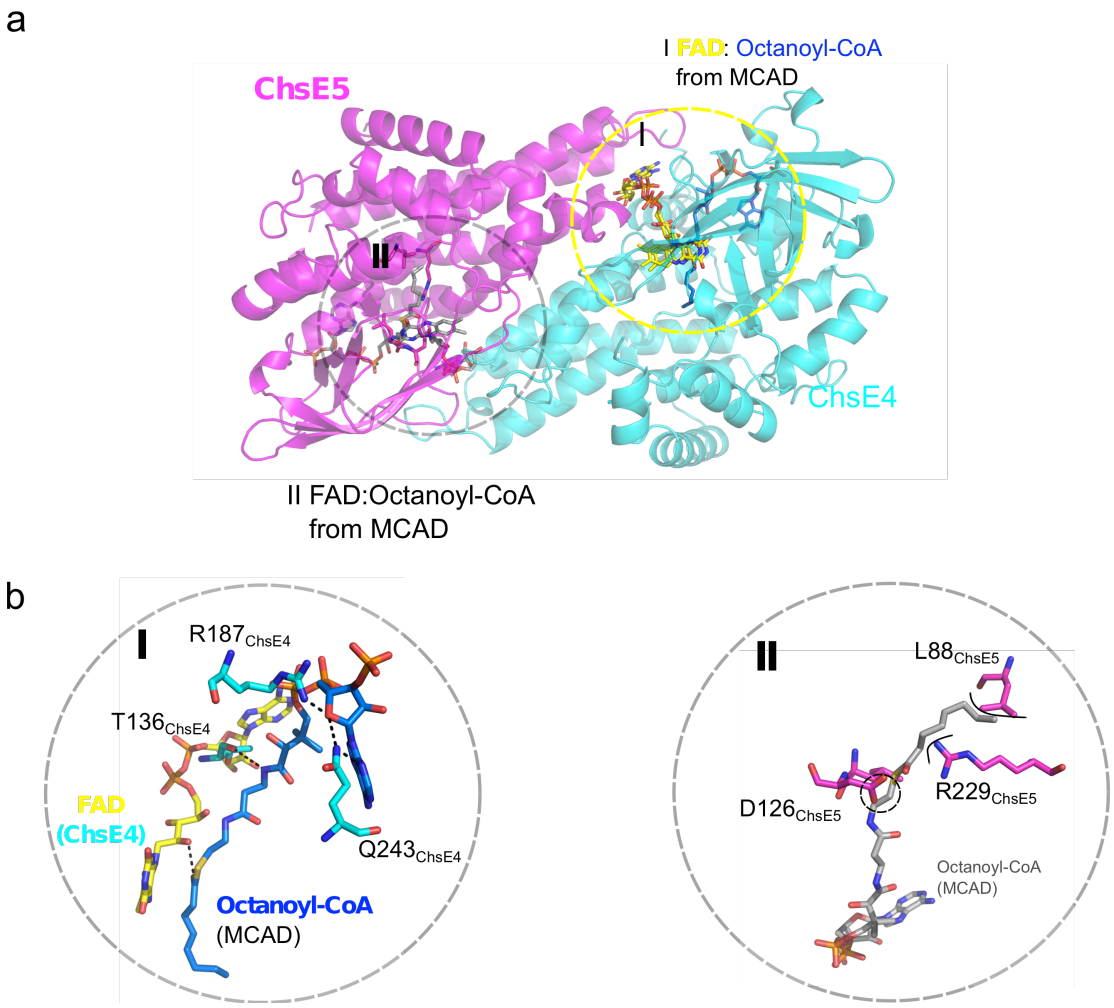


Figure 3-15. The acyl-CoA binding sites. (a) The ChsE4-ChsE5 heterodimer was superimposed onto the MCAD homodimer (PDB: 3mde) with two FAD:acyl-CoA binding sites. The ChsE4-ChsE5 heterodimer is shown with the two sets of FAD:acyl-CoAs from MCAD. One of the FAD cofactors overlays completely with the FAD from ChsE4, circled in I and colored in yellow and the octanoyl-CoA from MCAD is shown in blue. The other FAD: octanoyl-CoA binding site from MCAD is circled in II and shown in gray. (b) The highly conserved residues, T136, R187 and Q243 together with the FAD from ChsE4 interact with the CoA moiety in binding site **I**. The residues L88, D126, and R229 from ChsE5 would clash with the octanoyl-CoA, shown in circle **II**. In addition, the highly conserved CoA interaction residues T, R, and Q are not conserved in binding site II.

We then examined the superposition of ChsE4 and ChsE5 to further characterize the differences in binding sites in the two protomers. In addition to the presence of large polar residues that occlude the region of ChsE5 that typically would be an acyl-CoA binding tunnel, the superimposed structures of ChsE4 and ChsE5 reveal a conformational shift relative to ChsE4.

In ChsE5, the secondary structures α11′, α12′, α13′, β6′, β7′ and β3′ are all shifted relative to ChsE4, and instead of comprising a very flexible loop as in ChsE4, α7′ in ChsE5 occupies a relatively large space in what would have been the binding pocket. These shifts further reduce the ChsE5 binding cleft (Figure 3-16a).

The substrate-binding pocket of ChsE4-ChsE5 can accommodate bulky steroid substrates. Next, we docked the substrate (*25S*)-3-OCS-CoA into the ChsE4-ChsE5 binding pockets. Steric clashes precluded docking into the ChsE5 pocket, again suggesting that ChsE5 does not possess the ability to bind steroid substrates due to the reduced size of the cleft, which is filled with hydrophilic residues (Figure 3-16b). Conversely, the 3-OCS-CoA substrate could be readily docked into the ChsE4 site, demonstrating that the ChsE4 binding site is sufficiently large to accommodate a steroid structure (Figure 3-17).

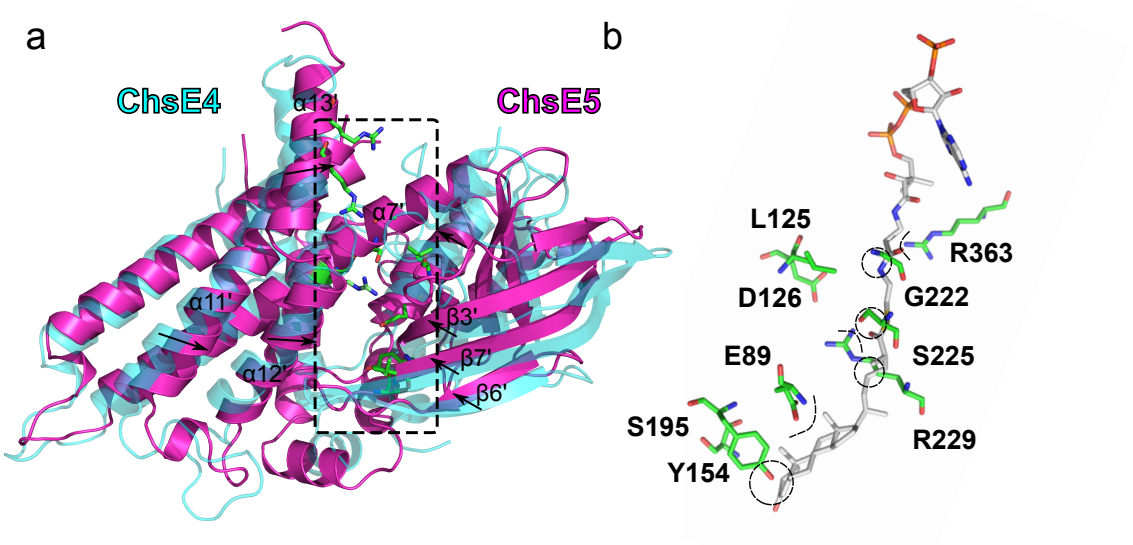


Figure 3-16. Substrate binding site analysis in ChsE5. ChsE5 was superimposed onto ChsE4 and the RMS value is 2.958 Å with 887 α-carbons aligned. The ChsE5 secondary structure that corresponds to the binding tunnel in ChsE4 is shifted into the tunnel relative to ChsE4; the shifting directions are shown in black arrows. The residues that surround the binding tunnel in ChsE5 are shown as sticks and are colored by atom type.

The 3-OCS-CoA steroid side chain aligns between the FAD cofactor and the putative active site base, Glu247. In homotetrameric ACAD structures, the active site base is located on an α-helix in close proximity to the proton to be abstracted. In contrast, Glu247 resides on a random coil that interrupts this α-helix (Figure 3-17a, 3-17b). Glu247 is too far from the docked structure for proton transfer to occur suggesting that a conformational change occurs upon substrate binding. Two hydrogen-bonding interactions are formed between the phosphate group of the CoA molecule and His175 and Arg381. NH3″ interacts with Thr130 and the thioester carbonyl group is stabilized by the backbone amide of Gly362 through hydrogen bonding. Except for a few polar interactions, 3-OCS-CoA is mainly surrounded by hydrophobic and aromatic residues (Figure 3-18).

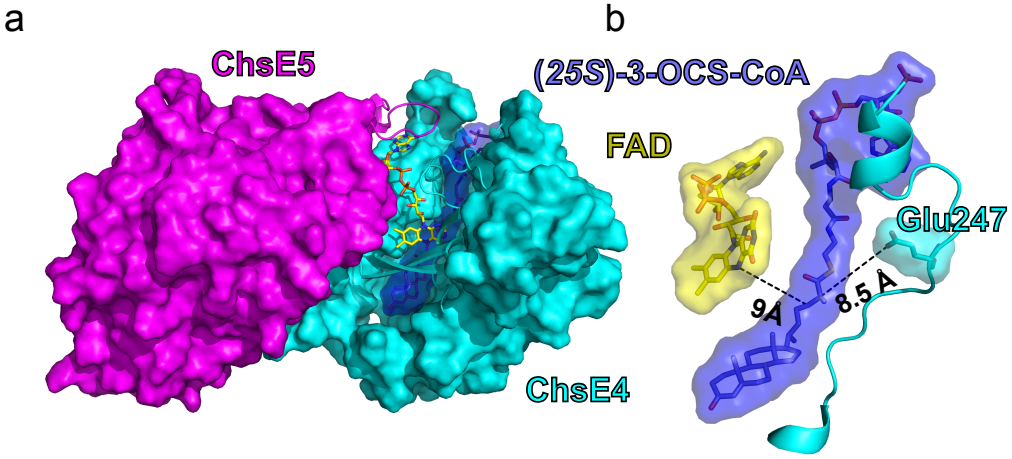


Figure 3-17. (*25S*)-OCS-CoA docked into ChsE4 and sandwiched between FAD and the active site base. (a) For clarity, half of the FAD binding tunnel is represented as a surface colored by chain, while the other half of the tunnel is drawn as a cartoon. (b) Orientation and proximity of FAD (yellow), the docked 3-OCS-CoA (blue) and the active site base (light blue). The distances between Glu247 and C25 in 3-OCS-CoA, and between C24 in 3-OCS-CoA and N5 in FAD are indicated as dashed lines and labeled.

Glu247 from ChsE4 is the active site base required for dehydrogenation. Consistent with the substrate docking results described above, alignment of the primary amino acid sequences of

ChsE4 and ChsE5 against nine well-characterized human ACADs revealed that ChsE4 possesses the typical catalytic general base glutamate in the conserved position, but ChsE5 does not.[68] In order to experimentally confirm that ChsE4-Glu247 is the active site base, we mutated ChsE4-Glu247 to Ala.

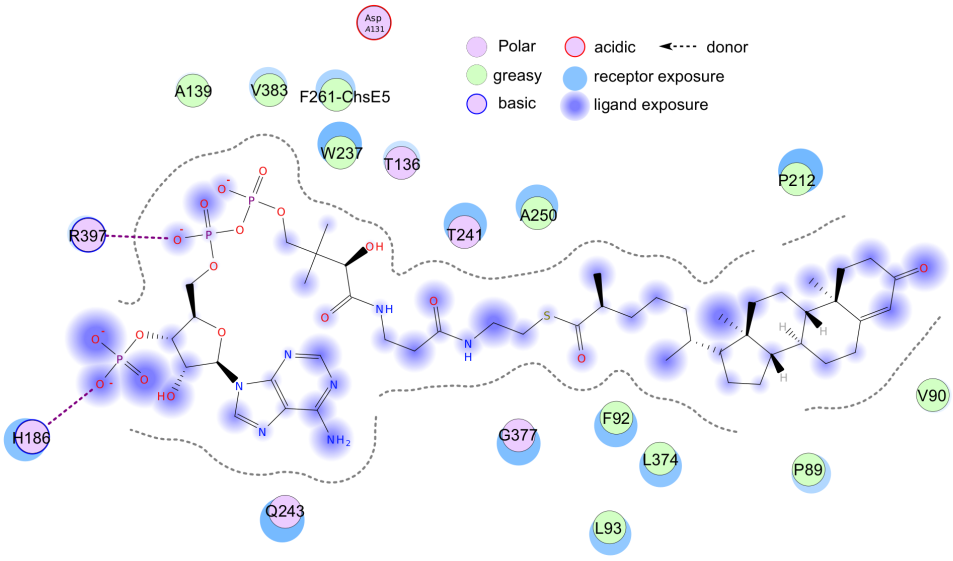


Figure 3-18. Scheme of the interactions between the ChsE4-ChsE5 heterodimer and (25S)-3-OCS-COA.

The purified mutant ChsE4$_{E247A}$-ChsE5 pulled down as a heteromeric complex, and had a characteristic FAD absorbance in the UV-visible spectrum (Figure 3-19). ChsE4$_{E247A}$-ChsE5 was characterized by analytical gel filtration and found to have the same elution profile as wild-type ChsE4-ChsE5, indicating that ChsE4$_{E247A}$-ChsE5 still forms a stable $\alpha_2\beta_2$ heterotetramer in solution. The activity of ChsE4$_{E247A}$-ChsE5 was assayed under the same conditions as ChsE4-ChsE5. However, no detectable activity was observed even at very high concentrations of enzyme (5 μM) and 3-OCS-CoA substrate (100 μM). Furthermore, no product formation was observed by MALDI-TOF spectrometry (Figure 3-4). These results indicate that Glu247 in ChsE4 is the active site general base required to afford a dehydrogenated product.

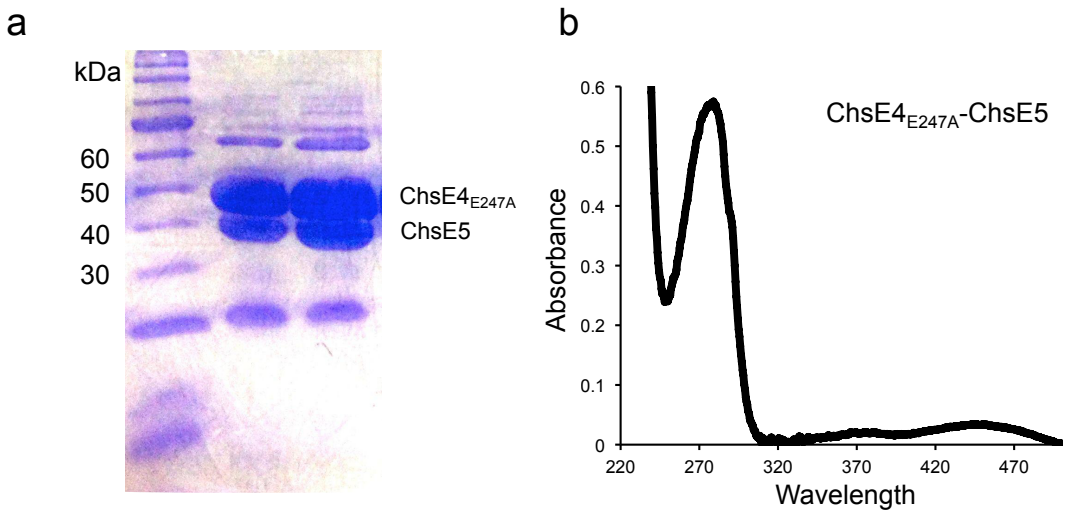


Figure 3-19. Purification and characterization of $ChsE4_{E247A}$-ChsE5. (a) SDS-PAGE gel analysis of the purified $ChsE4_{E247A}$-ChsE5 (b) The UV-visible spectrum of the purified $ChsE4_{E247A}$-ChsE5 showing the characteristic FAD absorbance.

Substrate binding tunnel comparison across ACADs. The ability of ChsE4-ChsE5 to bind and catalyze dehydrogenation of steroid acyl-CoA substrates led us to analyze further the structural details of its substrate-binding tunnel. The substrate binding sites of human isovaleryl-CoA dehydrogenase (i3VD), *Mycobacterium smegmatis* FadE13, pig medium chain acyl-CoA dehydrogenase (MCAD), rat short chain acyl-CoA dehydrogenase (SCAD), and *Megasphaera elsdenii* butyryl-CoA dehydrogenase (BCAD) were compared using Caver.[134] All of the binding sites are located at the enzyme core (Figure 3-20a). Of all the binding sites, ChsE4-ChsE5 has the longest tunnel-like binding site (Figure 3-20a).

A structure-based sequence alignment of the homologs revealed differences that reflect substrate specificity. Even though all of the ACADs have homologous structures, the distinctive loops between α4 and α5, α7, and α8 give ChsE4 the ability to bind bulkier substrates by proving both more flexibility and a larger amount of space (Figure 3-20b), supporting its unique biological function to degrade bulky cholesterol in *Mtb*.

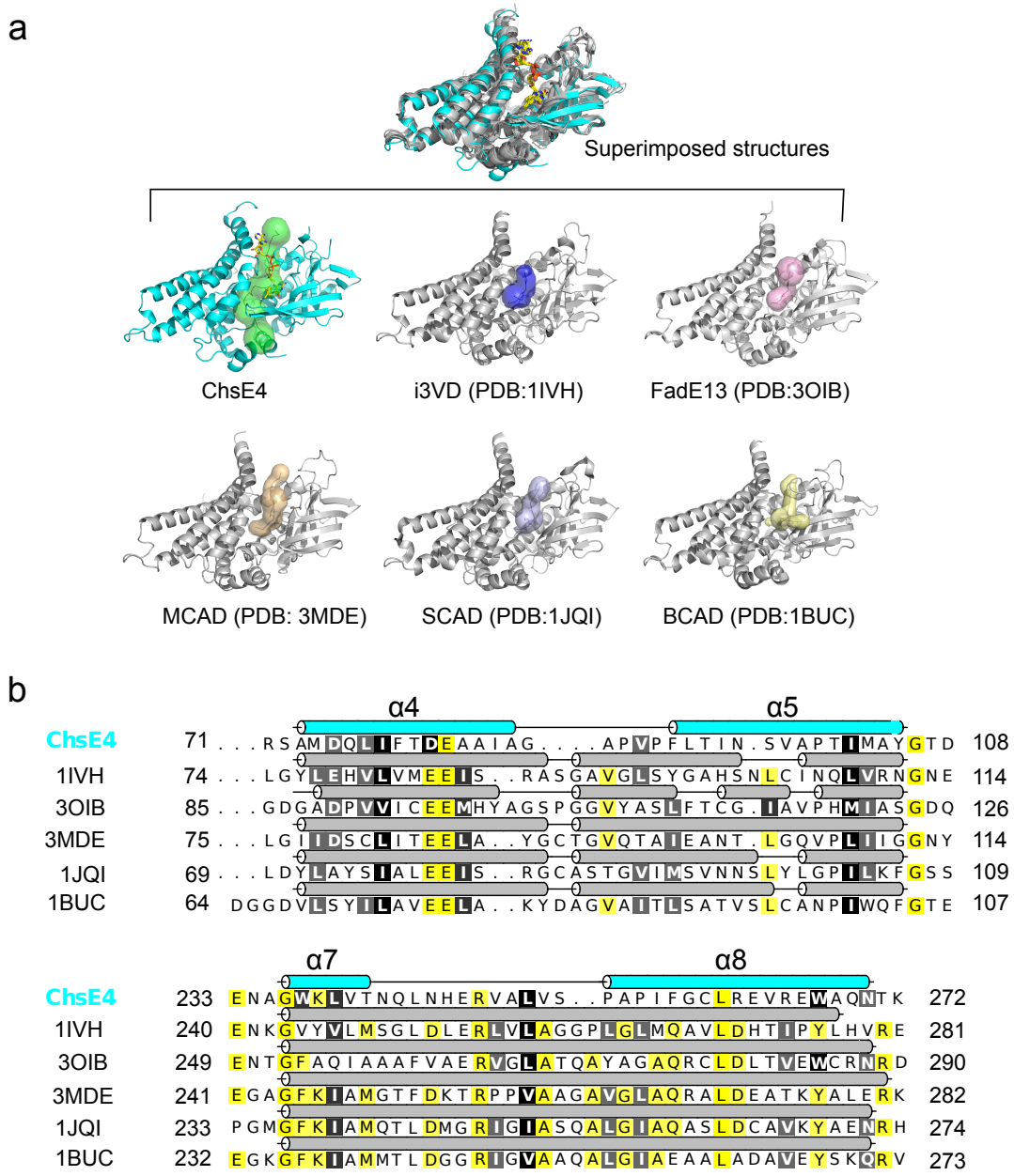


Figure 3-20. Substrate binding site comparison across species. (a) The structure of ChsE4, human isovaleryl-CoA dehydrogenase (i3VD), pig medium chain acyl-CoA dehydrogenase (MCAD), Mycobacterium smegmatis FadE13, rat short chain acyl-CoA dehydrogenase (SCAD) and Megasphaera elsdenii butyryl-CoA dehydrogenase (BCAD) are superimposed and their substrate binding sites have been identified by Caver. The substrate binding sites are shown as transparent surfaces. The RMS value between ChsE4 and i3VD is 2.030 Å with 975 α-carbons aligned; the RMS value between ChsE4 and 3OIB is 1.718 Å with 831 α-carbons aligned; the RMS value between ChsE4 and 3MDE is 2.362 Å with 942 α-carbons aligned; the RMS value between ChsE4 and 1JQI is 2.714 Å with 941 α-carbons aligned; the RMS value between ChsE4 and 1BUC is 2.086 Å with 870 α–carbons aligned. (b) Secondary structure sequence alignment

of ChsE4, i3VD, MCAD, SCAD, BCAD and FadE13. ChsE4 is colored in cyan, while the other secondary structure cartoons are colored in gray. Yellow highlighted residues are identical, black and gray highlighted residues are highly similar.

Substrate promiscuity of ChsE4-ChsE5 explains the *igr* phenotype. ChsE4-ChsE5 is the only ACAD in the KstR1 regulon that can dehydrogenate 3-OCS-CoA. Thus, ChsE4-ChsE5 controls metabolic flux into cholesterol side chain β-oxidation. However, ChsE4-ChsE5 also catalyzes the dehydrogenation of 3-OPC-CoA and 3-OCO-CoA, albeit with somewhat slower rates. The substrate specificity overlap of ChsE4-ChsE5 with ChsE1-ChsE2 and ChsE3 (Table 3-3 and Figure 3-2) requires that compensating enzyme activities must be considered in the evaluation of ACAD gene knockout phenotypes *in vivo* and *in vitro.*

The *in vivo* and *in vitro* phenotypes of the *igr* knockout have been evaluated.[48-50] Notably, initial growth of H37Rv:Δ*igr* in the mouse model of infection is attenuated.[48] The attenuation has been attributed to the inability of the Δ*igr* mutant to metabolize cholesterol fully.[49] A 3-OPC-CoA related metabolite accumulates in Δ*igr* mutant cultures grown in the presence of cholesterol for two weeks, consistent with loss of *igr*-encoded ChsE1-ChsE2 activity.[50] However, 24 weeks after infection *in vivo*, Δ*igr* mutant bacterial counts in the lung reach wild-type levels.[48]

Taking these results together, we infer that at early stages of infection, the catalytic activity of ChsE4-ChsE5 with 3-OPC-CoA is insufficient to overcome loss of ChsE1-ChsE2, but that at later stages ChsE4-ChsE5 is able to compensate the deficiency. This change may be attributed to a shift in metabolite concentrations. Although we do not know the concentration of the steroid CoA metabolites *in vivo*, typically intracellular substrate concentrations of metabolites are close to or higher than K_m of the respective catalyzing enzyme.[135] We propose that during the initial stages of infection, 3-OCS-CoA levels are sufficiently high that ChsE1-ChsE2 is required for metabolism of 3-OPC-CoA. As infection reaches the chronic phase, the supply of 3-OCS-CoA

becomes depleted, 3-OPC-CoA accumulates, and as the substrate specificities suggest (Table 3, Figure 2), ChsE4-ChsE5 can compensate for loss of ChsE1-ChsE2. We hypothesize that similarly, there are enoyl-CoA hydratases and retroaldolases that compensate loss of the *igr* operon once dehydrogenation of 3-OPC-CoA has been accomplished.

Conclusion. The substrate specificities of the KstR1-regulated ACADs for sterol side-chain degradation in *Mtb* have been elucidated. Only ChsE4-ChsE5 catalyzes the dehydrogenation of the 8-carbon side chain steroid intermediate, 3-OCS-CoA, which is required for the first cycle of steroid side chain β-oxidation. ChsE3 only catalyzes the dehydrogenation of the five-carbon steroid side chain intermediate, 3-OCO-CoA, and as previously demonstrated[50, 56] ChsE1-ChsE2 specifically catalyzes the dehydrogenation of the three-carbon steroid chain intermediate, 3-OPC-CoA.

ChsE4-ChsE5 has the broadest substrate specificity. While the steady-state kinetics of ChsE4-ChsE5 show a preference for dehydrogenation of an 8-carbon side chain substrate, both the 5-carbon and 3-carbon substrates are dehydrogenated at rates comparable to catalysis of their dehydrogenation by ChsE1-ChsE2 and ChsE3. These secondary activities allow kinetic compensation for loss of the second (ChsE3) or third (ChsE1-ChsE2) ACAD in steroid side chain β-oxidation. ChsE4-ChsE5 preferentially dehydrogenates steroid substrates, as demonstrated by the 80-fold lower apparent second order rate constant for octanoyl-CoA.

The structure of ChsE4-ChsE5 serves as a starting point for the development of specific inhibitors for this subfamily of structurally distinct ACADs. The ChsE4-ChsE5 substrate-binding site is much larger than those seen in host ACADs. Although the $\alpha_2\beta_2$ heterotetramer structure is highly conserved in Mycobacterial and Proteobacterial species, it is distantly related to typical

ACADs and no known human orthologs exist.[41, 56, 68] Appropriately designed inhibitors that prevent *Mtb* from generating downstream cholesterol catabolites acetyl-CoA and propionyl-CoA by blocking the first step of cholesterol side chain β-oxidation will allow direct investigation into the role this pathway plays in *Mtb* persistence and survival *in vivo*.

4. Acknowledgementss

The X-ray data collection and structural determination of ChsE4-ChsE5 were performed in collaboration with Dr. Kip Guja; The synthesis of 3-OCS-CoA and 3-OCO-CoA were accomplished by Rui Lu; The regio- and stereochemical characterization of dehydrogenated products of ChsE4-ChsE5's catalysis were conducted by Rui Lu. Biophysical characterization of ChsE3 was accomplished by Johnna St. Clair and Matthew Wipperman. Amber Bonds helped finish the steady-state kinetics of ChsE4-ChsE5 with 3-OCO-CoA as the substrate.

Chapter IV. Elucidating *igr* genes from *Mtb*

This chapter is adapted from a paper that was published in ACS Chemical Biology[111]

1. Introduction

In the process of defining genes required for survival of *Mtb* in macrophages by transposon library screening, several experiments highlighted the importance of the intracellular growth operon (*igr* operon), comprised of *Rv3540c* to *Rv3545c* in *Mtb*, for survival in macrophages.[42, 48] Deletion of the operon (H37Rv:Δ*igr*) results in attenuated growth in macrophages and in mice.[48] The genes of the *igr* operon were predicted to encode lipid β-oxidation enzymes.[136] However, the operon is not required for growth on even or odd chain fatty acids.[48] The location of the *igr* operon in the 82-gene cholesterol catabolism cluster (*Rv3492c* to *Rv3574*) suggested a potential function in cholesterol metabolism,[39] and the operon was found to be required for growth of *Mtb in vitro* in the presence of cholesterol.[49] Metabolite profiling of the H37Rv:Δ*igr* strain using isotopically-labeled cholesterol resulted in accumulation of a metabolite. The structure of the metabolite led to the hypothesis that the *igr* operon is required for β-oxidation of C20-C22 in the cholesterol side chain.[50]

The 6 genes (*Rv3540c-Rv3545c*) in the *igr* operon are annotated as a lipid transfer protein (Ltp2/Rv3540c), MaoC-like hydratases (Rv3541c, Rv3542c), acyl-CoA dehydrogenases (Rv3543c, Rv3544c) and a cytochrome P450 (cyp125/Rv3545c). Enzymatic and structural studies demonstrated that Cyp125 oxidizes cholest-4-en-3-one or cholesterol at C-26 to cholest-4-en-3-one-26-oic acid.[87] However, in the H37Rv strain employed in the *igr* studies, loss of this activity is compensated by Cyp142.[109] Biophysical characterization of the acyl-CoA dehydrogenases Rv3543c-Rv3544c, named ChsE1-ChsE2, showed that they form an obligate $\alpha_2\beta_2$ heterotetramer with two active sites and two FAD cofactor binding sites.[50, 56] ChsE1-ChsE2 catalyzes the C17-C20 dehydrogenation of steroid acyl-CoA thioester, 3-oxo-4-pregnene-20-

carboxyl-CoA.[50] All the above information established the importance of the *igr* operon for cholesterol metabolism by *Mtb*.

A BLAST search based on protein sequence similarities indicated that both Rv3541c and Rv3542c have MaoC-like hydratase domains, and in addition Rv3542c includes a DUF35/DUF35_N domain (Figure 4-1a). Here, we demonstrate that Rv3541c and Rv3542c form a heterodimer, two of which further dimerize to form an $\alpha_2\beta_2$ heterotetramer. The unique architecture of this heterotetramer is in stark contrast to previously characterized enoyl-CoA hydratase structural homologs, e.g., peroxisomal and bacterial enzymes, that form homodimers. We also demonstrate that Rv3541c-Rv3542c catalyzes the hydration of a steroid substrate, enoyl-CoA thioester 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA). We now refer to Rv3541c-Rv3542c as ChsH1-ChsH2 to emphasize its function in cholesterol side chain enoyl-CoA hydration.

Interestingly, the substrate binds in an unusual pocket that spans both ChsH1 and ChsH2, explaining why heterodimer formation is necessary, and the pocket is dynamic in size. Taken together with the distinctive acyl-CoA dehydrogenase (ACAD) quaternary structure of ChsE1-ChsE2,[56] our data suggests that *Mtb* metabolizes sterol side chains through a unique assembly of β-oxidation enzymes. The structures described here highlight opportunities for the design of inhibitors specifically targeting cholesterol catabolism in actinobacteria, particularly *Mycobacterium tuberculosis*.

Ltp2 is annotated as a lipid transfer protein or keto acyl-CoA thiolase from Tuberculist genolist. Regulated by the KstR1 repressor and located in the essential *igr* operon region, Ltp2's importance in cholesterol metabolism is self-evident. However, no reports are available up to

now covering its physiological function or 3D structure. A BLAST search based on protein sequence similarities indicated that Ltp2 belongs to sterol carrier protein X (SCPx) with specificity for branched acyl-CoAs (Figure 4-1b). Here, we establish that Ltp2 is more similar to a sterol carrier protein X (SCPx) thiolase and has to partner with the DUF35 domain from Rv3542c to catalyze the thiolase cleavage of 17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA (17-HOPC-CoA), generating androstenedione (AD) and propionyl-CoA.

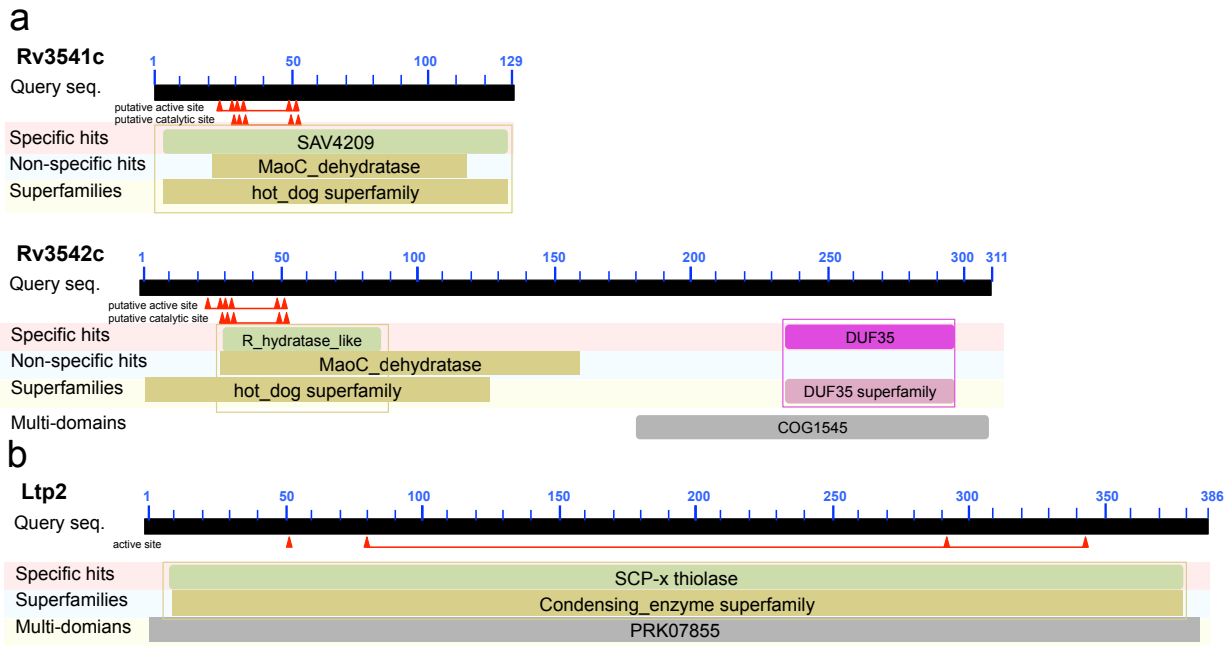


Figure 4-1. Annotations of Rv3541c, Rv3542c and Ltp2 from BLAST search based on their protein sequences. (a) SAV4209 is a *Streptomyces avermitilis* protein with a hot-dog fold that is similar to those of (*R*)-specific enoyl-CoA hydratases. MaoC like domain: the maoC gene is part of a operon with maoA which is involved in the synthesis of monoamine oxidase. Several bacterial proteins that are composed solely of this domain have (*R*)-specific enoyl-CoA hydratase activity. DUF35 OB-fold domain: this domain has no known function and is found in conserved hypothetical archaeal and bacterial proteins. COG1545: predicted nucleic-acid-binding protein containing a Zn-ribbon. (b) SCP-x thiolase domain is associated with sterol carrier protein (SCP)-x isoform and related proteins; Condensing_enzyme superfamily: family of enzymes that catalyze a (decarboxylating or non-decarboxylating) Claisen-like condensation reaction.

2. Experimental methods

Materials, strains, media, and general methods. Ferricenium hexafluorophosphate was purchased from Sigma-Aldrich (St. Louis, MO). Coenzyme A was purchased from MP Biomedicals (Solon, Ohio). Isopropyl β-D-1-thiogalactopyranoside was from Denville Scientific (Metuchen, NJ). Tryptone, HEPES, and TRIS were purchased from Fisher Scientific (Pittsburgh, PA). 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, factor Xa, and protein ladder were from New England Biolabs (Beverly, MA). HisTrap FF columns and Superdex 200 HiLoad 16/60 and 10/300 GL columns were from GE Healthcare Biosciences Corp. (Piscataway, NJ). Oligonucleotides were from IDT Inc. (Coralville, IA). Total genomic DNA of *M. tuberculosis* H37Rv was obtained from the TB Research Materials Facility at Colorado State University (Fort Collins, CO) (NIAD NO1-AI40091). MALDI mass spectra were acquired on a Bruker Autoflex II TOF/TOF. Big Dye DNA sequencing (Applied Biosystems, Foster City, CA; performed by the Stony Brook University Sequencing Facility) was used to verify the coding sequence of the expression plasmids. BL21(DE3) *E. coli* was obtained from BioRad. 2 × YT media is composed of 16 g tryptone, 10 g yeast extract and 5 g NaCl per liter. Buffer A: 20 mM Tris-HCl buffer pH 8.0, supplemented with 300 mM NaCl and 10 mM imidazole. Buffer B: 20 mM Tris-HCl buffer pH 8.0, supplemented with 300 mM NaCl and 500 mM imidazole. Buffer C: 50 mM Tris-HCl buffer pH 8.0, supplemented with 200 mM NaCl. The *casI* (*Ro05822*) gene sequence was synthesized by GenScript USA Inc. (Piscataway, NJ). Primary crystallization screens were from Hampton research.

Expression plasmid construction. *ChsH1*(*Rv3541c*) and *ChsH2* (*Rv3542c*) 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 similarly digested pET28b (Table 4-1). DNA sequencing of the plasmids confirmed that the sequence was correct and that no mutations were introduced during PCR. Active site mutants were prepared using quick change site-directed mutagenesis.[137] Briefly, the following primers, the mutated nucleotide colored in red, were designed to introduce the expected mutation. After thermal cycling amplification, 1 μL of the DpnI restriction enzyme (10 U/μL) was added to digest the parental strands. 2 μL of the DpnI treated DNA product was transferred into XL1-Blue competent cells. The mutations were confirmed by DNA sequencing.

pigr3H1$_{D29A}$ forward primer: 5'-GCTGGCTACCAGGGCCTTCCAGGATGTGC-3'

pigr3H1$_{D29A}$ reverse primer: 5'-CGACCGATGGTCCCGGAAGGTCCTACACG-3'

pigr3H1$_{H34A}$ forward primer: 5'-GGGACTTCCAGGATGTGGCTCATGATCGAGACAAGG-3'

pigr3H1$_{H34A}$ reverse primer: 5'-CCCTGAAGGTCCTACACCGAGTACTAGCTCTGTTCC-3'

The gene of *ltp2 (Rv3540c)* was 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 similarly digested pT7TEV-HMBP vector (Table 4-1). DNA sequencing of the plasmids confirmed that the sequence was correct and that no mutations were introduced during PCR.

Table 4-1. Expression constructs used in this work.

Construct name	Genes	Restriction sites	Purified enzyme	Source/Reference
pET28b	-	-	-	Novagen
pT7TEVHMBP	-	-	-	Garcia-Diaz lab
pChsH1	*Rv3541c*	NdeI/XhoI	N-His_6-ChsH1	This work[a]
pChsH2	*Rv3542c*	NdeI/NotI	N-His_6-ChsH2	This work[a]
pigr3	*Rv3542c, Rv3541c, Rv3540c*	NdeI/HindII	N-His_6-ChsH2-ChsH1	[50]
pChsH1ChsH2^N	*Rv3542c, Rv3541c, Rv3540c*	NdeI/HindII	N-His_6-ChsH2-ChsH1	This work[a]
pigr3H1$_{D29A}$	*Rv3542c, Rv3541c, Rv3540c*	NdeI/HindII	N-His_6-ChsH2-ChsH1$_{D29A}$	This work[a]
pigr3H1$_{H34A}$	*Rv3542c, Rv3541c, Rv3540c*	NdeI/HindII	N-His_6-ChsH2-ChsH1$_{H34A}$	This work[a]
pMBPltp2	*Rv3540c*	EcoRI/HindIII	N-His_6-MBP-Ltp2	This work[c]
pCasI	*Ro05822*	NdeI/XhoI	N-His_6-CasI	This work[b]

[a]Genes were cloned from H37Rv genomic DNA and ligated into pET28b with the indicated restriction sites to include an N-terminal His_6 fusion tag. [b]The *Rhodococcus jostii* RHA1 *casI (Ro05822)* gene was synthesized by Genescript and ligated into pET15b with the indicated restriction sites to include an N-terminal His_6 fusion tag. [C]Genes were cloned from H37Rv genomic DNA into pT7TEVHMBP vector with the indicated restriction sites to include an N-terminal His_6 fusion tag and MBP fusion.

To produce the C-terminal truncation of ChsH2, a factor Xa cleavage site, followed by 4 Gly residues, was introduced at Arg187 of ChsH2. Insertion was conducted by PCR with primers 5'- ATC GAC GGG CGT GGT GGT GGT CGT CCC TCG TCG TCG CGG GAC-3' and 5'-CAT CAT AGC GTC GGG ATC CAA ATC GTC AGG-3' to give construct *pChsH1ChsH2^N*. This work was performed by Dr. Suzanne Thomas.

Protein expression and purification. To express ChsH1 or ChsH2 individually, BL21(DE3) *E. coli* cells were transformed with construct *pChsH1* or *pChsH2*. 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} ~ 0.6 – 0.8 by the addition of 1 mM IPTG, and cells were grown for 20 h at 25 °C. Purified proteins were analyzed by reducing SDS-PAGE. CasI was expressed as described for ChsH1 or ChsH2 using construct *pCasI* and 100 μg/mL ampicillin in the media.

Construct *pMBPltp2* was used to express Ltp2 with the fusion of Maltose binding protein (MBP). 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} ~ 0.6 – 0.8 by the addition of 1 mM IPTG, and cells were grown for 20 h at 25 °C. Purified proteins were analyzed by reducing SDS-PAGE.

ChsH1-ChsH2 complex was obtained using construct *pigr3*^{50}. Expression was induced at OD_{600} ~ 0.6 – 0.8 by the addition of 1 mM IPTG and cells were grown 20 h at 25 °C. Cells were lysed by French press or sonication in Buffer A and cellular debris was removed by centrifugation at 125,000 × g for 1 h. ChsH1-ChsH2 was purified by IMAC, with HisTrap FF column and Buffers A and B. Protein was further purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with Buffer C. Active site mutants were expressed and purified following the method for *pigr3* using constructs *pigr3H1*$_{D29A}$ and *pigr3H1*$_{H34A}$.

ChsH1-ChsH2^{N} was expressed as described for *pigr3* and isolated by IMAC. ChsH2 was cleaved by factorXa at 23 °C in buffer C supplemented with 2 mM $CaCl_2$. Factor Xa protease reaction was monitored by SDS-PAGE and cleaved ChsH1-ChsH2^{N} was purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with Buffer C.

Solution-state biophysical analysis of ChsH1-ChsH2 and ChsH1-ChsH2^{N}. ChsH1-ChsH2 (10 mg/mL) and ChsH1-ChsH2^{N} (7 mg/mL) were analyzed by analytical gel-filtration on a Superdex 200 (10/300 GL) column (GE Healthcare). The column was equilibrated in Buffer C. Samples were eluted isocratically in Buffer C, monitoring at 220 and 280 nm. Several standard

proteins were analyzed under the same conditions to generate standard curves to estimate molecular weights of analyzed proteins.

Molecular weights were determined using analytical ultracentrifugation sedimentation equilibrium (Beckman Optima XL-A). ChsH1-ChsH2 (5.7 μM, 2.8 μM, and 1.4 μM) and ChsH1-ChsH2^N (7.3 μM, 3.7 μM, and 1.8 μM) were centrifuged at speeds of 10k, 20k, and 30k at 20 ºC. Scans were acquired after 18 and 20 h of centrifugation at each speed monitoring at 280 nm. The protein partial-specific volume of 0.7359 for ChsH1-ChsH2 and 0.7379 for ChsH1-ChsH2^N and a solvent density 1.0079 for buffer C were calculated using SEDNTERP. Data were fit globally to the ideal, single species model using Heteroanalysis to determine the molecular weight.

Protein complex stoichiometries of ChsH1-ChsH2 and ChsH1-ChsH2^N 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 ºC 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 ºC. MS spectra were deconvoluted using ESIprot 1.0[138] 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. Protein stoichiometries were determined from the ratio of the molar concentrations.

Synthesis of enoyl-CoAs. Trans-2-decenoyl-CoA and trans-2-octenoyl-CoA were synthesized using the mixed anhydride method.[139] Briefly, the mixed anhydride was prepared by

mixing the corresponding acid (0.2 mmol) with ethyl chloroformate (0.4 mmol) in 4 mL dry THF in the presence of TEA (56 μL). After approximately 25 min, the acid was completely converted into the corresponding anhydride according to TLC. The mixed anhydride was filtered though glass wool in a disposable Pasteur pipet into newly prepared lithium CoA solution (55 mg) dissolved in 5 mL of H_2O and THF (3:2 v/v) (pH 8). The reaction was stirred at room temperature (rt) for 2 days. The pH of the reaction mixture was adjusted to 3 and the unreacted acid was removed by extraction with ether. The enoyl-CoAs were purified using a semi-prep C18 column in a HPLC system using 20 mM ammonium bicarbonate, pH 8 with a linear gradient from 0 to 90% methanol in 20 mM ammonium bicarbonate. The purification process was monitored at 220 nm and 248 nm simutaneouly. The identities of collected peaks were confirmed by MALDI-TOF MS.

3-oxo-4-pregnadiene-20-carboxylic acid.[50, 114] Stigmasta-4, 22-diene-3-one (1 g) was dissolved in 200 mL CH_2Cl_2 and cooled to -78 °C. 1% (v/v) pyridine was added to selectively ozonolyze the double bond at the 22 position. O_2 was purged from the solution for 5 min, ozone was sparged until the solution turned blue, then O_2 was sparged until the blue color dissipated. Methyl disulfide was added and reaction was warmed to rt and allowed to stand for 12 h. The resulting aldehyde was purified by silica gel chromatography. Sodium chlorite (10 eq) was added to the aldehyde (dissolved in 10% H_2O/ACN) and stirred for 16 h. The reaction was monitored by Thin layer chromatrography (TLC). 3-oxo-4-pregnadiene-20-carboxylic acid was purified by silica gel chromatography. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.73 (s, 1H), 2.50-2.22 (m, 5H, 2.05-1.94 (m, 2H), 1.90-1.79 (m, 2H), 1.72 (dd, *J*=13.9, 5.3 Hz, 1H), 1.54 (s, 5H), 1.47-1.37 (m, 2H), 1.25 (d, *J*=6.9Hz, 3H), 1.19 (s, 3H), 0.98 (ddd, *J*=32.2, 22.8, 15.6 Hz, 4H), 0.74 (s, 3H) (Appendix Figure A-1).

3-oxo-4-pregnene-20-carboxyl CoA, 3-OPC-CoA. 3-oxo-4-pregnene-20-carboxylic acid (3-OPC) was dissolved to 50 mM in 94% ethanol containing 60 mM NaOH. The thioesterification was performed for 4 h at 22 °C in a 1 mL reaction volume containing 5mM 3-OPC, 10 μM CasI (Ro05822), 100 mM pH 7.4 HEPES, 5 mM $MgCl_2$, 2.5 mM ATP, and 1 mM CoASH.[97] The reaction was quenched by adding 1 mL of MeOH to precipitate the enzyme. The supernatant was isolated by centrifugation and filtration. 3-oxo-4-pregnadiene-20-carboxyl CoA was purified using a semi-prep C18 column in a HPLC system using 100 mM ammonium acetate, pH 4.5 with a linear gradient from 0 to 90% MeOH in 100 mM ammonium acetate to yield 3-OPC-CoA. The purification process was monitored at 220 nm and 248 nm simutaneouly. The identities of collected peaks were confirmed by MALDI-TOF MS.

3-oxo-pregna-4,17-diene-20-carboxyl-CoA, 3-OPDC-CoA. 3-oxo-4-pregnene-20-carboxyl CoA was re-dissolved to 2.8 mM in H_2O . The oxidation was performed for 3 h at 20 °C in 1mL of 0.5 mM 3-OPC-CoA, 3 μM ChsE1-ChsE2 (FadE28-FadE29), 100 mM pH 8.5 TAPS and 1 mM ferricenium hexafluorophosphate[56]. The product was purified using a semi-prep C18 column in a HPLC system in 100 mM ammonium acetate, pH 4.5 with a linear gradient of 90% MeOH. MeOH was removed under reduced pressure and H_2O was removed by lyophilization to yield 3-OPDC-CoA. The purification process was monitored at 220 nm and 248 nm simutaneouly. The identities of collected peaks were confirmed by MALDI-TOF MS.

Hydratase assay. ChsH1-ChsH2, ChsH1-ChsH2^N, $ChsH1_{D29A}$-ChsH2, and $ChsH1_{H34A}$-ChsH2 were assayed for hydratase activity with substrates octenoyl-CoA, decenoyl-CoA, and 3-oxo-pregna-4,17-diene-20-carboxyl CoA in 100 mM HEPES pH 7.4 buffer and 100 nM (50 nM for 3-oxo-pregna-4,17-diene-20-carboxyl CoA) enzyme at 25 ºC. Reactions were monitored at

263 nm and initial velocities were determined. Product formation was quantified using ε_{263nm} = 6,700 M^{-1} cm^{-1} that corresponds to the α,β-unsaturation of the enoyl-CoA substrates. Controls were run without enzyme or without substrate, and showed negligible decreases in absorbance at 263 nm. The rates of product formation were fit to the Michaelis-Menten equation (1) to determine K_M and k_{cat} for each enzyme/substrate pair. Hydration product formation was confirmed by analysis of the assay mixture by MALDI-TOF mass spectrometry.

$$v = V_m[S]/(K_m + [S]) \qquad (1)$$

Thiolase assay of MBP-Ltp2. 600 nM ChsH1-ChsH2 was pre-incubacted with 86 μM 3-oxo-pregna-4,17-diene-20-carboxyl CoA in 100 mM HEPES pH 7.4 buffer at 25 ºC for 2 hours. 10 μL of the assay mixture was collect as control and then MBP-Ltp2 (600 nM) was added to the assay. 10 μL of the assay mixture was collected after 1 hour incubation and 3 hours incubation in the presence of exogenously added MBP-Ltp2. All the collected samples were subjected to zip-tip extraction and then MALDI-TOF mass spectrometry.

Crystallization. Crystal screenings including Index, crystal screen, crystal screen 2, PEG/Ion, PEG/Ion 2, SaltRx 1, SaltRX 2, Natrix, Natrix 2, PEGRx1, PEGRx2, and MembFrac from Hampton research have been conducted in order to get crystals of full length ChsH1-ChsH2. The ChsH1-$ChsH2^N$ apoenzyme crystals were obtained by hanging drop vapor diffusion at rt. Briefly, 1 μL of 15 mg ml^{-1} protein was mixed 1:1 with a reservoir solution of 20%-24% PEG 3350, 20 mM $CaCl_2$, 20 mM $CdCl_2$, 20 mM $CoCl_2$ at pH 7 or pH 6.5 and equilibrated against 500 μL of the reservoir solution. The crystals were then harvested and transferred to a cryoprotectant solution containing 20 mM $CaCl_2$, 20 mM $CdCl_2$, 20 mM $CoCl_2$, 200 mM NaCl, 25% PEG 3350 at pH 7.0 and 17% glycerol. Co-crystallization: ChsH1-$ChsH2^N$ was mixed with

free CoA, octanoyl-CoA, decenoyl-COA, and 3-OPC-CoA. Crystals were obtained at 4 °C by hanging drop vapor diffusion, in mother liquor containing 1 mg/mL of each ligand, 20 mM $CaCl_2$, 20 mM $CdCl_2$, 20 mM $CoCl_2$, 200 mM NaCl at pH 7.0 and 25% PEG 3350. The crystals were harvested and transferred to a cryoprotectant solution containing 1 mg/mL of each ligand respectively, 20 mM $CaCl_2$, 20 mM $CdCl_2$, 20 mM $CoCl_2$, 200 mM NaCl, 25% PEG 3350 at pH 7.0 and 17% ethylene glycol. All crystals were flash cooled in liquid N_2 before data collection.

X-ray data collection and structure determination. Diffraction data for ChsH1-ChsH2^{N,} ChsH1-ChsH2^{N}:CoA, ChsH1-ChsH2^{N}:octanoyl-CoA and ChsH1-ChsH2^{N}:decenoyl-CoA were collected on beamline X25 of the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY) at a wavelength of 1.7 Å, and for ChsH1-ChsH2^{N}:3-OPC-CoA, data were collected on beamline 23ID-D of the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) at a wavelength of 1.075 Å. All datasets were processed using XDS[115] and Aimless[116] as implemented in autoPROC.[120] Four cadmium sites in ChsH1-ChsH2^{N} were located with SHELX,[140] phases were calculated to 1.54 Å via single-wavelength anomalous dispersion[119] using SHARP,[120] and automated modeling building with ARP/wARP[141] produced a starting model consisting of 85% of ChsH1 and 90% of ChsH2^{N}. For ChsH1-ChsH2^{N}:CoA, ChsH1-ChsH2^{N}:octanoyl-CoA, ChsH1-ChsH2^{N}:decenoyl-CoA and the ChsH1-ChsH2^{N}:3-OPC-CoA structure, phases were calculated via molecular replacement with Phaser, using ChsH1-ChsH2^{N} as a search model. Six cadmium sites were identified in the ChsH1-ChsH2^{N}:3OPC-CoA structure; four of them are located in the tetramer-tetramer packing interface, and two of them coordinate with Asp29/His34 in ChsH2^{N}, blocking the active site (Fig. 27b). For all structures, manual model building was carried out in Coot[142] followed by refinement with Refmac[125] and Phenix.[143] Final model quality was assessed using MolProbity or Phenix.[144]

Data collection and refinement statistics are shown in Table 2 and Table 3. This work was accomplished by collaborating with Dr. Kip Guja and Dr. Miguel Garcia-Diaz

Small-angle X-ray scattering of ChsH1-ChsH2 and ChsH1-ChsH2^N in solution. The freshly prepared ChsH1-ChsH2 was concentrated to 197 μM (18.89 mg/mL) and ChsH1-ChsH2^N was concentrated to 136 μM (9.3 mg/mL). Buffer C as references to subtract signals from background diffraction in both the cases for ChsH1-ChsH2 and ChsH1-ChsH2^N. Small-angle X-ray scattering (SAXS) data sets were collected for both the ChsH1-ChsH2 solutions (197 μM, 98.5 μM, 49.25 μM) and ChsH1-ChsH2^N (136 μM, 68 μM, 34 μM), respectively, at the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY) on Beamline X9 using an energy of 13.5 keV.[145] Solutions were transferred to a 96-well plate and placed in a water-thermostated sample holder (10 °C). Throughout the course of the experiment (about 30 sec per run), samples were flowed continuously through a 0.9 mm quartz capillary tube to minimize damage from X-ray radiation (~ 20 μL).

The scattering contribution of the ChsH1-ChsH2 or ChsH1-ChsH2^N protein complex was obtained by subtracting the buffer scattering profile from the protein solution scattering profile. Data processing was performed using a Python-based package developed at X9. The data were fit in GNOM to produce a well-behaved P(r) curve. P(r) curves were compared for three different concentrations in PRIMUS. Twenty protein shape (bead) models were reconstructed in DAMMIF, and then aligned and compared in DAMSEL in order to determine the most probable model. DAMSUP was used to align all models with the most probable model, and all the aligned models were averaged in DAMAVER to compute a probability map. DAMFILT was applied to filter the average model at a preset cut-off volume. The “damfilt” models were used to yield the final SAXS envelope structures.

Table 4-2. Data collection and refinement statistics for ChsH1-ChsH2^N and ChsH1-ChsH2^N:3-OPC-CoA

	ChsH1-ChsH2^N	ChsH1-ChsH2^N:3-OPC-CoA
Data collection		
Space group	*P* 2_1	*P* 6_1 2 2
Cell dimensions		
a, *b*, *c* (Å)	51.8, 134.1, 90.4	51.9, 51.9, 436.2
α, β, γ (°)	90, 91.6, 90	90, 90, 120
Resolution (Å)	37.6–1.52	72.70–1.76
R_{merge}	0.033 (0.434)[a]	0.068 (0.605)
R_{meas}	0.038 (0.585)	0.073 (0.642)
R_{pim}	0.026 (0.408)	0.024 (0.208)
$CC_{1/2}$	0.999 (0.886)	0.999 (0.852)
I / σI	19.2 (2.2)	18.8 (8.8)
Completeness (%)	99.2 (95.7)	100 (100)
Multiplicity	13.4 (11.2)	8.0 (8.8)
Wilson B-factor	17.04	21.19
Refinement		
Resolution (Å)	37.6–1.52	72.70–1.76
No. reflections	180,565	40,369
R_{work} / R_{free}	0.1768 / 0.2046	0.1895 / 0.2322
CC*	1.0 (0.976)	1.0 (0.963)
No. atoms		
Total	10,103	2,607
Protein	8,920	2,286
Ligand	-	76
Water	1163	263
B-factors		
Protein	24.0	34.8
Ligand	-	41.0
Water	33.7	40.6
R.m.s deviations		
Bond lengths (Å)	0.009	0.016
Bond angles (°)	1.15	1.57
PDB ID	4W78	4W7B

[a]Values in parentheses are for the highest-resolution shell.

Table 4-3. Data collection and refinement statistics for ChsH1-ChsH2^N: 3-OPDC-CoA, ChsH1-ChsH2^N:Octanoyl-CoA and ChsH1-ChsH2^N:Decenoyl-CoA.

	ChsH1-ChsH2^N: 3-OPDC-CoA	ChsH1-ChsH2^N: Octanoyl-CoA	ChsH1-ChsH2^N: Decenoyl-CoA
Data collection			
Space group	*P* 6_1 2 2	*P* 2_1	*P* 2_1
Cell dimensions			
a, *b*, *c* (Å)	51.851, 51.851, 436.702	51.777, 132.467, 53.029	51.842, 133.688, 90.563
α, β, γ (°)	90, 90, 120	90, 119.06, 90	90, 91.54, 90
Resolution (Å)	44.67 - 2.221	43.75 - 1.31	37.62 - 1.663
$I / \sigma I$	21.69 (3.53)[a]	16.48 (3.91)	14.16 (2.31)
Completeness (%)	100.00 (100.00)	98.87 (95.44)	99.18 (99.56)
Wilson B-factor	35.75	13.03	20.31
Refinement			
Resolution (Å)	44.67 - 2.221	43.75 - 1.31	37.62 - 1.663
Unique reflections	18662 (1785)	147574 (14240)	143065 (14336)
R_{work}	0.2472 (0.3478)	0.1794 (0.2261)	0.1728 (0.2381)
R_{free}	0.2722 (0.4134)	0.1920 (0.2469)	0.2033 (0.2709)
No. atoms			
Total	2443	5374	10100
Protein	2269	4597	9006
Ligand	77	119	140
Water	97	658	954
B-factors			
Protein	53.30	19.30	26.70
Ligand	40.70	32.90	33.00
Water	45.20	30.80	36.20
R.m.s deviations			
Bond lengths (Å)	0.005	0.012	0.011
Bond angles (°)	1.12	1.56	1.46
Ramachandran	95	98	98
Ramachandran	1.4	0	0.085
Clashscore	1.40	8.50	2.10

[a]Values in parentheses are for the highest-resolution shell.

3. Results and Discussion

The *igr* operon is required for conversion of the β-oxidation intermediate 3-oxo-pregn-4-ene-20-carboxyl-CoA to androstenedione, which concomitantly results in the formation of propionyl-CoA (Figure 4-2a, 4-2b).[50] The *igr* operon encodes two potential MaoC-like enoyl-CoA hydratases based on protein BLAST against non-redundant protein sequences database, ChsH1 (Rv3541c) and ChsH2 (Rv3542c). However, their role in cholesterol side chain degradation was unclear since known fatty acid β-oxidation cycles in bacteria, as well as eukaryotes, utilize a crotonase family member to catalyze hydration of fatty enoyl-CoA metabolites. MaoC-like enoyl-CoA hydratases are structurally distinct from crotonases, which in bacteria are employed in polyhydroxyalkanoate energy storage pathways instead of for energy metabolism.[146] In eukaryotes, MaoC-like enoyl-CoA hydratases are employed in peroxisomal β-oxidation of di- and trihydroxy-5β-cholestanic acid to chenodeoxycholic acid and cholic acid, respectively. However, these hydratases are part of a multifunctional enzyme (MFE) that includes a fused 3-hydroxy-acyl-CoA dehydrogenase domain.[147, 148] Both the bacterial and peroxisomal MaoC-like enoyl-CoA hydratases characterized to date are all homodimers, although the peroxisomal enoyl-CoA domains are larger than their bacterial homologs. Earlier pull-down experiments suggested that ChsH1 (Rv3541c) and ChsH2 (Rv3542c) form a heteromeric complex.[50] Given the unclear structural and functional homologies of ChsH1 and ChsH2 as well as their association, we undertook investigation of the structure and activity of the ChsH1-ChsH2 complex in order to understand its role in *Mtb* cholesterol metabolism.

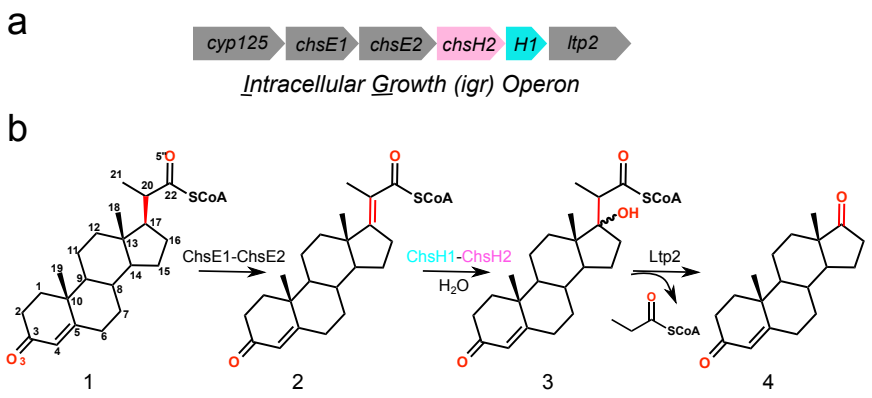


Figure 4-2. Biochemical function of individual *igr* genes. (a) The *igr* operon in *Mtb*; *chsH1* (*Rv3541c*) and *chsH2* (*Rv3542c*) are colored in cyan and pink, respectively. Other genes are shown in gray. (b) The biochemical function of the gene products encoded by the *igr* operon. **1**: 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA); **2**: 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA); **3**: 17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA (17-HOPC-CoA); **4**: androstenedione.

ChsH1-ChsH2 forms an obligate $\alpha_2\beta_2$ heterotetramer. Previously, we showed that ChsH1-ChsH2 can be isolated as a protein complex when coexpressed in *E. coli* from a single plasmid construct.[50] Initially, ChsH1 and ChsH2 were expressed individually as N-terminally His_6-tagged proteins in *E. coli*. However, both ChsH1 and ChsH2 precipitated directly after elution from an ion affinity chromatography (IMAC) column indicating ChsH1 and ChsH2 are not stable under these expression conditions.

Construct *pigr3*[50] (Table 4-1) was expressed heterologously to provide ChsH1-ChsH2 that was isolated by IMAC and further purified using gel filtration chromatography. SDS-PAGE gel analysis confirmed the co-isolation of ChsH1 and ChsH2 (Figure 4-3a and 4-3b). Analytical gel filtration chromatography demonstrated that ChsH1-ChsH2 forms a stable oligomeric complex in solution. Further analysis by analytical ultracentrifugation (AUC) equilibrium experiments established that ChsH1-ChsH2 forms a complex with a molecular weight of 99 kDa (Figure 4-4a). Given the individual molecular weights of ChsH1 and N-terminally His_6-tagged ChsH2 of

14 kDa and 36 kDa, respectively, the ChsH1-ChsH2 complex is an $\alpha_2\beta_2$ heterotetramer. To confirm the stoichiometry of the complex, the assembly was analyzed by LC/UV/MS (Figure 4-4b). Integration of the ChsH1 and ChsH2 peaks and calculation of their relative molar concentration provided a relative molar stoichiometry of one ChsH1 to one ChsH2, consistent with formation of an $\alpha_2\beta_2$ heterotetramer. This work was performed by Dr. Suzanne Thomas.

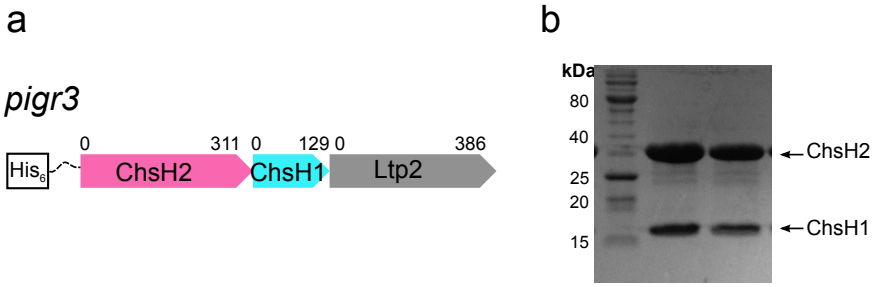


Figure 4-3. ChsH1 forms a enzyme complex with ChsH2. (a) Constructs for expression of ChsH1-ChsH2. An N-terminal 6×His tag is encoded at the 5' end of the first gene. The number of residues encoded by each gene is labeled. (b) SDS-PAGE analysis of proteins isolated by IMAC. Protein identities were confirmed by tryptic digest and MALDI-TOF MS fingerprinting.

We undertook solution of the structure of this unique heteromeric enoyl-CoA hydratase that metabolizes steroids. In the process of crystallizing full length ChsH1-ChsH2, diffraction quality crystals were obtained. The crystals diffracted to 1.54 Å resolution (Table 4-2) and the structure was solved using single wavelength anomalous dispersion. The crystals contained four $\alpha\beta$ heterodimers (two $\alpha_2\beta_2$ heterotetramers) in the asymmetric unit (Figure 4-5a). The two tetramers in one ASU are not identical. Helix α1, α5 and loop I, II in ChsH2^N shift positions between the two tetramers. The relative positions of ChsH1 are more conserved compared to ChsH2^N (Figure 4-5b). The electron density was of excellent quality, and allowed a model of full length ChsH1 and the N-terminal domain of ChsH2 (180 amino acids out of 311 amino acids) to be built (Figure 4-3 and Figure 4-6).

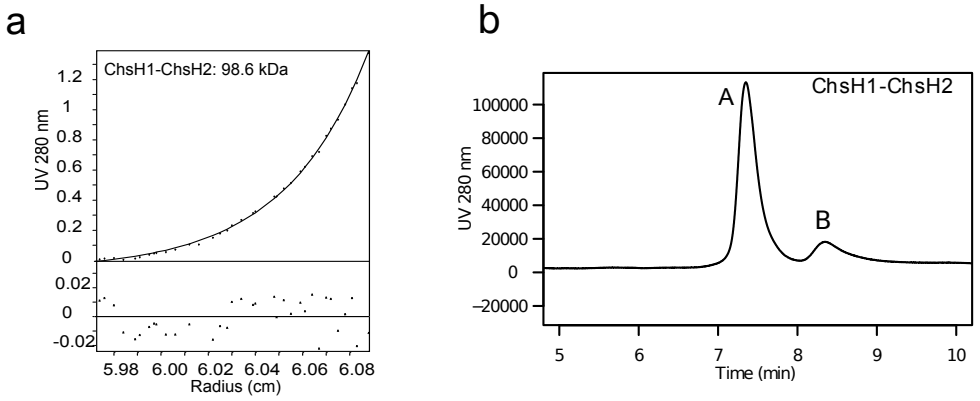


Figure 4-4. ChsH1-ChsH2 are heterotetramers in solution. (a) ChsH1-ChsH2 (5.7 μM, 2.8 μM, and 1.4 μM) was centrifuged at speeds of 10k, 20k, and 30k rpm at 20 ºC in an analytical ultracentrifuge. A representative fit is shown. The solid line shows the fit of the data to the ideal species model and the residuals of the fit are graphed below the fit. The best global fit provided molecular weights of 98.6 kDa for ChsH1-ChsH2. (b) Reverse phase LC/UV/MS chromatograms of ChsH1-ChsH2. Peaks A and B were identified as ChsH2 and ChsH1, respectively by deconvolution of multiple charged states in the corresponding ESI+ MS spectra. The absorbance of peaks were integrated and relative concentrations were determined from the calculated extinction coefficients of ChsH2 and ChsH1 [ε_{280} (ChsH2)= 59993 M^{-1} cm^{-1} and ε_{280} (ChsH1)= 16960 M^{-1} cm^{-1}].

The absence of further density for the C-terminal domain of ChsH2 prompted us to analyze the protein content of our crystals. Further analysis by SDS-PAGE and MALDI-TOF mass spectrometry confirmed that an approximately 130 amino acid C-terminal fragment of ChsH2 was lost through proteolysis prior to crystallization. We henceforth refer to this fragment as $ChsH2^{C}$.

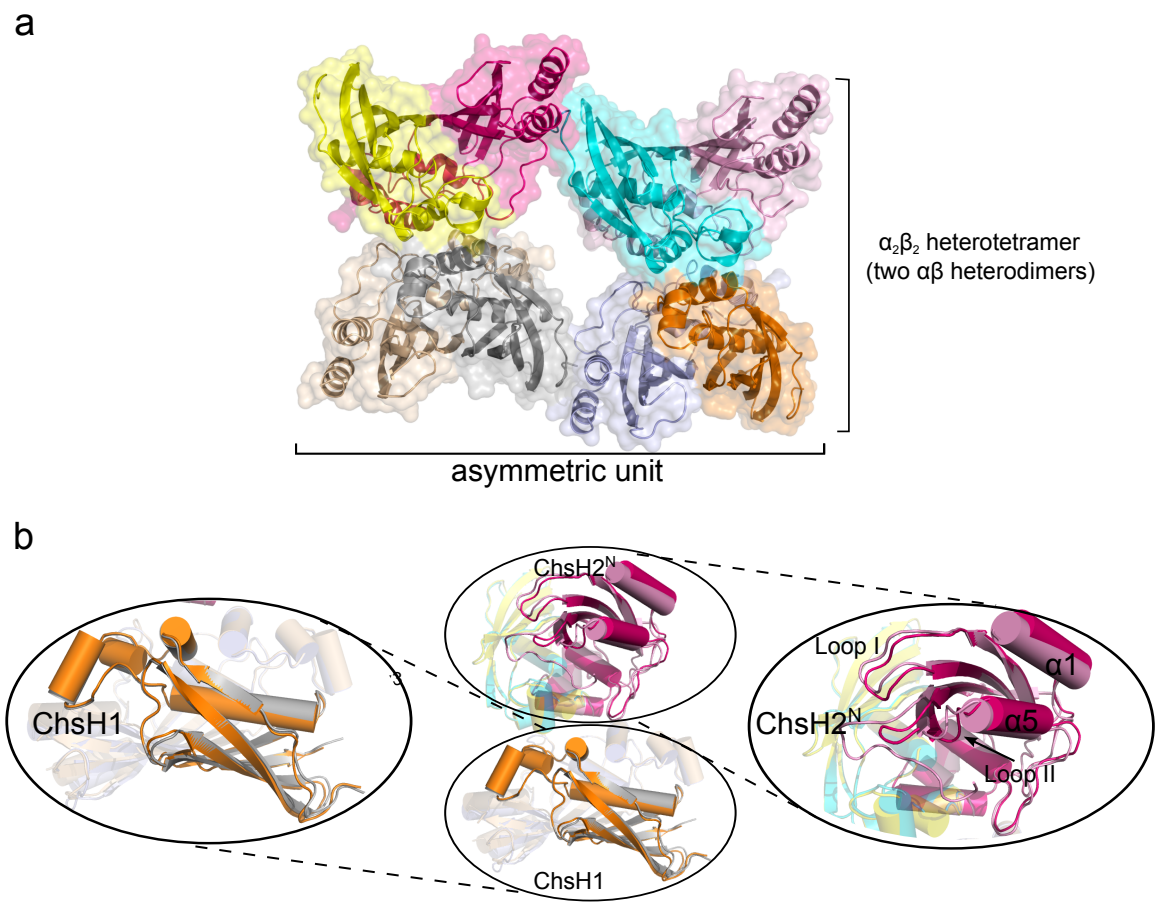


Figure 4-5. Overall architecture of ChsH1-ChsH2^N. (a) Two $\alpha_2\beta_2$ heterotetramers of ChsH1-ChsH2^N in one ASU. Surface representation of the structure corresponds to chain color. (b) Superimposition of two heterotetramer structures from one asymmetric unit. One set of overlayed ChsH1 chains (orange and gray) and ChsH2^N chains (magenta and pink) are labeled. The other set of overlayed ChsH1 and ChsH2^N chains is rendered transparent. The conformation of ChsH1 is conserved without any obvious variation between heterotetramers. The variations of secondary structures in ChsH2^N are labeled.

Purification and characterization of ChsH1-ChsH2^N. In order to reliably obtain the protein complex ChsH1-ChsH2^N and to reproduce the crystallization, an expression plasmid, *pChsH1ChsH2^N* (Table 4-1), was constructed that introduced a factor Xa cleavage site followed by 4 Gly residues after Arg187 of ChsH2 (Figure 4-6). The ChsH1-ChsH2 protein complex was obtained upon cistronic expression using the construct *pChsH1ChsH2^N*. The full-length protein was purified using IMAC and gel filtration chromatography and then cleaved with factor Xa to

provide ChsH1-ChsH2^N as confirmed by SDS-PAGE gel and analytical gel filtration (Figure 4-6b). The molecular weight of ChsH1-ChsH2^N was determined to be 65 kDa by AUC (Figure 4-7a) consistent with the predicted molecular weights of ChsH1 and His_6 tagged ChsH2^N (14 kDa and 23 kDa, respectively). The relative molar stoichiometry of the subunits remained 1:1 based on LC/UV/MS analysis (Figure 4-7b). Therefore ChsH1-ChsH2^N still forms an $\alpha_2\beta_2$ heterotetramer. The C-terminal fragment of ChsH2, referred to as ChsH2^C, precipitated during factor Xa cleavage. This work was performed by Dr. Suzanne Thomas.

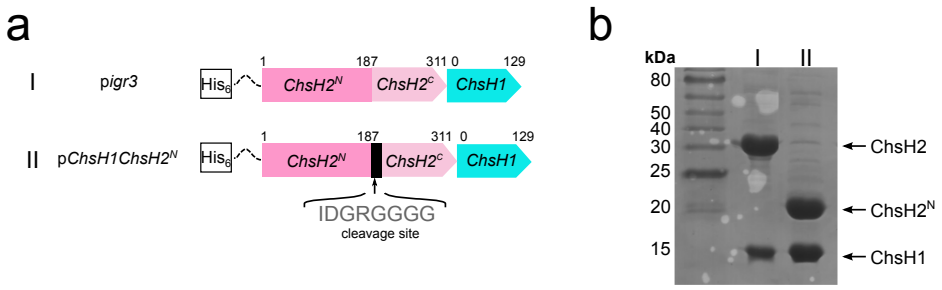


Figure 4-6. Construct for expression of ChsH1-ChsH2^N. (a) Constructs for expression of ChsH1-ChsH2 and ChsH1-ChsH2^N. An N-terminal 6×His tag is encoded at the 5' end of the first gene of each construct. The *ltp2* gene following *ChsH1* is not shown for clarity. The number of residues encoded by each gene is labeled. The factor Xa cleavage sequence is introduced after the Arg187 codon in the *chsH2* gene. (**b**) SDS-PAGE analysis of system I isolated by IMAC, and of system II isolated by IMAC, then subjected to factor Xa cleavage. Protein identities were confirmed by tryptic digest and MALDI-TOF MS fingerprinting.

Furthermore, we were able to obtain crystals of the new protein complex ChsH1-ChsH2^N. These crystals diffracted to 1.54 Å and yielded a structure identical to that originally obtained.

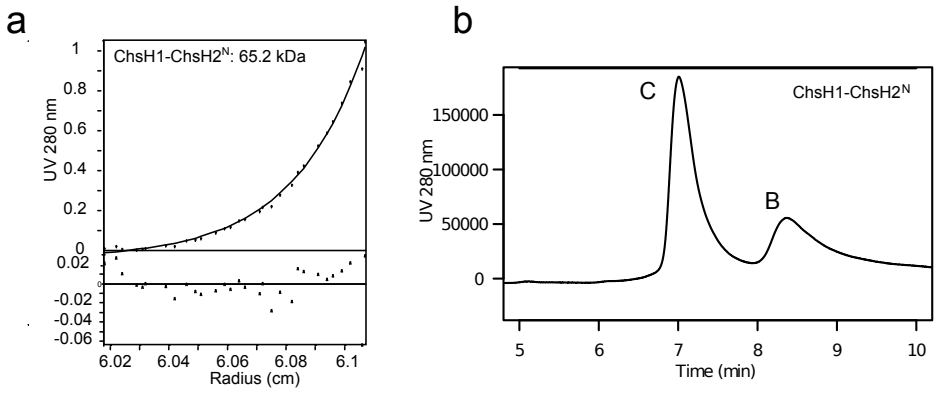


Figure 4-7. ChsH1-ChsH2^N are heterotetramers in solution. (a) ChsH1-ChsH2^N (7.3 μM, 3.7 μM, and 1.8 μM) were centrifuged at speeds of 10k, 20k, and 30k rpm at 20 ºC in an analytical ultracentrifuge. A representative fit for each sample is shown. The solid line shows the fit of the data to the ideal species model and the residuals of the fit are graphed below the fit. The best global fit provided molecular weights of 65.2 kDa for ChsH1-ChsH2^N. (b) Reverse phase LC/UV/MS chromatograms of ChsH1-ChsH2^N. Peaks B and C were identified as ChsH1 and ChsH2^N respectively by deconvolution of multiple charged states in the corresponding ESI+ MS spectra. The absorbance of peaks were integrated and relative concentrations were determined from the calculated extinction coefficients of ChsH1 and ChsH2^N [ε_{280} (ChsH1)= 16960 M^{-1} cm^{-1}, and ε_{280} (ChsH2^N)= 33523 M^{-1} cm^{-1}].

The structure of ChsH1-ChsH2^N reveals a unique MaoC-like enoyl-CoA hydratase architecture. Consistent with annotations based on BLAST alignments, the crystal structure shown here establishes that ChsH1-ChsH2^N belongs to the MaoC-like enoyl-CoA hydratase family. This family contains homodimeric MaoC-like enoyl-CoA hydratases that are comprised of a β-sheet that wraps around a central α-helix, which is known as a hot-dog fold. A second structural motif in MaoC-like hydratases is a three helix segment, referred to as an active site housing segment, and this motif is usually fused with the hot-dog fold domain in the MaoC-like hydratase family (Figure 4-8).[149]

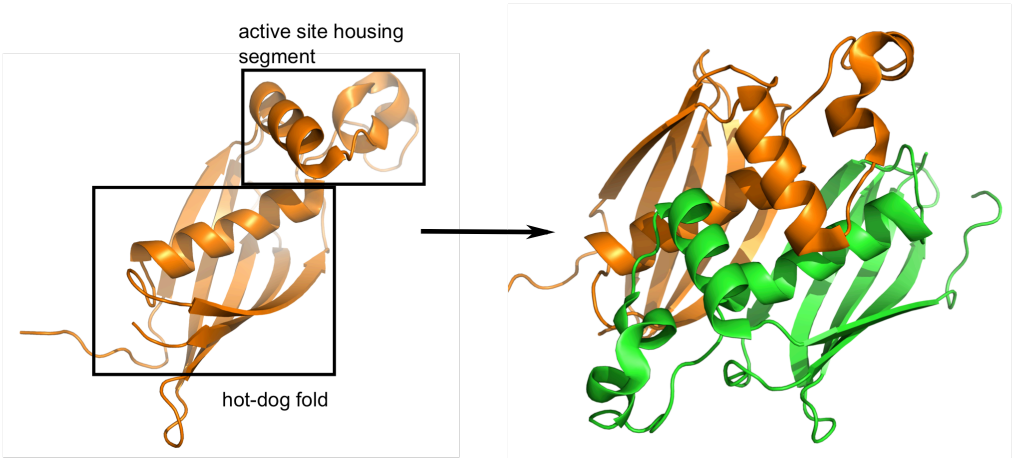


Figure 4-8. Homodimeric MaoC-like enoyl-CoA hydratase from *Aeromonas caviae* (PDB code:1IQ6). Each monomer consists of a hot-dog fold, which a β-sheet that wraps around a central a-helix, and a second motif bearing the active site. Two monomers dimerize to form homodimer.

Surprisingly, rather than the prototypical homodimeric structure, ChsH1-ChsH2^{N} forms a dimer of a unique αβ heterodimeric assembly comprising one standard and one nonstandard MaoC-like hydratase. In the ChsH1-ChsH2^{N} tetramer, one ChsH1 chain and one ChsH2^{N} chain form an intimate heterodimer. Two of these heterodimers combine to form the heterotetramer (Figure 4-9a and 4-9b). As far as we know, ChsH1-ChsH2 is the first example of an $\alpha_2\beta_2$ heteromeric MaoC-like enoyl-CoA hydratase.

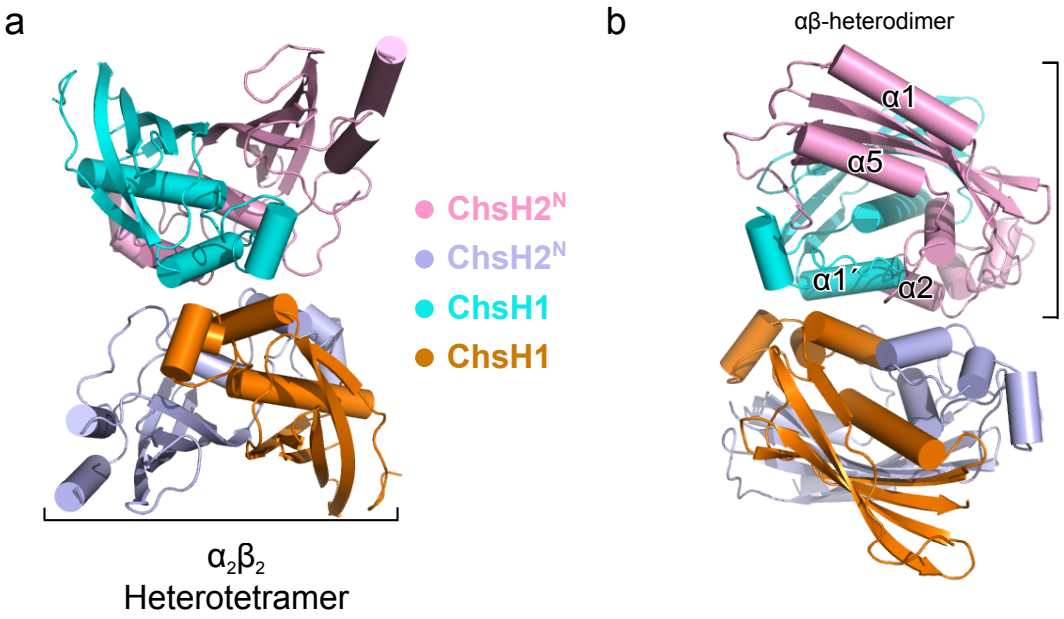


Figure 4-9. ChsH1-ChsH2^N forms a dimer of a unique αβ heterodimeric assembly comprising one standard and one nonstandard MaoC-like hydratase. (a) (b) ChsH1-ChsH2^N adopts a heterotetrameric assembly with two ChsH1 chains and two ChsH2^N chains. One ChsH1 chain and one ChsH2^N chain fold into a heterodimer and a pair of heterodimers assembles into a tetramer. The identities of the chains are labeled. The structure in (b) is obtained by a 90° rotation of the structure in (a) around the Y-axis.

The interaction between the two heterodimers involves helices α1' from ChsH1 and α2 from ChsH2^N. The helices from one heterodimer form a compact interface with the other heterodimer through salt bridges and hydrogen bonding interactions (Figure 4-10a, 4-10b and 4-10c). Arg37 from ChsH1 in one heterodimer forms a double salt bridge with Asp38 from the second ChsH1. Because of the two fold rotational symmetry in the dimer-dimer interface, four salt bridges are formed. Six hydrogen bonds are contributed from helices α2 of the ChsH2^N chains (Figure 4-10e and 4-10f).

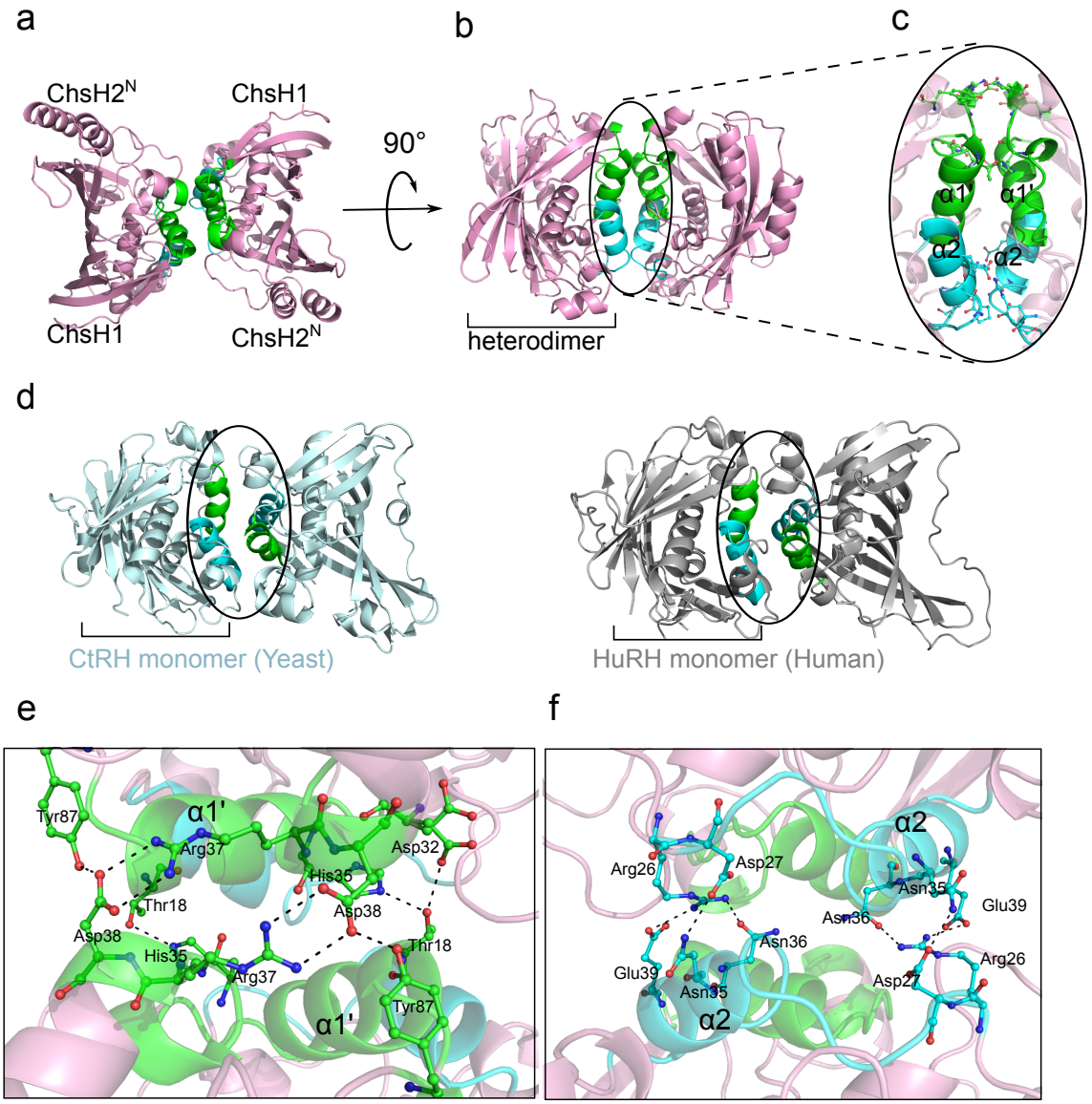


Figure 4-10. Dimer-dimer interface in ChsH1-ChsH2^N heterotetramer. (a) Overall structure of ChsH1-ChsH2^N tetramer. The chain identities are labeled. All the chains are colored in pink except the secondary structures in the dimer-dimer interface. Helices a1' from ChsH1 (green) and helices a2 from ChsH2^N (cyan) build up the heterodimer-heterodimer interface. (b) The structure from (a) is rotated 90° about the X-axis. The secondary structure elements in the interface are labeled. (c) Magnification of the dimer-dimer interface. Residues that are responsible for the dimer-dimer interactions are shown as sticks. Interacting residues from ChsH1 chains and ChsH2^N chains are colored in green and cyan, respectively. (d) The monomer-monomer interface in AcRH and HuRH. Helices from the C-terminal of AcRH or HuRH (green) and helices from the N-terminal of AcRH or HuRH (cyan), which are counterparts of a1' and a2, respectively, constitute the monomer-monomer interface.[150, 151] (e) Tyr87 and Thr18 from one ChsH1 chain form hydrogen bonds with Asp38 and His35, respectively, from the second ChsH1 chain. One conformation of Asp32 in the first ChsH1 chain forms hydrogen bonds with Thr18 of

the second ChsH1 chain. (**f**) Arg26 and Asp27 from one ChsH2^N chain hydrogen bond with Glu39 and Asn36, and Asn35, respectively from the other ChsH2^N chain.

Importantly, this heterodimer-heterodimer interface is clearly divergent with respect to other MaoC-like hydratase homologs. A DALI[152] search revealed homologous MaoC-like hydratases from several species, ranging from bacteria (*Aeromonas caviae*; AcRH) and yeast (*Candida tropicalis*; CtRH) to mammals (*Homo Sapiens*; HuRH). Interestingly, the ChsH1-ChsH2^N heterotetramer is comparable to a CtRH homodimer and a HuRH homodimer. The two monomers in CtRH or HuRH dimerize through a similar four-helix bundle (Figure 4-10d). However, in those cases, the helices from one monomer interact with the helices from the other monomer orthogonally[151] in a very different structural arrangement from ChsH1-ChsH2 (Figure 4-10b and 4-10d).

The ChsH1-ChsH2^N heterodimer is structurally distinct from its MaoC-like hydratase homologs. In the ChsH1-ChsH2^N heterodimer, five β-strands (β1'–β3'–β4'–β5'–β2') from ChsH1 and four β-strands (β1–β4–β3–β2) from ChsH2^N form an antiparallel nine-stranded β-sheet. Helix α3' from ChsH1 and α4 from ChsH2^N are above the antiparallel β-sheet in the viewpoint shown in Figure 4-11. Helices α2–η1–α3 from ChsH2^N and α1'–η1'–α2' from ChsH1, form a third layer on top of α3' and α4, and are almost two-fold symmetric (Figure 4-11a and 4-11b). This architecture and the extensive interface between the monomers imply that ChsH1 and ChsH2^N form an obligate heterodimer.

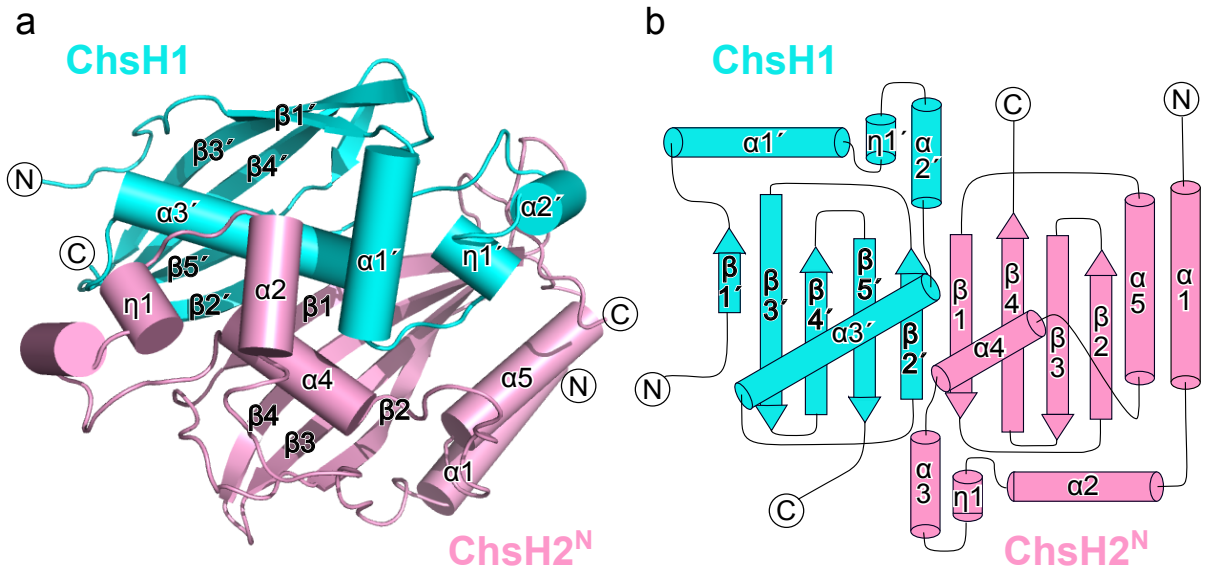


Figure 4-11. ChsH1-ChsH2^N heterodimer. (a) Schematic representation of one heterodimer from the tetramer. Each chain possesses a conserved "hot-dog" fold. α-helices and β-strands from ChsH1, ChsH2^N are shown as ribbons and arrows, respectively; C-termini and N-termini are labeled. 3_{10}-helices are labeled η. (b) Topology representation of one heterodimer. The color-coding is the same as in Figure 4-9. Secondary structure was assigned using DSSP.

The overall structure of the ChsH1-ChsH2^N heterodimer from *Mtb* is disparate from its homologs across different species. The ChsH1-ChsH2^N heterodimer has two hot-dog folds in total. Interestingly, a similar architecture is present in other structural homologs, but the two hot-dog folds are present as a single monomer (CtRH or HuRH), or a homodimer (AcRH) (Figure 4-12 and Table 4-4). The fold of ChsH1 is most similar to the C-terminal domain of the CtRH or HuRH monomer and corresponds to the structure of a single chain of AcRH (Figure 4-12 and Table 4-4). Surprisingly, ChsH2^N, although similar to the N-terminal domains of CtRH and HuRH monomers, deviates from the standard hot-dog fold of AcRH. In addition, both the structure and sequence have low similarity in the active site housing segment.

Table 4-4. Comparison of MaoC-like enoyl-CoA hydratases across species.

RH	CHD	ICHD	Number of binding sites	Substrate (aliphatic chain)	Function
AcRH homodimer[146]	Both monomers	None	2	4-6 carbon[153]	Produce (*3R*)-hydroxylacyl-CoA in PHA biosynthesis
CtRH monomer[150]	C-CtRH	N-CtRH	1	10-22 carbon	Peroxisomal b-oxidation of fatty acid to produce (*3R*)-hydroxyacyl-CoAs
HuRH monomer[151]	C-HuRH	N-HuRH	1	Methyl-branched;	Breakdown of very-long or methyl-branched fatty acids; bile acid synthesis
ChsH1-ChsH2 heterodimer	ChsH1	ChsH2	1	steroid 3-carbon side chain	β-oxidation of cholesterol in *Mtb*

CHD: complete hot-dog fold; ICHD: incomplete hot-dog fold; AcRH: *Aeromonas caviae* MaoC-like enoyl-CoA hydratase; CtRH: *Candida tropicalis* MaoC-like enoyl-CoA hydratase; HuRH: *Homo sapiens* MaoC-like enoyl-CoA hydratase; C-CtRH; C-HuRH: C-terminal domain of CtRH; C-terminal domain of HuRH; N-CtRH; N-HuRH: N-terminal domain of CtRH; N-terminal domain of HuRH

ChsH1 possesses a canonical MaoC-like hydratase hot-dog fold providing the active site, but ChsH2^{N} possesses a modified hot-dog fold that confers the ability to bind bulky substrates. ChsH1 includes a complete hot-dog fold and an active site housing segment (Figure 4-13a). Five antiparallel β strands (β1'-β3'-β4'-β5'-β2') act as a bun, wrapping around the sausage-like central helix α3' to form a standard hot-dog fold (Figure 4-13a), characterized by the four-turns in the central α-helix.[150] The region from α1', η1' to α2' comprises the active site housing segment. The strictly conserved active site residues, Asp29 and His34 (Figure 4-14), located in the active site housing segment (α1'–η1'–α2') in ChsH1 (Figure 4-13a), are almost in identical positions with respect to the active sites in AcRH, CtRH and HuRH (Figure 4-14a).

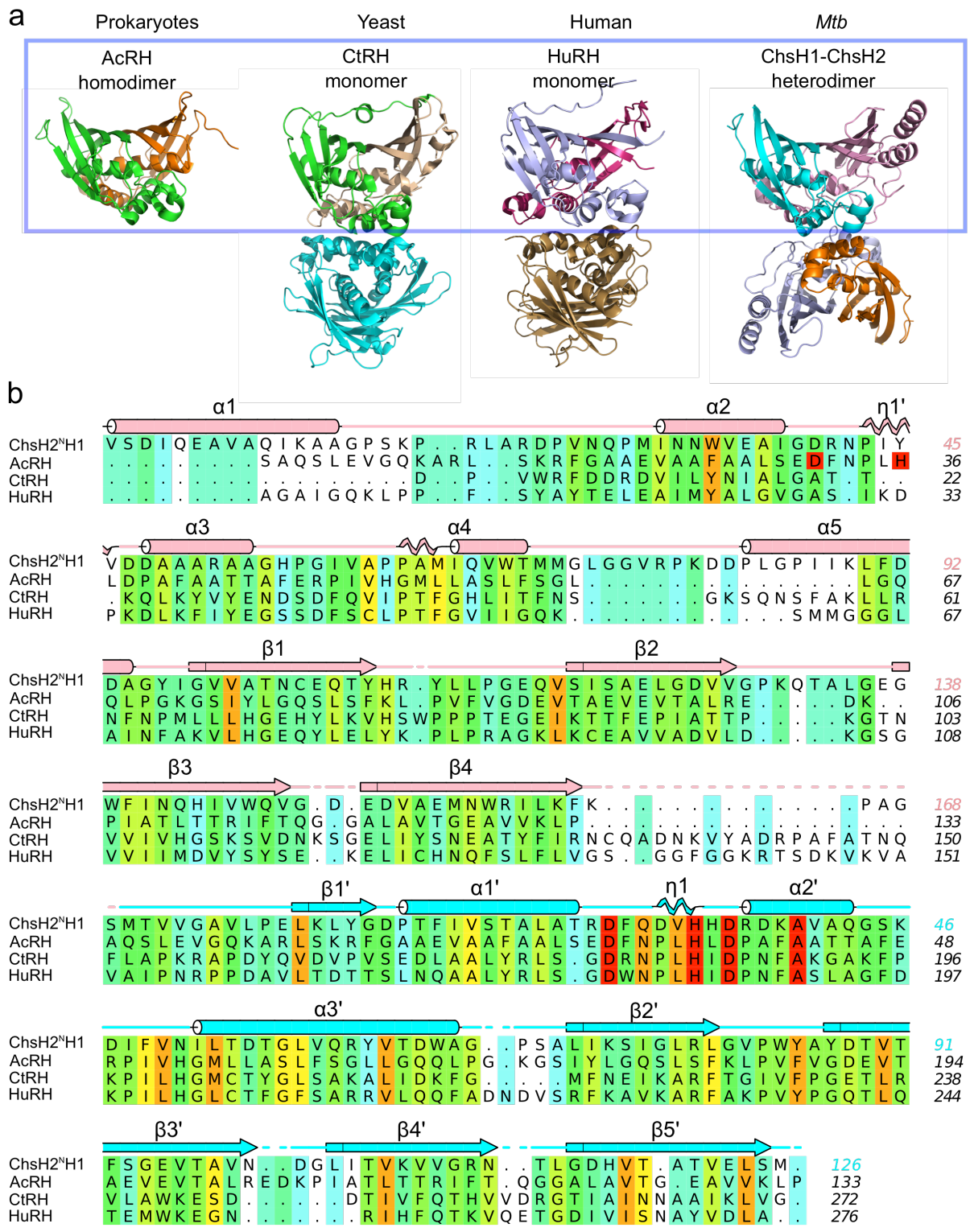


Figure 4-12. The 3D structural comparison and protein sequence alignment between ChsH1-ChsH^N with its homologs across different species (a) The overall structure of the ChsH1-ChsH2^N heterodimer from *Mtb* is disparate from its homologs across different species. (b) Structural

sequence alignment between ChsH1-ChsH2^N dimer, AcRH homodimer, CtRH monomer (PDB code: 1PN2), and HuRH monomer (PDB code: 1S9C). The sequence of ChsH2^N (1-178) and the sequence of ChsH1 (1-127) overlay with the N-terminal and C-terminal sequences of the monomers, respectively. The heterodimer sequence is referred to as ChsH2^{N}H1. Secondary structure elements are labeled and color coded following Figure 4-11. The residues are colored based on their conservation. The calculated RMS is 1.969 Å with 969 atoms aligned after the superimposition between AcRH monomer and ChsH1-ChsH2^N dimer. The calculated RMS is 2.331 Å with 818 atoms aligned after the superimposition between CtRH monomer and ChsH1-ChsH2^N dimer. The calculated RMS is 1.826 Å with 1009 atoms aligned after the superimposition between HuRH monomer and ChsH1-ChsH2^N dimer. The protein sequence of ChsH2^{N}H1 dimer shares 47.4% sequence identity with CtRH monomer and shares 47.1% with HuRH monomer.

To establish that Asp29 and His34 correspond to the active site of the ChsH1-ChsH2 complex, Asp29 was mutated to Ala, and His34 was mutated to Ala in independent constructs (Table 4-1). The purified protein complexes were analyzed by SDS-PAGE and by analytical size exclusion chromatography. They both had the same elution profile as wild-type ChsH1-ChsH2. Therefore, ChsH1$_{D29A}$-ChsH2 and ChsH1$_{H34A}$-ChsH2 still form $\alpha_2\beta_2$ heterotetramers. The activities of these two mutant protein complexes were tested with octenoyl-CoA, decenoyl-CoA and 3-OPDC-CoA. ChsH1$_{D29A}$-ChsH2 retained 2%-4% of the wild type activity. Moreover mutation of His34 to Ala abolished the catalytic activity within detection limits (less than 0.0062% activity) (Figure 4-14b).

In striking contrast to ChsH1, the structure of ChsH2^N varies substantially with respect to the standard MaoC-like hydratase structure, e.g., ChsH1, AcRH or the C-terminal domains of CtRH or HuRH. Only four antiparallel β-strands wrap around the central helix α4. Furthermore, helix α4 contains two turns, and one turn is a 3_{10} helix (Fig. 4-15a). The short central helix (α4) renders the hot-dog fold incomplete (Figure 4-15b and 4-15c). This short, modified helix results in an open binding pocket formed by the 4-stranded β-sheet. Helix α1 and α5 are two new flexible elements (Figure 4-15) that are not present in other known MaoC-like hydratase

structures. Even though the active site housing segment (α2– η1–α3) is structurally conserved, the active site His is replaced with a Tyr (Figure 4-15).

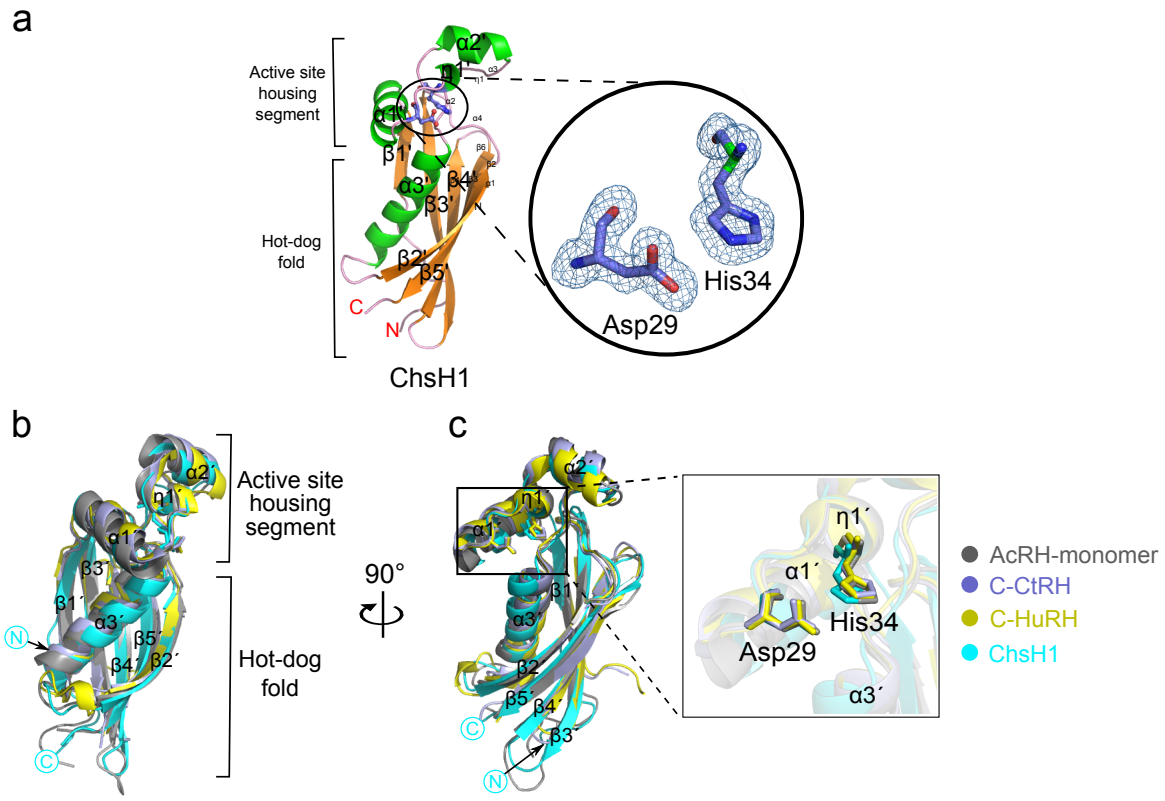


Figure 4-13. ChsH1 possesses a canonical MaoC-like hydratase hot-dog fold providing the active site. (a) ChsH1 possesses a standard hot-dog fold. The cartoon representations are colored by secondary structure. N-termini and C-termini are labeled. In ChsH1, the five stranded b-sheet (β1'-β3'-β4'-β5'-β2') wraps around α3', forming a hot-dog fold and the proposed active site housing segment is comprised of α1', η1' and α2'. The potential active site residues are colored blue and the simulated annealing omit maps of those residues are calculated to reduce the effects of model bias. Electron density is shown as mesh (σ=3). (b) (c) 3-D structural conservation between AcRH monomer (gray), C-terminal domain of CtRH (C-CtRH) (residue number: 163-280; purple), C-terminal domain of HuRH (C-HuRH) (residue number: 174-298; yellow) and ChsH1 chain (cyan). The conserved active sites are enlarged in the box.

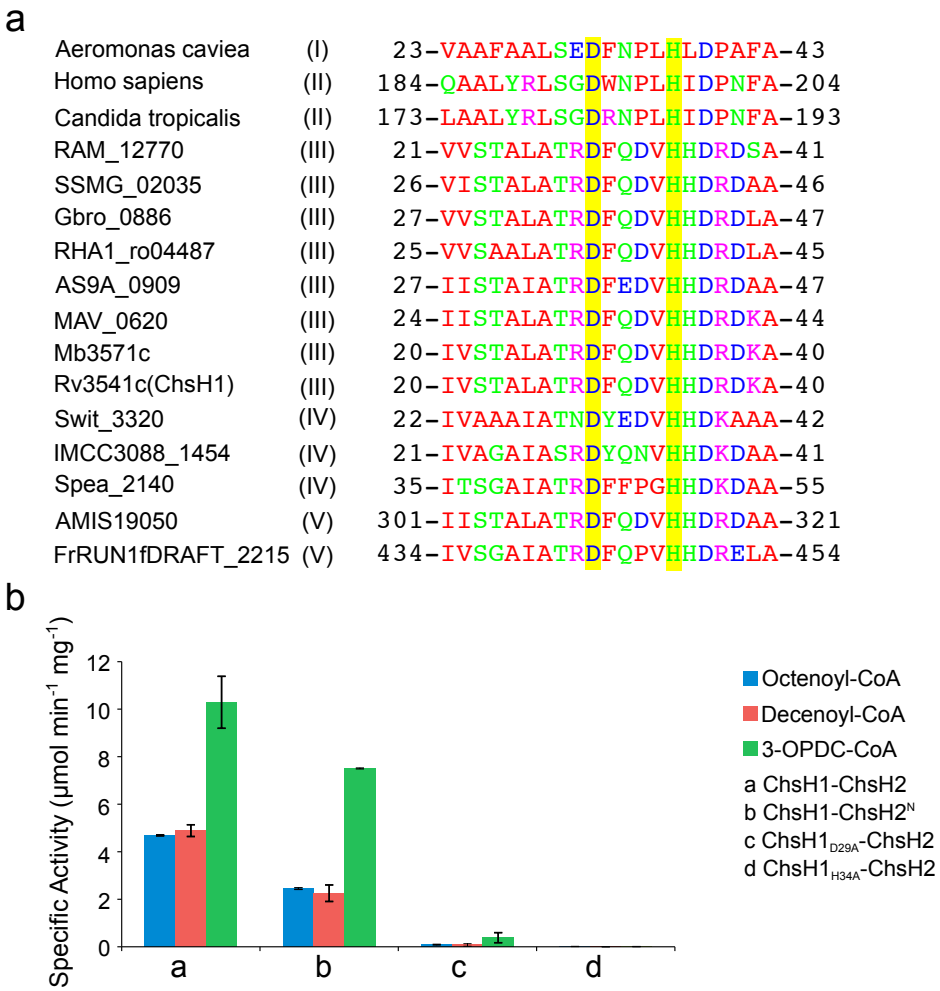


Figure 4-14. Protein sequence alignment of ChsH1 and its homologs across species and specific activity data for ChsH1-ChsH2, ChsH1-ChsH2^{N}, ChsH1$_{D29A}$-ChsH2 and ChsH1$_{H34A}$-ChsH2. (a) Active site housing segment sequence alignment of ChsH1 against other MaoC-like enoyl-CoA hydratases across species. All the organisms included in the alignment match the organisms in the phylogenetic tree in Figure 4-30. The operonic organizations of genes encoding MaoC-like enoyl-CoA hydratases were labeled following Figure 4-30. All the aligned residues are colored by similarity (ClustalW). The highly conserved putative active siteresidues are highlighted yellow. The sequence accession numbers from NCBI for each organism are *Aeromonas Caviea* (O32472.1), *Homo sapiens* (PDB: 1S9C), *Candida tropicalis* (PDB:1PN2), RAM_12770 (*Amycolatopsis mediterranei S699*; YP_005530501.1), SSMG_02035 (*Streptomyces sp. AA4*; ZP_07277995.1), Gbro_0886 (*Gordonia bronchialis DSM 43247*; YP_003272092.1), RHA1_ro04487 (*Rhodococcus jostii RHA1*; YP_704431.1), AS9A_0909 (*Amycolicicoccus subflavus DQS3-9A1*; YP_004492161.1), MAV_0620 (*Mycobacterium avium 104*; YP_879900.1), Mb3571c (*Mycobacterium bovis AF2122/97*; NP_857210.1), ChsH1 (*Mycobacterium tuberculosis H37Rv*; NP_218058.1), Swit_3320 (*Sphingomonas wittichii RW1*; YP_001263804.1), IMCC3088_1454 (*Gamma proteobacterium IMCC3088*; ZP_08270994.1), Spea_2140 (*Shewanella pealeana ATCC 700345*; YP_001501995.1), AMIS19050 (*Actinoplanes missouriensis 431*; YP_005461641.1), FrEUN1fDRAFT_2215 (*Frankia sp. EUN1f*;

ZP_06412519.1). **(b)** Specific activity data for ChsH1-ChsH2, ChsH1-ChsH2^{N}, ChsH1$_{D29A}$-ChsH2 and ChsH1$_{H34A}$-ChsH2 assayed with octenoyl-CoA, decenoyl-CoA and 3-OPC-CoA.

The distinctively short central α-helix (α4) is a source of another major difference (Figure 4-15), and has the most important functional implication. The short helix α4 and the following long loop generate a bigger space in the hot-dog fold compared to other MaoC-like hydratases (Figure 4-16), and the presence of this space is critical to allow the ChsH1-ChsH2^{N} heterodimer to bind the steroid substrate. Consistently, comparison of this structural feature in the structures of ChsH1-ChsH2^{N} and its homologs (Figure 4-16) reveals that as the central α-helix of the hot-dog fold becomes shorter, the binding cavity becomes wider and sufficiently spacious to accommodate increasingly larger and bulkier substrates. It appears that upon shifting from linear alkyl chain and branched alkyl chain substrates to the rigid rings of a steroid system, the protein architecture transitions from small monomer to extended monomer, and finally to a heterodimer (Figure 4-16).

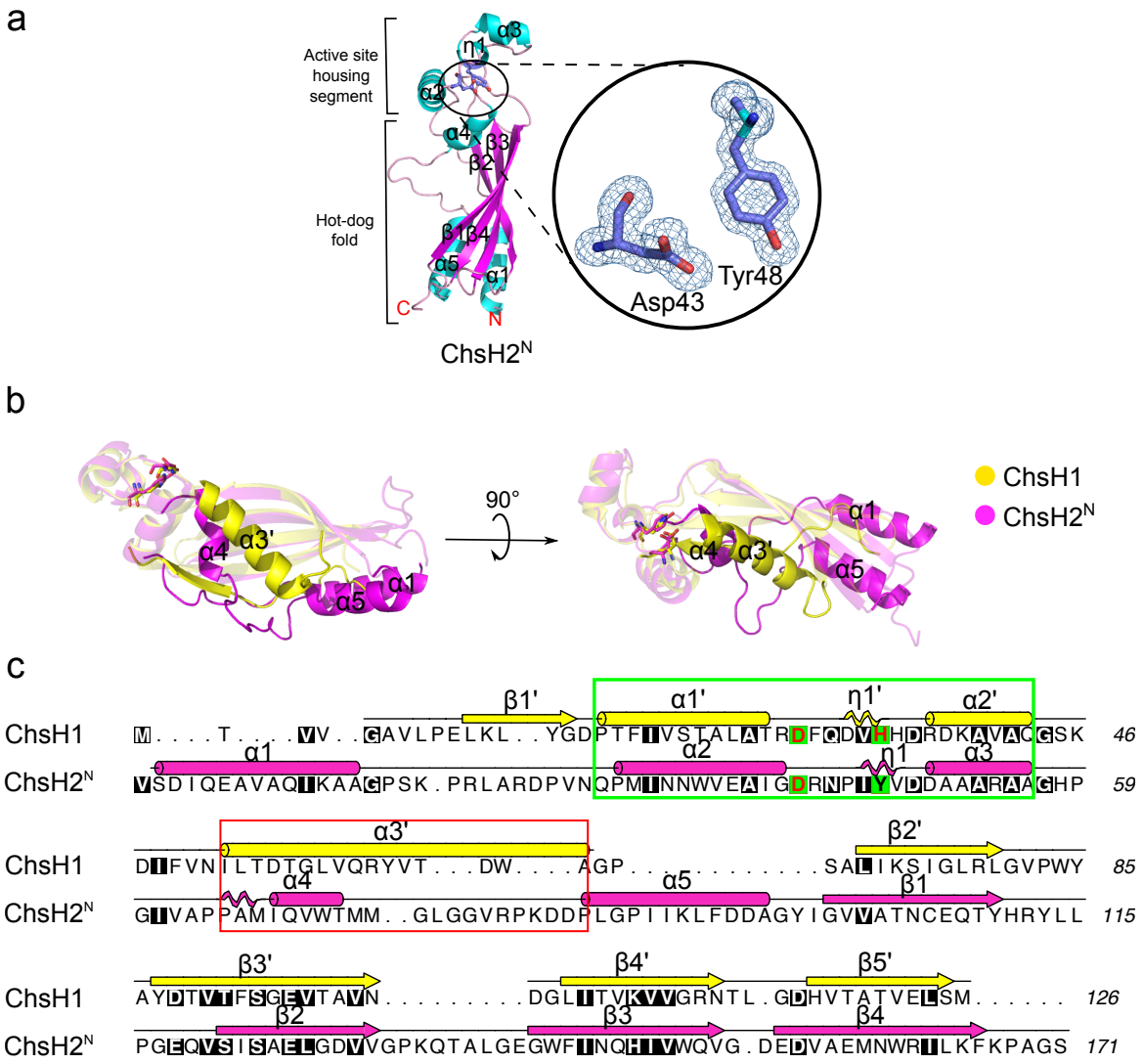


Figure 4-15. Structure of ChsH2^N and secondary structure sequence alignment between ChsH1 and ChsH2^N. (a) ChsH2^N adopts a nonstandard hot-dog fold. The cartoon representations are colored by secondary structure. N-termini and C-termini are labeled. In ChsH2^N, a four stranded b-sheet (b1-b4-b3-b2) wraps around a4, assembling into a nonstandard hot-dog fold; α2, η1 and α3 fold to form a potential active site housing segment. The potential active site residues are colored blue and the simulated annealing omit maps of those residues are calculated to reduce the effects of model bias. Electron density is shown as mesh (σ=3). (b) Cartoon representation of superimposed ChsH1 (yellow) and ChsH2^N (magenta). Structurally similar parts are rendered transparent and the variations are marked. The structure on the right is obtained by rotating the structure on the left 90° about the X-axis. Potential active site residues are shown as sticks. (c) Secondary structural sequence alignment between ChsH1 and ChsH2^N. Identical residues are highlighted in black. The secondary structure elements from ChsH1 (yellow) and from ChsH2^N (magenta) correspond to (a) and (b). Residues that constitute active site housing segments are highlighted in the green box. The central α-helices from ChsH1 and ChsH2^N are highlighted in the red box.

For example, in the structure of N-CtRH complexed with its physiologic product, *(3R)*-hydroxy-octanoyl-CoA, the ligand is accomodated in the interface between the nonstandard N-terminal hot-dog fold and the complete C-terminal hot-dog fold.[150] The pantothenate and adenosine triphosphate of the ligand are exposed to solvent, and the ten-carbon acyl chain points towards the shorter central α-helix in the N-terminal nonstandard hot-dog fold.[150] Interestingly, the ChsH1-ChsH2^{N} dimer has a similar, yet distinct, substrate binding mode; ChsH1 provides the active site and ChsH2^{N} provides the binding pocket (Figure 4-22b). The available space in the suggested binding tunnel of the hot-dog fold is restricted by the rigid central helix, so the length of the helix influences substrate preferences (Figure 4-16, Table 4-4). The existence of a nonstandard hot-dog fold with shorter central helices in the N-terminal domains of CtRH, HuRH and ChsH2^{N} chain suggests that one active site has evolved into a binding site to accommodate bulky substrates concomitant with the sacrifice of one set of active site catalytic residues (Figure 4-16, Table 4- 4).

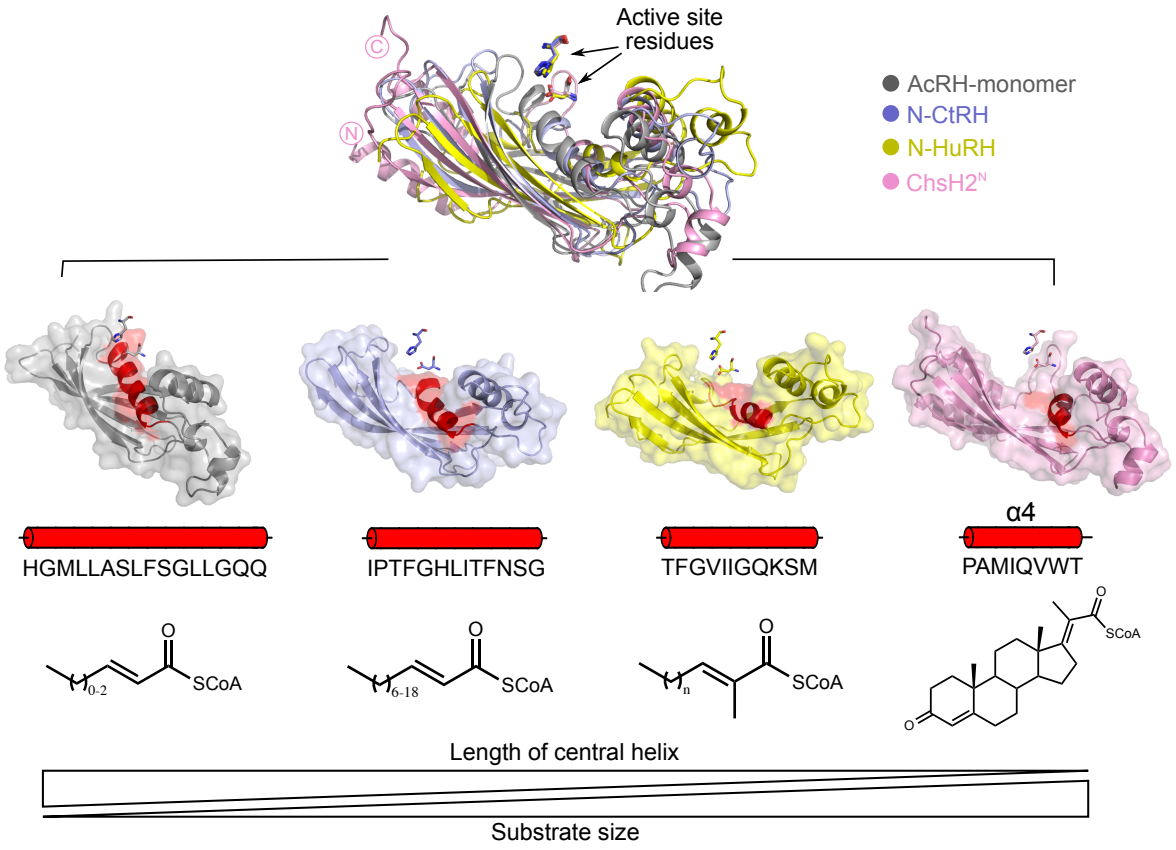


Figure 4-16. Comparisons of MaoC-like enoyl-CoA hydratases across species. 3D structural comparison between AcRH monomer (gray), N-terminal domain of CtRH (N-CtRH) (residue number: 1-135; purple), N-terminal domain of HuRH (N-HuRH) (residue number: 1-146; yellow), and ChsH2^N chain (pink). Active site residues from the other chain of AcRH, C-terminal domains of CtRH and HuRH or ChsH1, are shown as sticks. The central α–helices are colored red. Molecular surfaces are rendered transparent. The length of the central α-helices correspond to the 3-D structures above them. For simplicity, 3_{10} helices and α-helices are both shown as cylinders. Protein sequences shown underneath are the amino acids comprising the central helices. The sizes of the central helices dictate the substrate preferences. The preferred substrates correspond to the structure above them. AcRH (PDB code: 1IQ6); CtRH (PDB code: 1PN2); HuRH (PDB code: 1S9C).

The standard central helix of 16 amino acids in the complete hot-dog fold confers specificity for 4 to 6 carbon enoyl-CoAs.[154] In CtRH, the bent central helix of 13 amino acids in the N-terminal hot dog fold enables accomodation of long-chain enoyl-CoAs (C_{10}-C_{22}) in the binding site.[150] The shorter central α-helix of 11 amino acids and bigger pocket in the N-terminal hot-dog fold of HuRH extends the accepted enoyl-CoA CoA esters chain length up to C_{26} and methyl-

branched enoyl-CoAs. The central helix α4 of 8 amino acids and adjacent flexible loop in $ChsH2^N$ generates an even more flexible active site with a larger pocket. The enlarged pocket in ChsH1-$ChsH2^N$ makes it possible to accomodate steroid CoA thioesters as substrates, consistent with its function in cholesterol degradation in *Mtb*.[50] (Figure 4-16, Table 4- 4)

Small angle X-ray scattering revealed the structural information of full length ChsH1-ChsH2. A BLAST search revealed that $ChsH2^C$ belongs to the DUF35 superfamily, whose function remains to be determined. SSO2064 (PDB code: 3irb) is the first structural representative of the DUF35 superfamily. Protein sequence alignements between $ChsH2^C$ and SSO2064 suggested that they share very high similarities in terms of their sizes and amio acids conservations (Figure 4-17a). Small angle X-ray scattering was empolyed to unravel the structural information of full length ChsH1-ChsH2 in order to understand the relative orientation of $ChsH2^C$. The molecular shape from SAXS for ChsH1-$ChsH2^N$ agrees with the high resolution crystal structure. Meanwhile, the molecular shape obtained from SAXS for full length ChsH1-ChsH2 revealed an extra arm, which is consistant with the addition of $ChsH2^C$. We used the crystal structure of SSO2064 and modeled it onto the extra arm in the full length molecular envelope with high quality superimpostition (Figure 4-17).

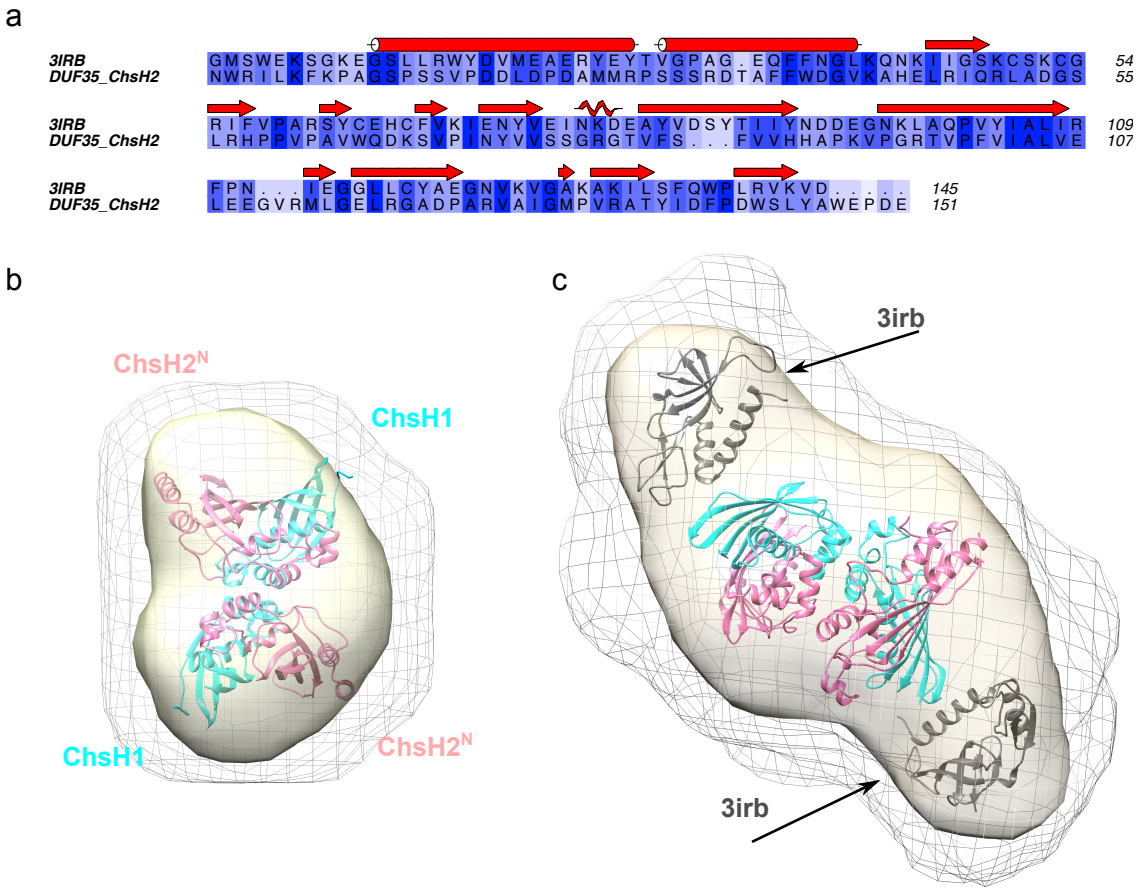


Figure 4-17. Small Angle X-ray Scattering (SAXS) generates envelopes for the solution state of both ChsH1-ChsH2^{N} and ChsH1-ChsH2. (a) Protein sequence alignment between ChsH2^{C} (the DUF35 domain from ChsH2) and a representative structure from DUF35 family (PDB code: 3irb); the secondary structure elements are imported based on the structure of 3irb. (b) (c) The molecular envelope of solution state ChsH1-ChsH2^{N} and ChsH1-ChsH2 was determined from the experimental scattering profile using the ASTAS package. The meshed envelope is the model generated from all the aligned models averaged by DAMAVER and is contoured to 30 Å, which is the most probable map. The final envelope is shown in yellow solid surface. It is generated by DAMFILT, which filtered the avergaed model at a preset cut-off volume (here is the averaged volumes of all the models). The final damfilt models were contoured to 30 Å, and then the high-resolution crystal structure was subsequently docked into the molecular envelope in Chimera with a correlation value of 0.8578 for (b). (c) The crystal structure of ChsH1-ChsH2^{N} was fitted into the envelope with a correlation value of 0.9041. Two 3irb monomers were fitted into remaining spaces highlighted by arrows with correlations values of 0.9829.

ChsH1-ChsH2 prefers to turnover steroid substrates while a truncated form of ChsH1-ChsH2^{N} is sufficient for catalytic activity. It has been demonstrated that ChsE1-ChsE2 catalyzes the dehydrogenation of 3-OPC-CoA (chapters II and III). We assayed 100 nM and 1

μM of ChsH1-ChsH2 protein complex with 3-OPDC-CoA. The assay mixture was subjected to MALDI-TOF mass spectrometry analysis after 15 min incubation at 25 °C. The hydrated product of 3-OPDC-CoA is observed (Figure 4-18).

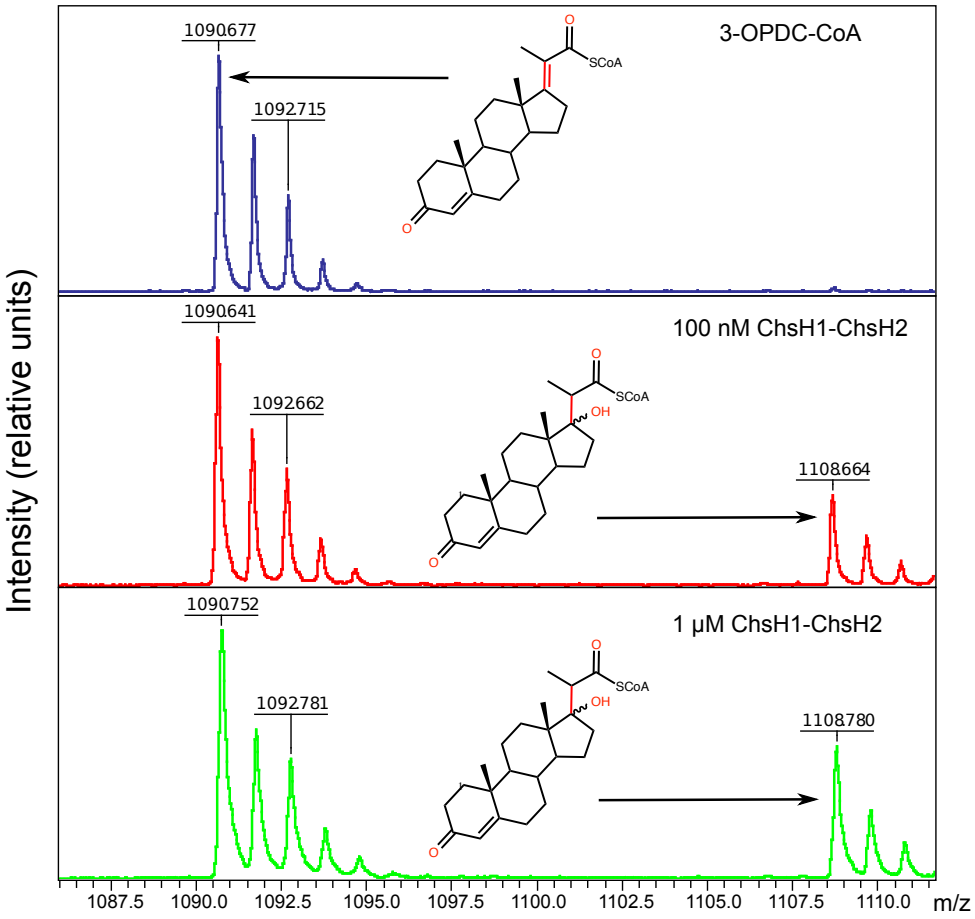


Figure 4-18. MALDI-TOF mass spectra confirmed ChsH1-ChsH2 catalyzed hydratation of 3-OPDC-CoA. 100 nM and 1 μM of ChsH1-ChsH2 protein complex were assayed with 3-OPDC-CoA). The assay mixture was subjected to MALDI-TOF mass spectrometry analysis after 15 min incubation at 25 °C. The hydrated product of 3-OPDC-CoA is observed.

We assayed the ChsH1-ChsH2 enzyme complex with three different substrates: octenoyl-CoA, decenoyl-CoA and 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA). The hydration process was monitored spectrophotometrically and the formation of the hydrated products was confirmed by MALDI-TOF mass spectrometry. Although ChsH1-ChsH2 catalyzed

the hydration of all three substrates, the highest turnover activity was observed with 3-OPDC-CoA (Figure 4-14b). We also obtained the kinetics parameters in order to explain the specific role that ChsH1-ChsH2 plays in cholesterol metabolism in *Mtb*. The results showed that ChsH1-ChsH2 possessed about 10 times higher activity towards 3-OPDC-CoA, one of the cholesterol metabolite intermediates, relative to fatty acids enoyl-CoAs (Figure 4-19 and Table 4-5). Therefore, ChsH1-ChsH2 has the ability to bind bulky substrates and preferentially catalyzes the hydration of steroid enoyl-CoA compared to aliphatic enoyl-CoAs, thus establishing a role in cholesterol metabolism for this enzyme.

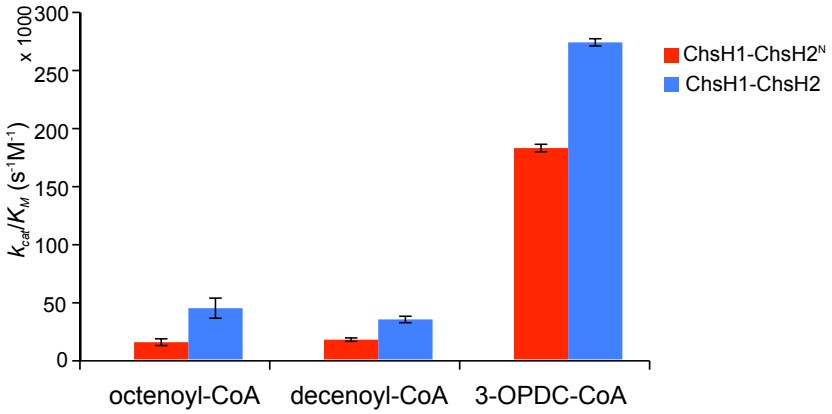


Figure 4-19. Catalytic efficiencies of ChsH1-ChsH2 and ChsH1-ChsH2^N towards octenoyl-CoA, decenoyl-CoA and 3-OPDC-CoA, respectively.

In order to test whether ChsH1-ChsH2^N is still catalytically active, we analyzed its catalytic activity with the substrates described above. Importantly, ChsH1-ChsH2^N retained around 70% of its catalytic activity, indicating that ChsH2^C is not required for catalysis of hydration (Figure 4-14b, Figure 4-19 and Table 4-5). Our hypothesis is that ChsH2^C helps bind CoA substrate and facilitate its utilization. Therefore, we think the existence of ChsH2^C will enhance ChsH1-ChsH2's enzyme activity.

Table 4-5. Enzyme specificities of ChsH1-ChsH2 and ChsH1-ChsH2^N, respectively.

ChsH1-ChsH2			
Substrate	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1}\mu M^{-1}$)
Octanoyl-CoA	28.6 ± 5.30	12.6 ± 0.803	(4.41 ± 0.864) ×10^5
Decenoyl-CoA	27.7 ± 2.13	9.53 ± 0.255	(3.44 ± 0.280) ×10^5
3-OPDC-CoA	9.88 ± 1.07	27.0 ± 1.07	(2.73 ± 0.314) ×10^6
ChsH1-ChsH2^N			
Octanoyl-CoA	21.6 ± 3.94	3.20 ± 0.223	(1.48 ± 0.289) ×10^5
Decenoyl-CoA	20.7 ± 1.73	3.51 ± 0.089	(1.70 ± 0.148) ×10^5
3-OPDC-CoA	9.38 ± 1.58	17.1 ± 0.990	(1.82 ± 0.324) ×10^6

The ChsH1-ChsH2^N heterotetramer forms a complex with 3-oxo-4-pregnene-20-carboxyl CoA. In order to investigate the structural basis of preferential hydration of a steroid enoyl-CoA instead of aliphatic enoyl-CoAs, we co-crystallized ChsH1-ChsH2^N with a substrate analog, 3-oxo-4-pregnene-20-carboxyl CoA (3-OPC-CoA) (Figure 4-20). The crystals diffracted to a resolution of 1.76 Å (Table 4-2) and molecular replacement was used to solve the ChsH1-ChsH2^N:3-OPC-CoA structure using the apo-protein structure as the template.

The atomic picture of the ChsH1-ChsH2^N:3-OPC-CoA complex established that one binding site exists per ChsH1-ChsH2^N heterodimer. In each heterodimer, the binding pocket is located at the ChsH1-ChsH2^N interface and intrudes into ChsH2^N (Figure 4-20). Binding of 3-OPC-CoA results in a large quaternary structural change compared to the apo-ChsH1-ChsH2^N complex (Figure 4-21a and 4-21b). The major changes occur in the ChsH2^N chain. Movements of α1, α5, loop I, and II (6 to 7 Å) and the smaller movements of β1–β4–β3–β2 generate enough space to accommodate the 3-OPC-CoA ligand in the tunnel (Figure 4-21a and 4-21b). We posit that the

heterodimer interface serves to enable the large structural change required for binding the polycyclic steroid system which could not occur in a monomeric binding site like that seen in CtRH or HuRH.

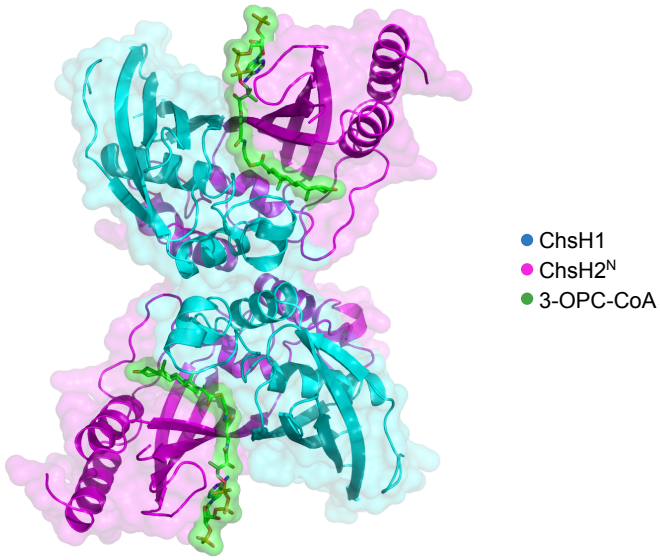


Figure 4-20. ChsH1-ChsH2^N complexed with 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA). Overall atomic picture of ChsH1-ChsH2^N: 3-OPC-CoA. The holo-heterotetramer is in the same orientation as Fig. 4-9a. A single heterotetramer is in one asymmetric unit (ASU). ChsH1 chains are colored blue, and ChsH2^N chains are colored magenta. Ligand 3-OPC-CoA is rendered as sticks and colored by atom. Molecular surfaces are rendered transparent.

The ligand is in a boomerang conformation, which is similar to the orientation of 3-hydroxy-octanoyl-CoA in CtRH.[150] The "boomerang" sits between the ChsH1 chain and the ChsH2^N chain with its elbow, cysteamine, positioned across β1 and β2' (Figure 4-22). The whole boomerang bends toward ChsH2^N with the coenzyme A moiety exposed to solvent and the four-ring steroid system buried inside ChsH2^N (Figure 4-21). From the ChsH1-ChsH2^N:3-OPC-COA structure, the C17-C20 bond, which is the site of hydration, sits in apposition to the active site residues Asp29/His34 from ChsH1 (Figure 4-22).

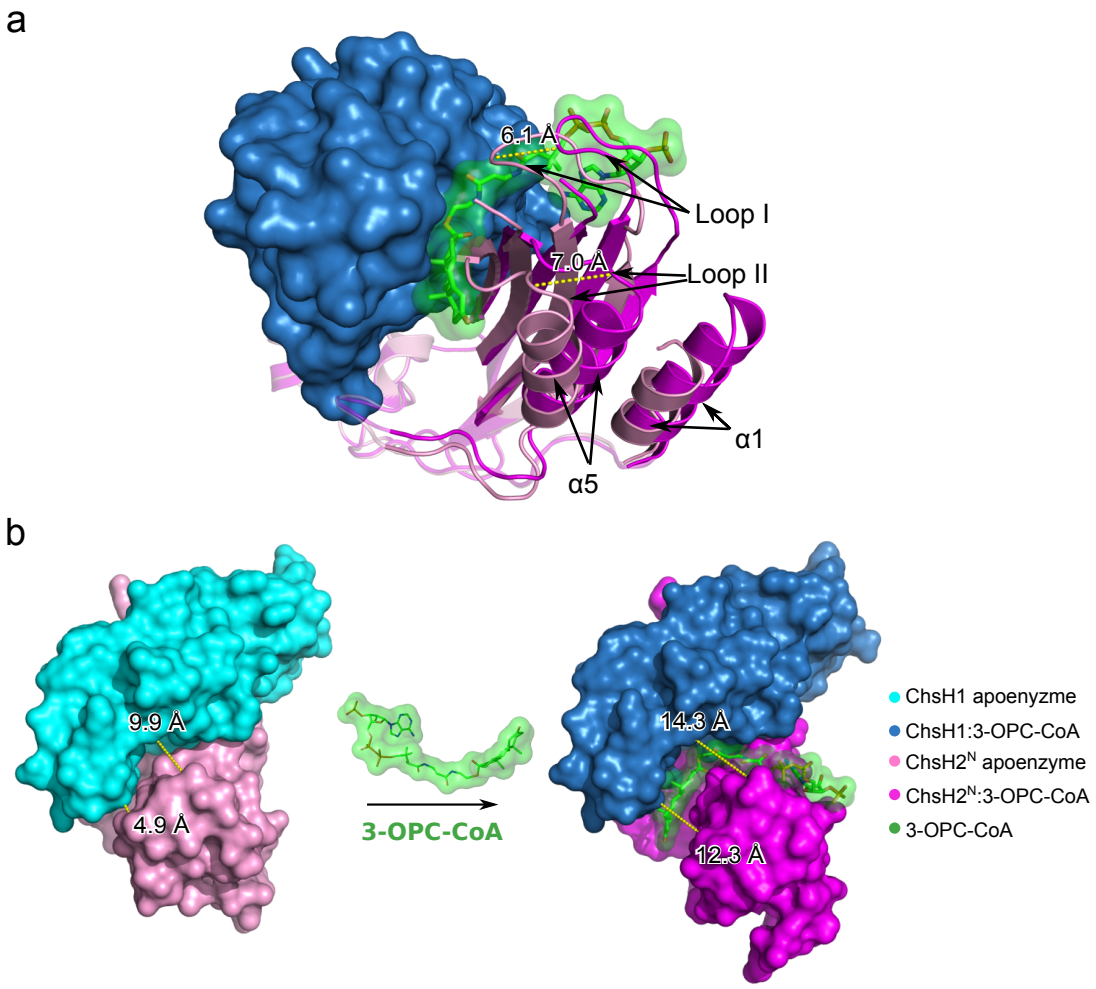


Figure 4-21. Comparison between apo-ChsH1-ChsH2^N and ChsH1-ChsH2^N:3-OPC-CoA. (a) Superimposed heterodimer apo-ChsH1-ChsH2^N structure and ChsH1-ChsH2^N:3-OPC-CoA structure. Superimposed ChsH1 chains are colored in blue and their surface is shown. The ChsH2^N chains from Apo-ChsH1-ChsH2^N and ChsH1-ChsH2^N:3-OPC-CoA are shown in pink and magenta, respectively. The variations occur in the segments of the ChsH2^N chains which are labeled. (b) The difference in the opening of the binding pocket before and after 3-OPC-CoA binding.

Binding interactions with ChsH2^N anchor the coenzyme A moiety to the protein. The amine group (N1") is hydrogen bonded to Asn106 in β1, N2'' is hydrogen bonded to Arg163 in β5, and O3'' is stabilized by Ala137 from loop II. The elbow of the boomerang is stabilized through a hydrogen bond between N4'' nitrogen and Gly81 from β2' in ChsH1. The steroid part of the "boomerang" has fewer specific interactions with the enzyme compared to the coenzyme A

moiety, consistent with the steroid's hydrophobicity. The thioester O5" is stabilized through a hydrogen bonding interaction with Ala104-ChsH2^{N}. No polar interactions were found that stabilize the O3-C3 ketone (Figure 4-22).

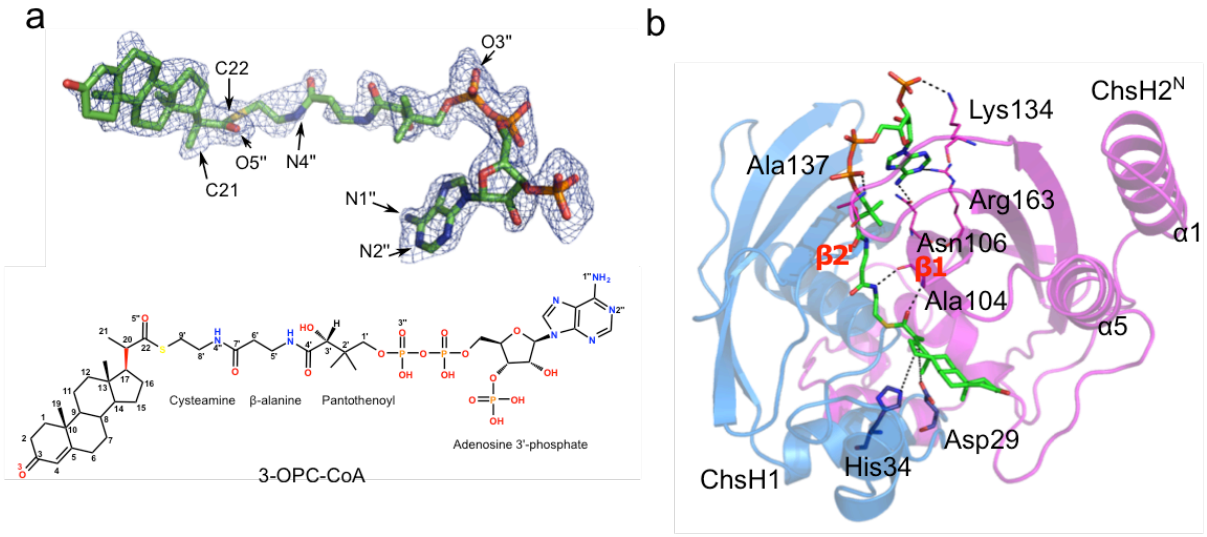


Figure 4-22. 3-OPC-CoA binding interactions in the ChsH1-ChsH2^{N} heterodimer. (a) A simulated annealing *Fo-Fc* omit map of 3-OPC-CoA contoured at 2.5 σ was calculated to reduce the effects of model bias. A chemical structure of 3-OPC-CoA is shown and colored by atom. (b) Ligand-binding interactions in the ChsH1-ChsH2^{N} heterodimer. The hydrogen bonds are shown as black dashes. Residues that interact with 3-OPC-CoA are labeled. 3-OPC-CoA fits between β1 from ChsH2^{N} and β2'from ChsH1.

The effects of metals identified in ChsH1-ChsH2^{N} and ChsH1-ChsH2N:3-OPC-CoA structures. Cadmium sites were identified in both the apo-ChsH1-ChsH2^{N} and ChsH1-ChsH2^{N}:3-OPC-CoA structures. In the apo-ChsH1-ChsH2^{N} structure, cadmium ions were located around the active site as well as at the ASU packing interface. In all of the ligand bound structures, cadmium ions were identified between the active site and the α,β carbons of enoyl/acyl-CoAs that block the interaction of the active site (Asp29, His34) with the substrate analog. The structure of ChsH1-ChsH2^{N}:3-OPC-CoA has been used as an example to investigate the effects of metals (Figure 4-23b). The distance between the active site and the α, β carbons of

the substrate analog (8-9 Å) suggests that the cadmium ions bind where the water of hydration is expected to be during catalysis. To confirm whether the metals are introduced by the crystallization process or are required by the enzyme for catalytic activity, ChsH1-ChsH2 was assayed with octenoyl-CoA in the presence of 20 mM EDTA. There was no obvious change in catalytic activity upon addition of EDTA (Figure 4-23a). However, upon addition of 20 mM or even 2.5 mM cadmium, ChsH1-ChsH2 lost almost all catalytic activity with octenoyl-CoA. The catalytic activity was recovered upon addition of EDTA and cadmium ions simultaneously (Figure 4-23a), thereby demonstrating that the metal ions from the crystallization inhibit the activity of ChsH1-ChsH2 by blocking the hydration of the substrate. In addition, calcium and cobalt ions from the crystallization process inhibit ChsH1-ChsH2 catalytic activity to some extent at 20 mM. However, no inhibition was observed at 2.5 mM.

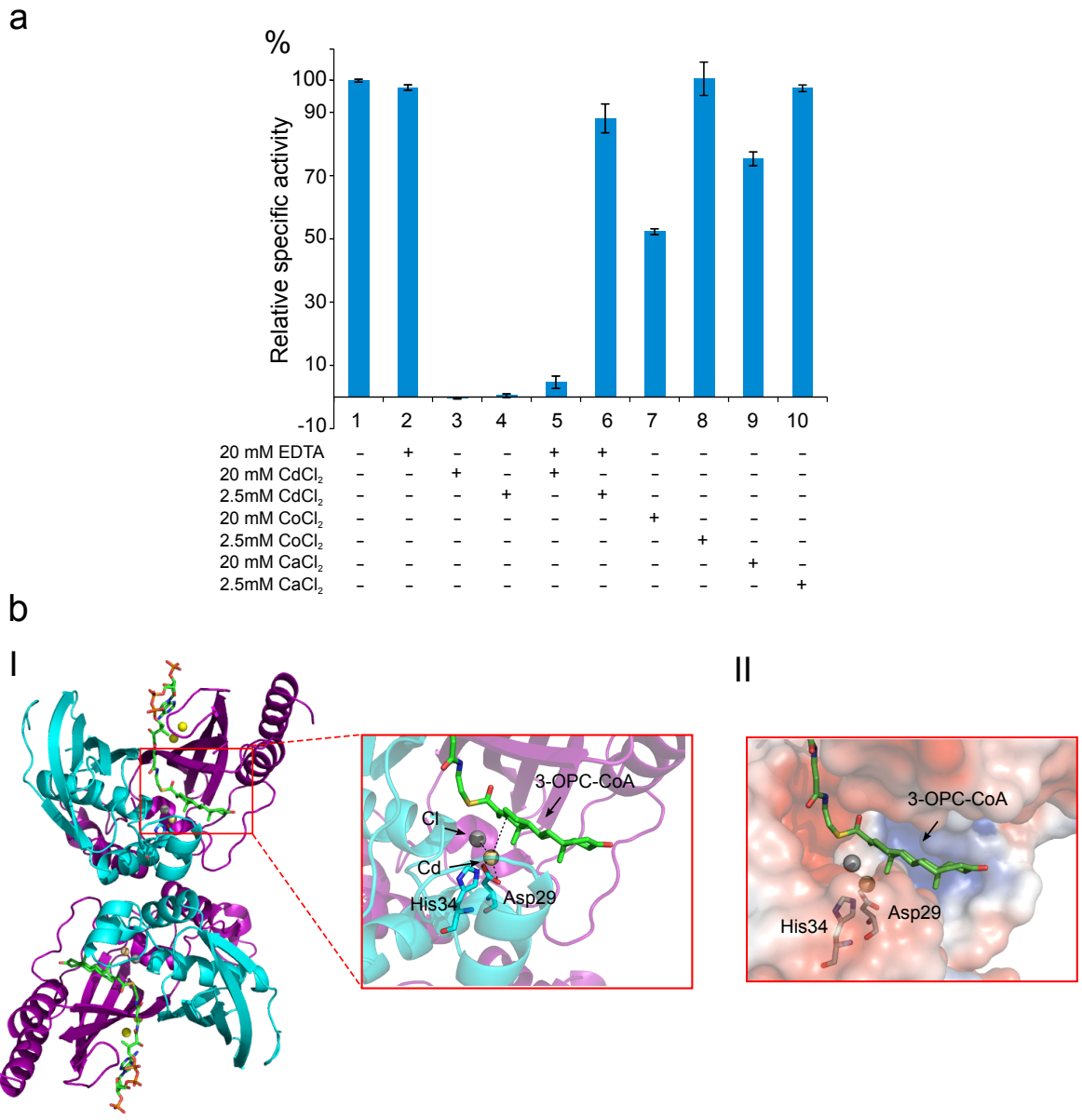


Figure 4-23. The effects of metals used for crystallization on catalytic activity. (a) The effects of metal ions on the activity of ChsH1-ChsH2. (b) I, 6 Cd^{2+} ions and 2 Cl^{-1} are identified in one of the ChsH1-ChsH2^{N} heterotetramers. One Cd^{2+} is located in the middle of the active site near the C17-C20 position of 3-OPC-CoA. Cl^{-1} coordinates with Cd^{2+}. II, electrostatic map shows an obvious negative pocket in which Cd^{2+} is located.

The ChsH1-ChsH2^{N} heterotetramer forms a complex with octanoyl-CoA. We co-crystallized ChsH1-ChsH2^{N} with another substrate analog, octanoyl-CoA. The crystals diffracted to a resolution of 1.31 Å (Table 4-3) and molecular replacement was used to solve the ChsH1-ChsH2^{N}:octanoyl-CoA structure using the apo-protein structure as the template.

Similar to ChsH1-ChsH2^{N}:3-OPC-CoA, there is one heterotetramer in one ASU (Figure 4-24a). Two octanoyl-CoA molecules are observed in one tetramer and they also bind at the interface of ChsH1 and ChsH2^{N}. After binding octanoyl-CoA, ChsH1-ChsH2^{N} does not display an obvious conformational change. However, the binding pocket at the interface of ChsH1 and ChsH2^{N} becomes smaller relative to the apo-ChsH1-ChsH2^{N} structure (Figure 4-24a, Figure 4-24b and Figure 4-24c). In addition, the hydrophobic eight carbon tail from octanoyl-CoA bound deeply into ChsH2^{N} and is tightly sealed by ChsH1-ChsH2^{N} heterodimer, which is different compared to the widely open binding channel in the ChsH1-ChsH2^{N}:3-OPC-CoA structure (Figure 4-24c). The structure of ChsH1-ChsH2^{N}:octanoyl-CoA was superimposed onto the structure of ChsH1-ChsH2^{N}:3-OPC-CoA. It showed that the movement of the corresponding substrate binding pocket resulted from the movement of the ChsH2^{N} chain (Figure 4-24c). No obvious conformational changes were observed for the ChsH1 chains (Figure 4-24c).

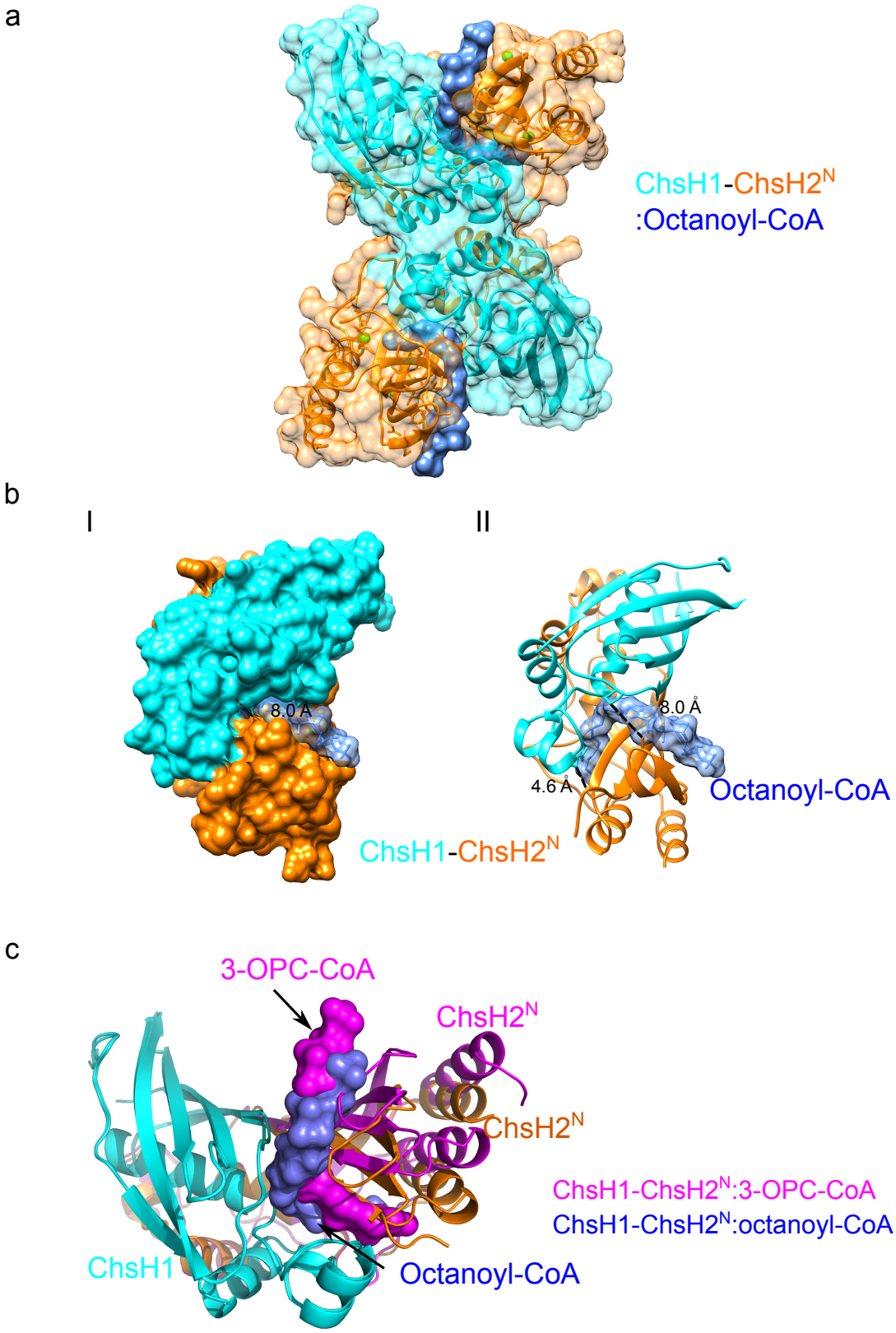

a
ChsH1-ChsH2N
:Octanoyl-CoA
b
I
II
8.0 Å
8.0 Å
4.6 Å
Octanoyl-CoA
ChsH1-ChsH2N
c
3-OPC-CoA
ChsH2N
ChsH2N
ChsH1-ChsH2N:3-OPC-CoA
ChsH1-ChsH2N:octanoyl-CoA
ChsH1
Octanoyl-CoA

Figure 4-24. ChsH1-ChsH2^N complexed with octanoyl-CoA. (**a**) Overall atomic picture of ChsH1-ChsH2^N: octanoyl-CoA. The holo-heterotetramer is in the same orientation as Fig. 4-9a. A single heterotetramer is in one asymmetric unit (ASU). ChsH1 chains are colored cyan and ChsH2^N chains are colored orange. Ligand octanoyl-CoA is shown in surface representation with no transparency and colored in blue. Two ligands bind simultaneously in one tetramer. (b) The difference in the opening of the binding pocket after octanoyl-CoA binding. Heterodimer of ChsH1-ChsH2^N:octanoyl-CoA. The distance between the same residues was measured across the the pocket opening relative to the structure of ChsH1-ChsH2^N:3-OPC-CoA. The distance between ChsH2^N-Ala137 and ChsH1- Pro83 and the distance between ChsH2^N-Pro168 and ChsH1-Gln43 are 8.0 Å and 4.6 Å, respectively, with octanoyl-CoA bound. (c) The structure of ChsH1-ChsH2^N:octanoyl-CoA is superimposed onto ChsH1-ChsH2^N:3-OPC-CoA. ChsH1 chains from there two structure are colored in cyan. ChsH2 chains from ChsH1-ChsH2^N:octanoyl-CoA and ChsH1-ChsH2^N:3-OPC-CoA were colored in orange and magenta, respectively.

The coenzyme A moiety is similarly exposed to the solvent (Figure 4-25b). The hydrogen bonding interactions between the coenzyme A moiety and the protein are the same relative to the ChsH1-ChsH2^N:3-OPC-CoA structure except that two residues, Leu80 and Gly81, contribute two extra hydrogen bonding interactions (Figure 4-25b), which is consistent with the tighter binding pocket (Figure 4-24c). No polar interactions were observed that interacted with the equivalent moiety of 3-OPC, the fatty acid tail (Figure 4-25b and 4-25c). A simulated annealing *Fo-Fc* omit map of octanoyl-CoA and active site residues (ChsH1-His34, ChsH1-Asp29) was generated using Phenix to reduce the effect of model bias (Figure 4-25a). H34 and D29 were exposed to the acyl-CoA α, β position, where the reaction is supposed to happen. Interestingly, a water molecule is located between the acyl-CoA α,β position and active site. This water molecule may be used for hydration during catalysis (Figure 4-25a). However, a metal ion (Cd^{2+}) was identified in the middle of the active site and acyl-CoA α,β position, thus blocked the catalysis reaction.

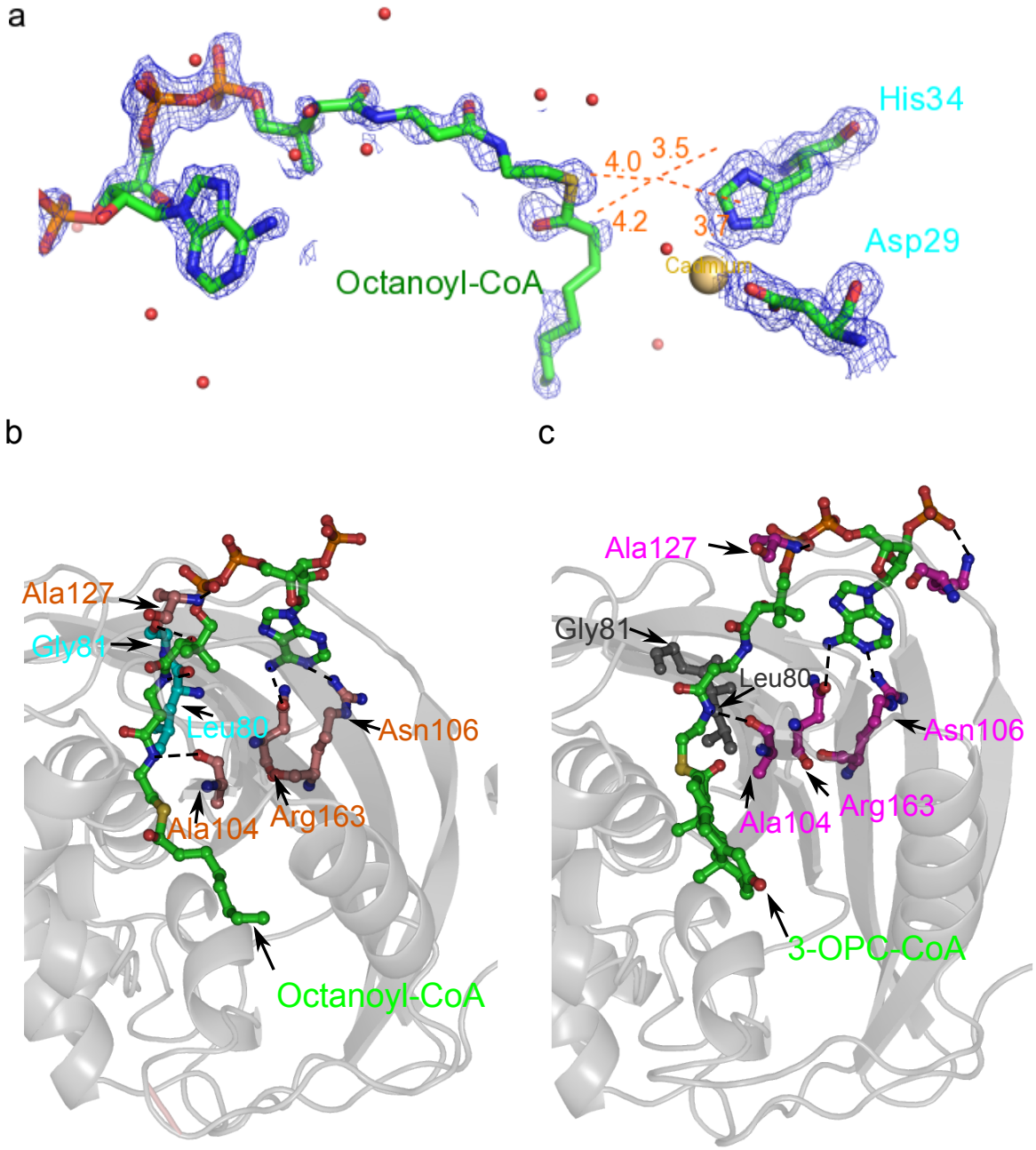


Figure 4-25. Omit map of octanoyl-CoA and its binding interactions with ChsH1-ChsH2^{N}. (a) Simulated annealing *Fo-Fc* omit map of octanoyl-CoA and active residues His34, Asp29 contoured at 1 σ was calculated to reduce the effects of model bias. The distances of Asp29, His34 with C17-C20 double bond are shown. Water molecules that are in hydrogen bonding distances (less than 3.5 Å) are shown. The cadmium ion introduced from crystallization condition is shown in yellow sphere. (b) (c) Residues from both ChsH1 and ChsH2^{N} that interact with octanoyl-CoA or 3-OPC-CoA through hydrogen bonding are shown in stick and labeled.

Polar interaction involved residues from ChsH2 and ChsH1 were labeled in orange and cyan, respectively. ChsH1 and ChsH2^N were shown in cartoon and rendered in transparent.

The ChsH1-ChsH2^N heterotetramer forms a complex with 3-oxo-pregna-4,17-diene-20-carboxyl-CoA. The crystal structure of ChsH1-ChsH2^N co-crystalized with its substrate, 3-OPDC-CoA, was also obtained. The crystals diffracted to a resolution of 2.22 Å (Table 4-3) and molecular replacement was used to solve the ChsH1-ChsH2^N:3-OPDC-CoA structure using the apo-protein structure as the template.

The structure of ChsH1-ChsH2^N: 3-OPDC-CoA is similar to ChsH1-ChsH2^N:3-OPC-CoA. They share the same space group (Table 4-2 and Table 4-3) and they share very similar structures. Briefly, one ASU contains one $\alpha_2\beta_2$ heterotetramer with two ChsH1 chains and two ChsH2^N chains and two 3-OPDC-CoA molecules were identified and modeled into the two equivalent binding sites per tetramer (Figure 4-26). Even though the binding pocket after 3-OPDC-CoA binding is much more open relative to octanoyl-CoA and decenoyl-CoA, it is less open relative to the structure with the saturated analog, 3-OPC-CoA bound (Figure 4-21b and Figure 4-26b). The difference suggests that some of the 3-OPDC-CoA may be hydrated due to catalysis. In order to investigate whether the hydration reaction happens upon 3-OPDC-CoA binding, polar interactions with 3-OPDC-CoA were analyzed and compared with the 3-OPC-CoA structure. Unfortunately, all the involved residues are identical and the distances between the active site and the C17-C20 bond is still too long to conclude that a catalytical active complex is formed (Figure 4-27). This may be due to the existence of metal ions (Cd^{2+}). Therefore, in order to investigate the catalytic structure of ChsH1-ChsH2 towards 3-OPDC-CoA, new crystallization conditions need to be screened to exclude the artificial influences of metal ions.

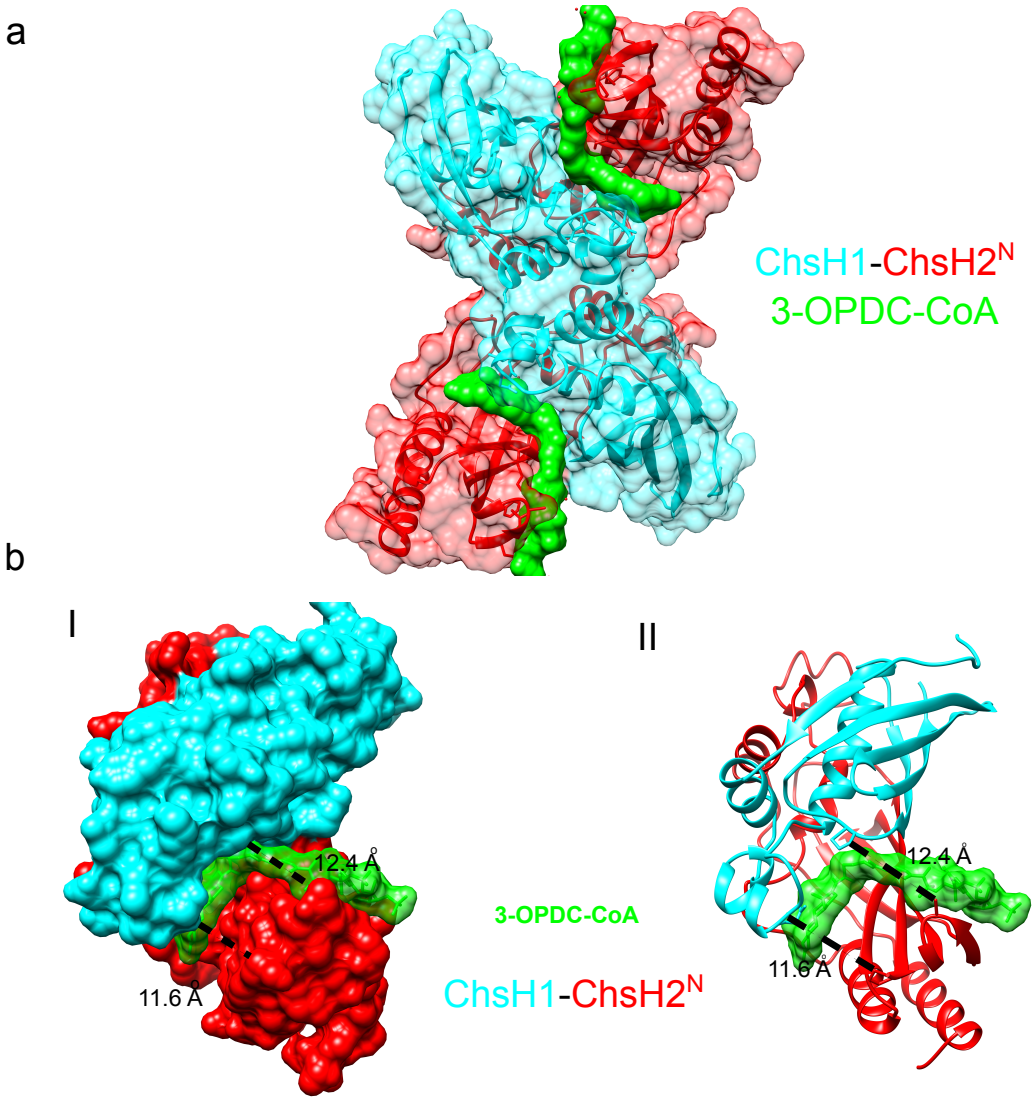


Figure 4-26. ChsH1-ChsH2^N complexed with 3-OPDC-CoA. (a) Overall atomic picture of ChsH1-ChsH2^N: 3-OPDC-CoA. A single heterotetramer is in one asymmetric unit (ASU). ChsH1 chains are colored cyan and ChsH2^N chains are colored red. Ligand 3-OPDC-CoA is shown in surface representation with no transparency and colored in green. Two ligands bind simultaneously in one tetramer. (b) Heterodimer of ChsH1-ChsH2^N: 3-OPDC-CoA. The distance between ChsH2^N-Ala137and ChsH1-Pro83 and the distance between ChsH2^N-Pro168 and ChsH1-Gln43 are 12.4 Å and 11.6 Å, respectively.

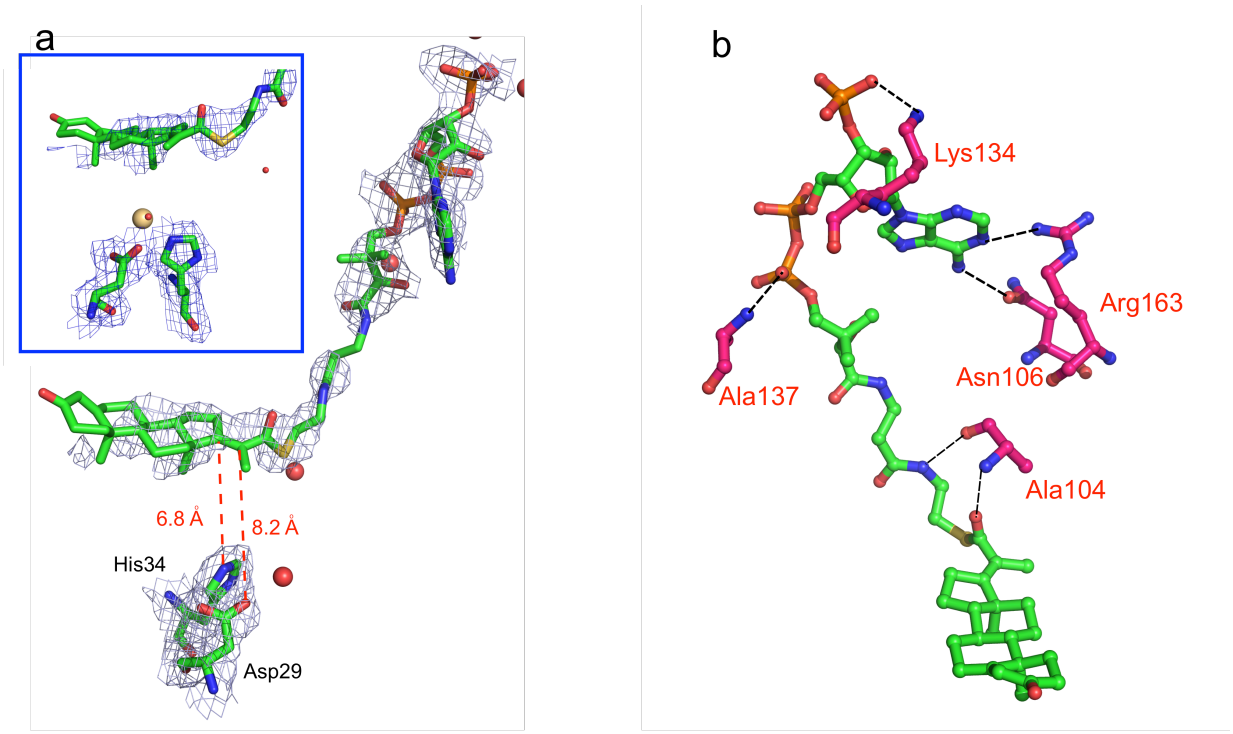


Figure 4-27. Omit map of 3-OPDC-CoA and its binding interactions with ChsH1-ChsH2^N. (a) Simulated annealing *Fo-Fc* omit map of 3-OPDC-CoA and active residues His34, Asp29 contoured at 1 σ was calculated to reduce the effects of model bias. The distances of Asp29, His34 with C17-C20 double bond are shown. Water molecules that are in hydrogen bonding distances (less than 3.5 Å) are shown. The Cd^{2+} located in the middle of active site and substrate is shown in yellow sphere. A chemical structure of 3-OPDC-CoA is shown. (b) Residues that are interacting with 3-OPDC-CoA through hydrogen bonding are shown in stick and labeled.

The ChsH1-ChsH2^N heterotetramer forms a complex with decenoyl-CoA. The co-crystallized structure of ChsH1-ChsH2^N with another substrate, decenoyl-CoA, has also been obtained from the same crystallization conditions. The crystals diffracted to a resolution of 1.67 Å (Table 4-3) and molecular replacement was used to solve the ChsH1-ChsH2^N:decenoyl-CoA structure using the apo-protein structure as the template.

Different to ChsH1-ChsH2^N:3OPC-CoA and ChsH1-ChsH2^N:octanoyl-CoA, there are two heterotetramers in one ASU in the ChsH1-ChsH2^N:decenoyl-CoA structure as in the apo-ChsH1-ChsH2^N structure. Interestingly, there is only decenoyl-CoA identified in one tetramer. This

binding mode suggested that the two binding site in one heterotetramer are equivalent and do not necessarily take substrate at the same time. The binding pocket is opened slightly wider compared to ChsH1-ChsH2^{N}:octanoyl-CoA, which is consistent with decenoyl-CoA being slightly larger than octanoyl-CoA (Figure 4-28). The CoA moiety of decenoyl-CoA is in the same binding mode as octanoyl-CoA (Figure 4-29b). An omit map was made to show the density of decenoyl-CoA, the active site was modeled in close proximity with the α, β double bond position in decenoyl-CoA, suggesting the reaction may occur in the crystal. However, anomalous peaks were observed in the middle. The omit maps of anomalous peaks were calculated and contoured to 7 σ. The strong signal from these anomalous peaks mixed with the density of the hydrophobic chain of decenoyl-CoA and made it impossible to model the decenoic acid correctly. Three anomalous peaks were observed here, which is different to the structures of ChsH1-ChsH2^{N}:3-OPC-CoA and ChsH1-ChsH2:octanoyl-CoA, we think that the occupancy of each anomalous peak is less than one and the binding position of the metal is not fixed (Figure 4-29). Again, new crystallization screens needs to be performed in order to exclude the artificial effects from metals.

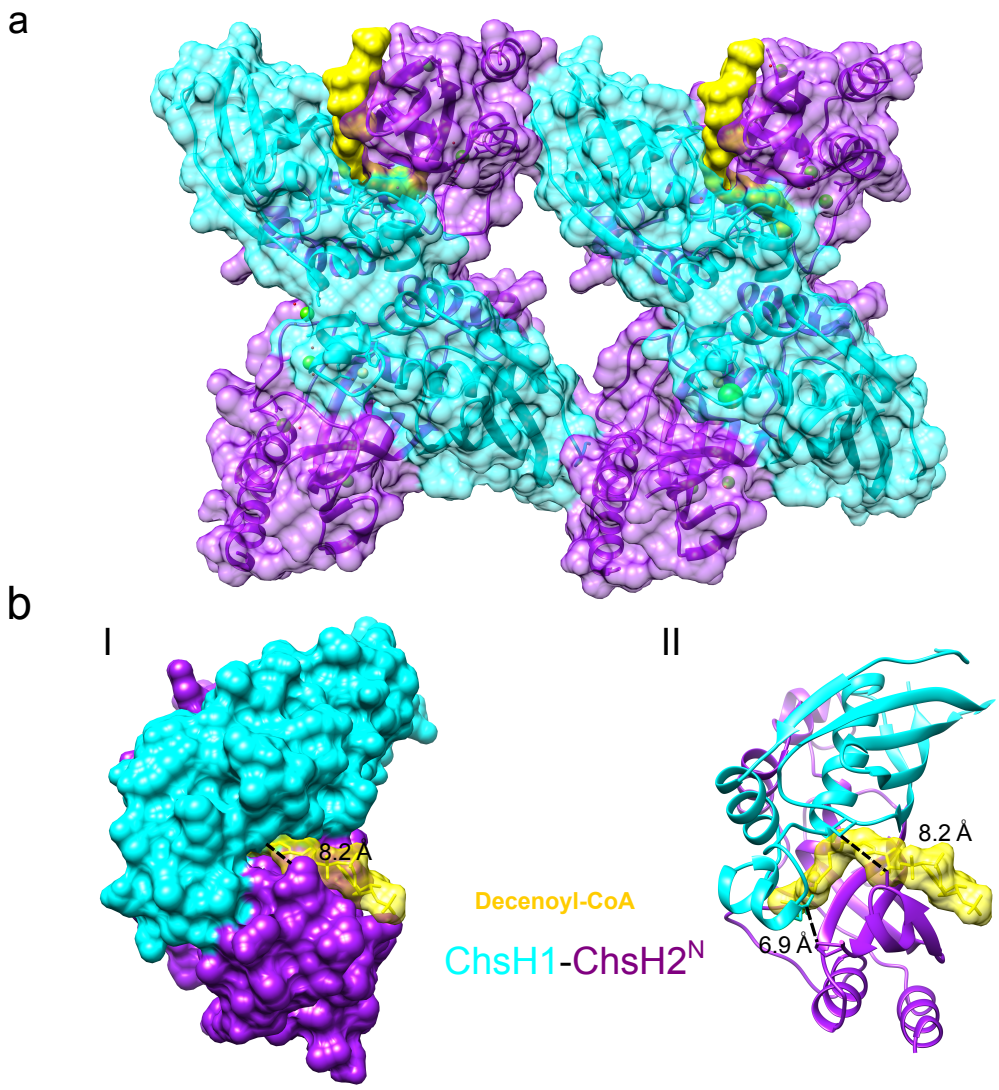
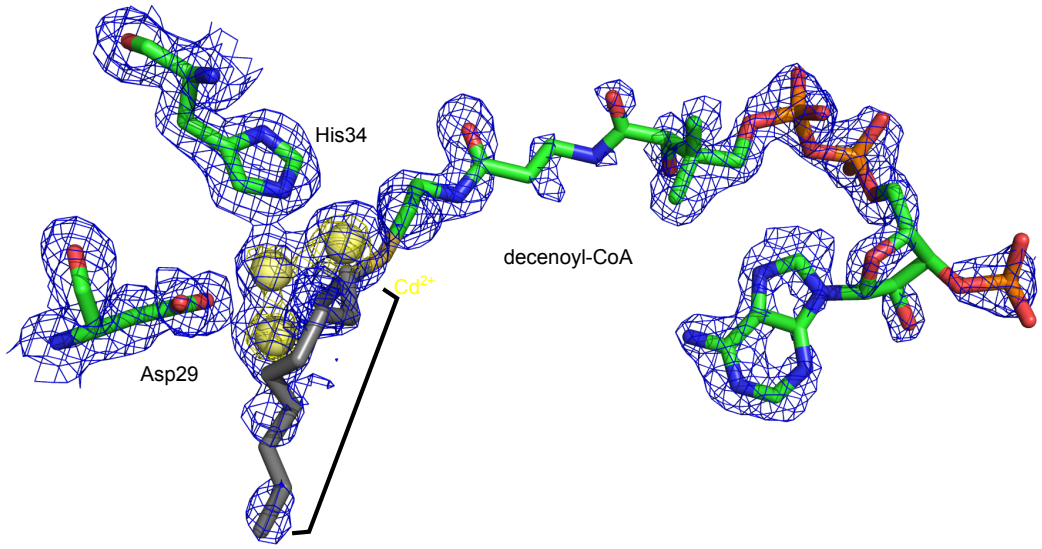


Figure 4-28. ChsH1-ChsH2^N complexed with decenoyl-CoA. (a) Overall atomic picture of ChsH1-ChsH2^N: decenoyl-CoA. Two heterotetramers are in one asymmetric unit (ASU), which is similar to apo-ChsH1-ChsH2^N. ChsH1 chains are colored cyan and ChsH2^N chains are colored purple. Ligand decenoyl-CoA is shown in surface representation with no transparency and colored in yellow. Only one ligand is observed bound in one tetramer. (b) Heterodimer of ChsH1-ChsH2^N:decenoyl-CoA. The distance between ChsH2^N-Ala137and ChsH1- Pro83 and the distance between ChsH2^N-Pro168 and ChsH1-Gln43are 8.2 Å and 6.9 Å, respectively.

a


His34
decenoyl-CoA
Cd2+
Asp29

b

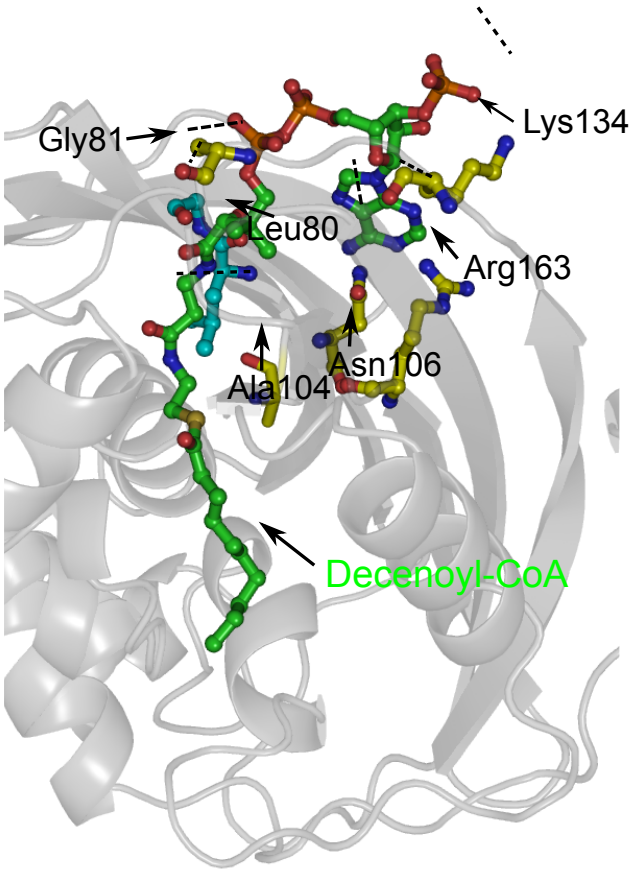

Gly81
Lys134
Leu80
Arg163
Asn106
Ala104
Decenoyl-CoA

Figure 4-29. Omit map of decenoyl-CoA and the active site in ChsH1-ChsH2^N. (a) Simulated annealing *Fo-Fc* omit map of 3-OPDC-CoA and active residues His34, Asp29 contoured at 1 σ was calculated to reduce the effects of model bias. The omit map of anomalous signal for Cd^{2+} was calculated, contoured at 7 σ and shown in yellow. The Cd^{2+} is shown in yellow sphere. Incorrectly modeled atoms are colored in grey. (b) Residues from both ChsH1 and ChsH2^N that interact with decenoyl-CoA through hydrogen bonding are shown in stick and labeled. Polar interaction involved residues from ChsH2 and ChsH1 were labeled in yellow and cyan, respectively. ChsH1 and ChsH2^N were shown in cartoon and rendered in transparent.

Phylogenetic relationships of ChsH1-ChsH2. Our characterization of the unusual ChsH1-ChsH2 architecture prompted us to assess whether the heteromeric assembly might also exist in other organisms and if so, whether those organisms were known to metabolize cholesterol. The presence of a hot-dog motif is insufficient to identify MaoC-like hydratase candidates because it is present in other enzyme families.[149] However, sequences in the active site housing segment are highly conserved and in combination with the hot-dog fold clearly define MaoC-like hydratases (Figure 4-14a). By comparing the MaoC-like hydratase sequences from human, fungi, and bacteria, we identified a generalized MaoC-like hydratase motif [VIL]-[AVI]-[SA]-X-[AY]-[ILRA]-[AL]-[ST]-[RGEN]-D-[FYWR]-[NQEF]-[PDN]-[VLG]-H-[ILH]-[PRK]-[ANDE]-X-A, which is located in the active site housing segments (Figure 4-13b and Figure 4-14a).

We used this motif to identify additional family members through individual BLAST queries with the protein sequences of ChsH1 and ChsH2 against the non-redundant protein sequence database. Then we examined their genomic environment to assess the prevalence of the heterotetrameric MaoC-like hydratase motif. We found five different genomic contexts for ChsH1 and ChsH2 homologs (Figure 4-30a).[155] Importantly, inclusion of genomic proximity in the search for homologs allowed us to identify three different motifs for encoding structural homologs of ChsH1 and ChsH2. All of these homologs are found in bacterial families known to catabolize steroids.[68]

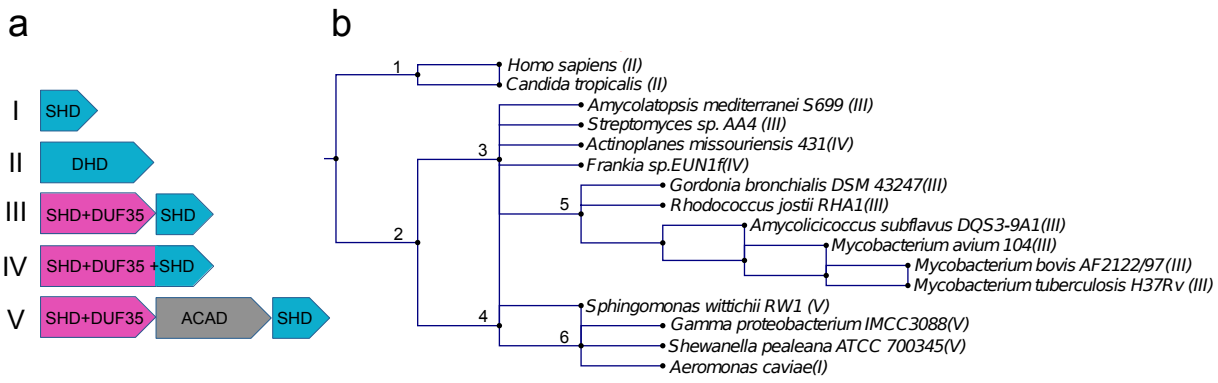


Figure 4-30. Phylogenetic relationships of ChsH1-ChsH2. (**a**) Homologs of ChsH1-ChsH2 were identified in five different genomic contexts by BLAST. The arrow lengths are scaled to gene lengths. SHD (cyan): single hot-dog fold; DHD (cyan): double hot-dog; SHD+DUF35 (magenta): Single hot-dog fold fused with DUF35/DUF35_N domain; ACAD (gray): acyl-CoA dehydrogenase. (**b**) Phylogenetic tree for five different genomic contexts of *chsH1-chsH2* gene homologs. 1: Domain Opisthoknot; 2:Domain Bacteria; 3: Phylum Actinobacteria; 4: Phylum Proteobacteria; 5: Suborder Corynebacterineae; 6: Class Gammaproteobacteria; Representative organisms were selected for each category. The Roman numerals in parentheses after each organism correspond to the gene organization in (a).

The two most common classes of homolog are single hot-dog fold and double hot-dog fold MaoC-like hydratases encoded without any neighboring MaoC-like enoyl-CoA hydratase genes. Typically, they form homodimeric MaoC-like enoyl-CoA hydratases like AcRH or CtRH as described above. These are evolutionarily most distant to *chsH1* and *chsH2* and represent the previously identified MaoC-like hydratases. The third category contains two homologous genes in the same adjacent context as *chsH1*, a single hot-dog fold, and *chsH2*, a single hot-dog fold fused to a DUF35/DUF35_N domain, which is hypothesized to bind and deliver acyl-CoA moieties in acyl-CoA-utilization processes.[155, 156] This arrangement is predominantly in Corynebacteria, which includes Mycobacteria (Figure 4-30b).

In less closely related Actinobacteria, the *chsH1-chsH2* motif exists as a single fused gene. The sequence alignments suggest that these proteins will form a homodimer that is structurally similar to the ChsH1-ChsH2 heterotetramer, i.e. a Rosetta fusion protein.[157, 158] Whether these

homologs catalyze the hydration of a steroid enoyl-CoA or a structurally simpler steroid metabolite remains to be determined. The fifth category is found in Proteobacteria and is comprised of operons that encode a *chsH2* homolog separated from a *chsH1* homolog by an acyl-CoA dehydrogenase (ACAD/FadE) homolog. Interestingly, in *Agrobacterium tumefaciens* strain C58, these three proteins are encoded as a single Rosetta fusion protein suggesting that the enoyl-CoA hydratase uses the acyl-CoA dehydrogenase product as a substrate, as is the case for the *igr*-encoded enzymes (Figure 4-2a, 4-2b). However, the organization of the *igr* operon is different; the acyl-CoA dehydrogenase genes (*chsE1* and *chsE2*) are adjacent to *chsH2* (Figure 4-2a). In addition, we have found no evidence for a tight association between the ChsH1-ChsH2 enzyme and the ChsE1-ChsE2 enzyme in our work.[50]

Within *Mtb,* 14 proteins are annotated as possessing hot-dogs folds and belong to the MaoC-like-hydratase family in *Mtb*.[159-161] Analysis of their sequences suggests that they either assemble as single hot-dog folds or double hot-dog folds and form homodimers like AcRH, and CtRH respectively. We found no other *Mtb* homologs like ChsH1 (Rv3541c) and ChsH2 (Rv3542c), which would be predicted to form a dimer of heterodimers. This unique utilization of a heterotetrameric MaoC-like hydratase is in contrast to the presence of six cholesterol-regulated heterotetrameric acyl-CoA dehydrogenases in *Mtb*.[68] We conclude that in *Mtb,* this structurally unique heterotetrameric enoyl-CoA hydratase is only employed in the final step of cholesterol side chain β-oxidation.

Putative role of the DUF35/DUF35_N domain. Electrostatic potential map calculations were undertaken to investigate potential interaction sites with additional proteins or ligands. The electrostatic potential maps clearly revealed that the Coenzyme A moiety resides in a charged environment and the hydrophobic 4-ring system is buried in a neutral environment (Figure 4-31),

as one would expect. The majority of the enzyme complex surface is relatively negative (Fig 4-31) with one exception; there is a large area of positive potential circled on the surface of $ChsH2^N$ (Figure 4-31). $ChsH2^C$ is a member of the DUF35/DUF35_N family. SSO2064 is a representative of the DUF35/DUF35_N family for which a structure has been solved (PDB code: 3irb). $ChsH2^N$ ends with a long and flexible loop, which would be followed by $ChsH2^C$. There is a distinctive groove with a hydrophobic and negative potential in SSO2064.[155] This groove matches the hydrophobic 4-ring carbon system and positive potential of the $ChsH2^N$ chain circled in Figure 4-31. This prediction is consistent with the molecular envelope of full length ChsH1-ChsH2 docking result in Figure 4-17, where the structure of SSO2064 was used to fit into the molecular shape of $ChsH2^C$.

The DUF35/DUF35_N motif is hypothesized to bind and deliver acyl-CoA moieties in acyl-CoA-utilization processes.[155, 156] This proposal is consistent with the formation of propionyl-CoA as the end product of β-oxidation of 3-OPC-CoA. It is possible that propionyl-CoA binding to ChsH1-ChsH2 may serve a regulatory function. $ChsH2^N$ ends with a long and flexible loop, which would be followed by $ChsH2^C$. The high flexibility of the $ChsH2^C$ domain is proposed to deliver generated propionyl-CoAs to other enzymes, for example, enzymes that function in fatty acid biosynthetic pathways. Therefore, we hypothesize that the $ChsH2^C$ domain helps close the substrate tunnel to either positively or negatively regulate catalytic function.

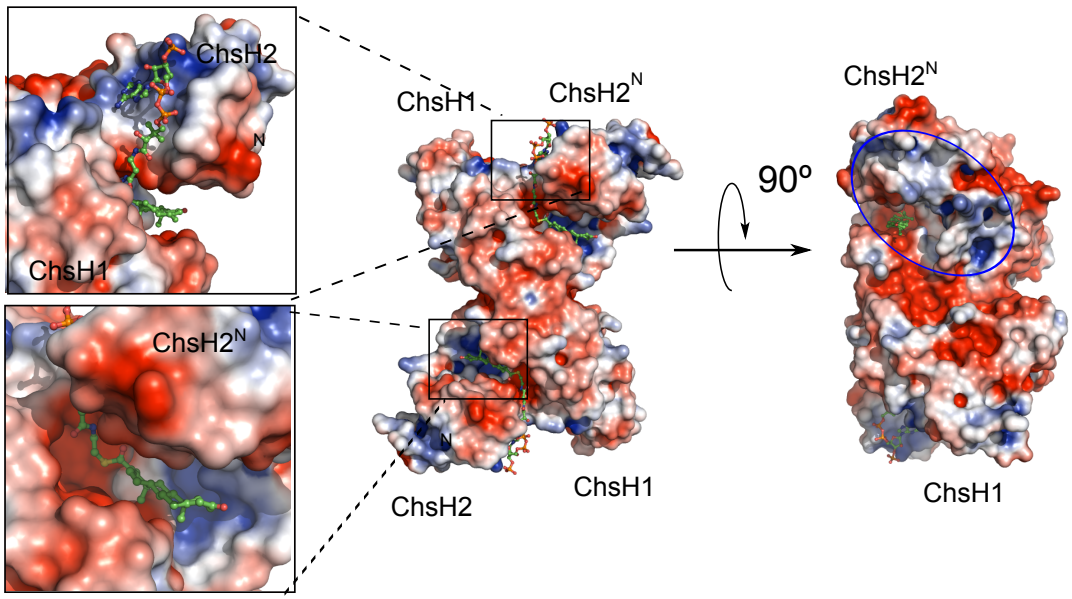


Figure 4-31. Electrostatic surface potential map of ChsH1-ChsH2^N:3-OPC-CoA. The electrostatic maps were generated using APBS in Pymol, positive potential is blue (+3 kTe^{-1}) and negative potential is red (-3 kTe^{-1}). 3-OPC-CoA is colored by atom.

Ltp2 functions as a sterol carrier protein (SCPx) thiolase in the presence of DUF35. There are five genes in the *Mtb* genome that are annotated as lipid transfer proteins (Ltp). They are *ltp1* (*Rv2790c*), *ltp2* (*Rv3540c*), *ltp3* (*Rv3523*), *ltp4* (*Rv3522*), and a probable nonspecific *ltp* (*Rv1627c*). All the *ltp* genes are regulated by KstR1 except for *ltp1*, which suggests that they may function in the cholesterol degradation pathway. Interestingly, with no exception, all the KstR1 regulated *ltp* genes are adjacent to a gene that is annotated as a DUF35 domain containing protein. The special gene organization suggests that they may function together.

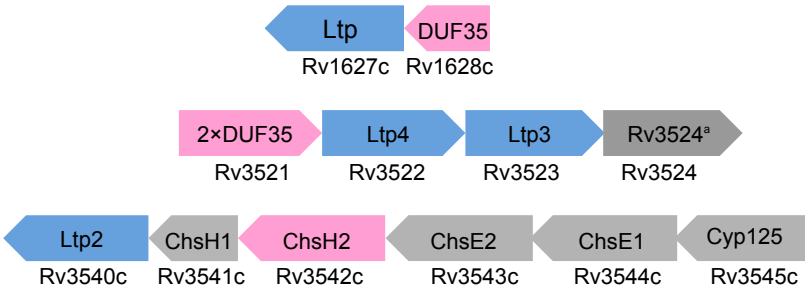


Figure 4-32. KstR 1 regulated *ltp* genes and their operonic gene organizaitons in the *Mtb* genome. [a]Rv3542 belongs to the NHL superfamily and is probably a conserved membrane protein.

Ltp2 has been successfully purified as a fusion of MBP (Figure 4-33). In gel-trypsin digestion/MALDI-TOF mass spectrometry confirmed the identity of the protein (Figure 4-33). MBP-Ltp2 was assayed for thiolase activity with 17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA (17-HOPC-CoA) in 100 mM TAPS buffer pH 8.0. Product formation was investigated by analysis of the assay mixture by MALDI-TOF mass spectrometry. However, no androstenedione or propionyl-CoA was observed. The result showed that MBP-Ltp2 was not active towards 17-HOPC-CoA under the condition we used.

Surprisingly, by analyzing the assay mixture of full length ChsH1-ChsH2 with 3-OPDC-CoA as the substrate by MALDI-TOF mass spectrometry, we observed both the hydrated product, 17-HOPC-CoA, as well as propionyl-CoA. Propionyl-CoA is one of the products after Ltp2 catalyzes cleavage of 17-HOPC-CoA (Figure 4-2b). In order to exclude the possibility that the substrate used here may be contaminated by propionyl-CoA, we analyzed the reaction mixture of ChsH1-ChsH2^{N} with 3-OPDC-CoA as the substrate. Propionyl-CoA was not generated (Figure 4-35). Therefore, we propose that a small amount of Ltp2 has been co-purified with ChsH1-ChsH2 when it is expressed from the *pigr3* construct (Figure 4-3a). Ltp2 may only

be active when in the presence of the DUF35 domain, which is in the ChsH2^{C} domain. This result is consistent with the predicted function of the DUF35 domain that it is responsible for delivering propionyl-CoA,[155] and the DUF35 domain may be required for Ltp2 to associate with ChsH1-ChsH2. More work needs to be done to detect the existence of Ltp2 in the purified ChsH1-ChsH2 protein from the *pigr3* construct.

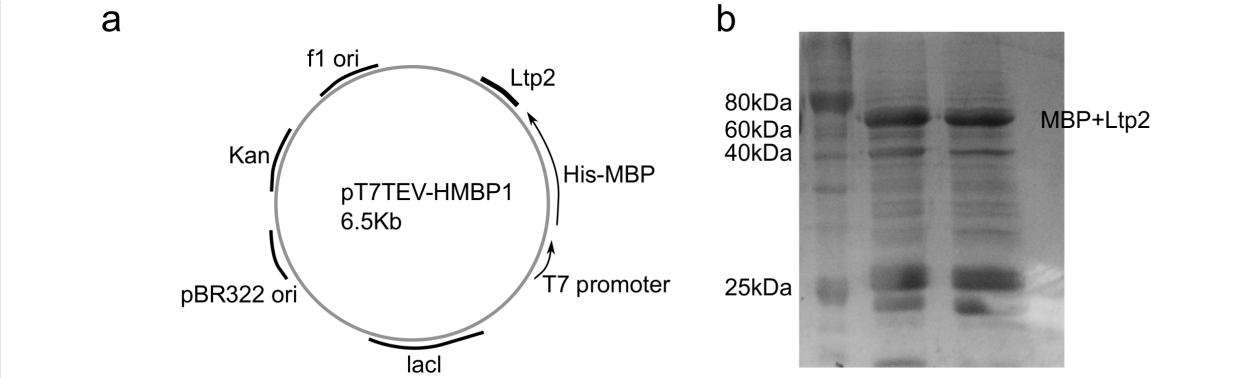


Figure 4-33. Ltp2 was fused to MBP and purified. (a) The contruct map of Ltp2 in the T7TEV-HMBP vector. (b) SDS-PAGE gel of Ltp2 fused with MBP.

ChsH1-ChsH2^{N} and full length ChsH1-ChsH2 have different propionyl-CoA binding properties. If DUF35 domain helps deliver the generated propionyl-CoA from the Ltp2 cleavage, ChsH1-ChsH2^{N} and ChsH1-ChsH2 will have different propinoyl-CoA binding affinities. In order to compare their propinoyl-CoA binding properties, isothermo titration calorimetry (ITC) was applied.

ChsH1-ChsH2^{N} (23 μM) was kept at a constant temperature of 25 °C while 1.7 mM propinoyl-CoA was titrated into the protein solution in a 5 μL increments. Binding was observed to be exothermic (Figure 4-34a). ChsH1-ChsH2 (46 μM) was also equilibrated at a constant temperature of 25 °C while 5 mM propionyl-CoA was titrated into the protein solution in a 5μL

increments. Interestingly, no propinoyl-CoA binding was observed for full length ChsH1-ChsH2 (Figure 4-34b).

ChsH1-ChsH2^{N} can bind CoA thioesters, like octanoyl-CoA, decenoyl-CoA, 3-OPC-CoA, and 3-OPDC-CoA, so it is not surprising that it can also bind propinoyl-CoA as a substrate. But full length ChsH1-ChsH2 did not display an obvious isothermo binding of propinoyl-CoA. We think that it is possible the binding of propinoyl-CoA for ChsH1-ChsH2 is in a very fast reversible equilibrium and the existence of DUF35 is playing an essential role for the binding.

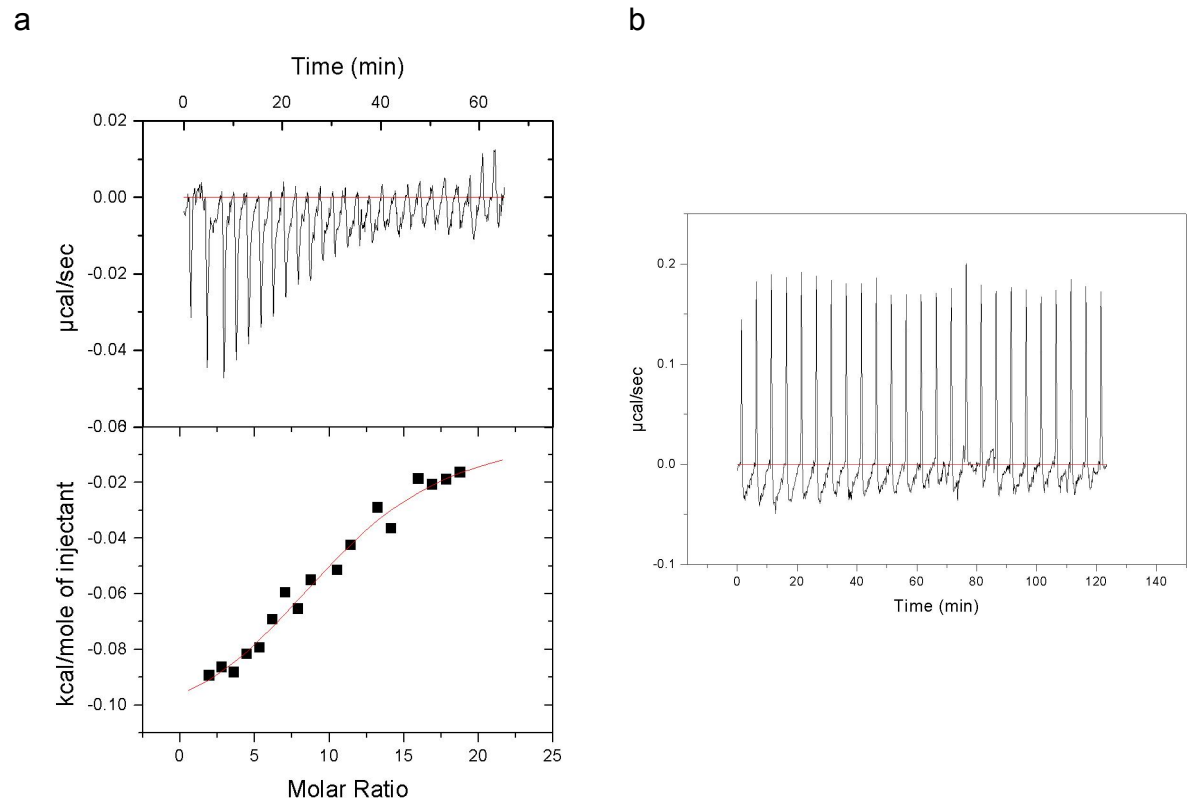


Figure 4-34. Analyzed ITC results for ChsH1-ChsH2^{N} and ChsH1-ChsH2 for titration of propinoyl-CoA. (a) The binding isotherm for titration of propinoyl-CoA (1.7 mM) into the solution of ChsH1-ChsH2^{N} (23 µM) (b) The binding isotherm for titration of propionyl-CoA (5 mM) into the solution of full length ChsH1-ChsH2 (46 µM).

In order to understand whether propionyl-CoA is a regulatory molecule regulating the activity of ChsH1-ChsH2, we tested the hydration activity of ChsH1-ChsH2 in the presence of propinoyl-CoA. ChsH1-ChsH2 (100 nM) was assasyed for the hydratase assay with 3-OPDC-CoA as the substrate in 100 mM HEPES (pH 7.4) buffer at 25 °C with or without 50 μM propinoyl-CoA. Prelimilary data showed that the specific activity in the presence of propinoyl-CoA did not change, the values are as described in Figure 4-23. Assay conditions can be optimized in the future to measue the activity. For example, propinoyl-CoA can be pre-incubated with the enzyme, and 3-OPDC-CoA is added to the assays to initiate reactions.

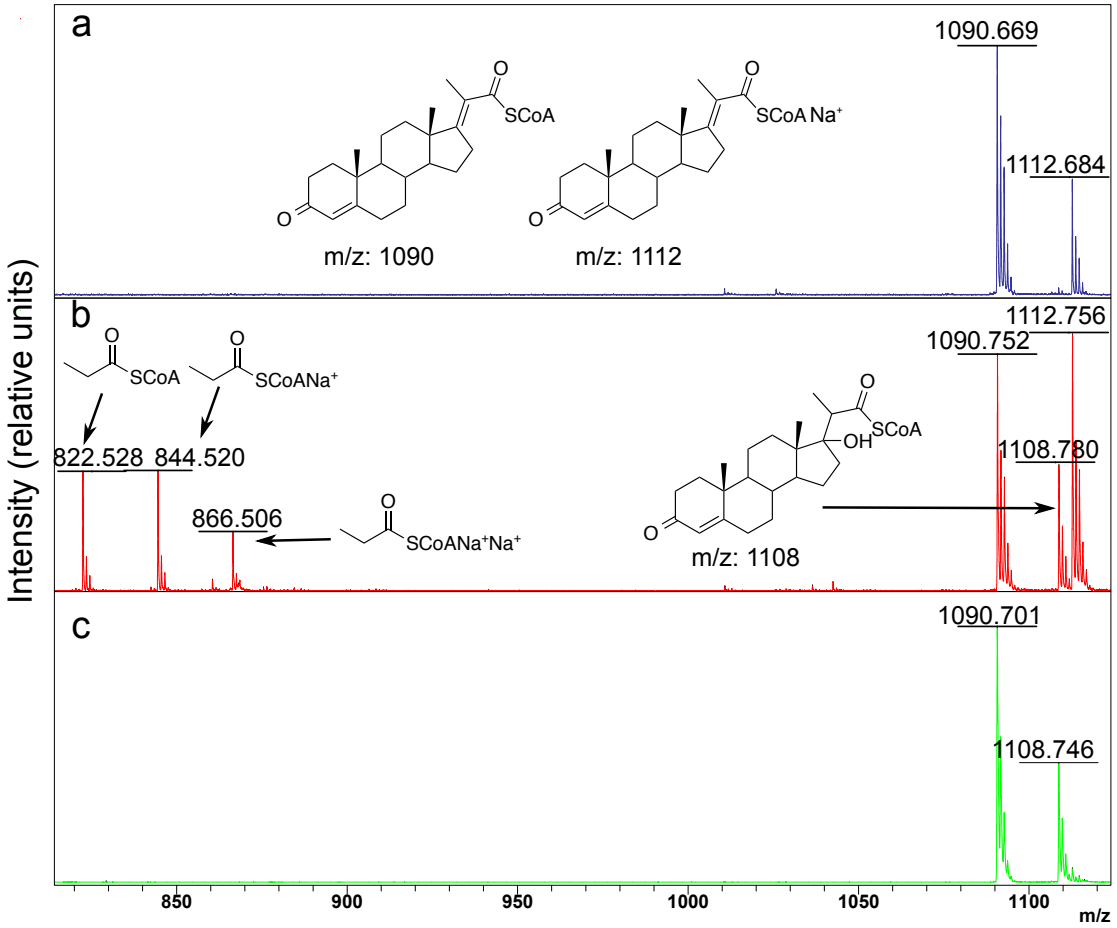


Figure 4-35. MALDI-TOF mass spectra of (a) 3-OPDC-CoA, (b) 600 nM ChsH1-ChsH2 purified from *pigr3* construct with 86 μM 3-OPDC-CoA as the substrate, and (c) 600 nM ChsH1-ChsH2^{N} purified from *pigr3* construct with 3-OPDC-CoA as the susbtrate

To validate our hypotheis that the propionyl-CoA from the assay of full length ChsH1-ChsH2 with 3-OPDC-CoA as the substrate resulted from trace amount of Ltp2, we tested whether more propionyl-CoA could be generated if MBP-Ltp2 was added exogenously. ChsH1-ChsH2 purified from the *pigr3* construct was incubacted with 3-OPDC-CoA as the substrate for 2 hours. The generation of propionyl-CoA was detected by MALDI-TOF MS (Figure 4-36). MBP-Ltp2 was added exogenously to the assay mixture and the abundance of the propionyl-CoA was monitored after 1 hour and 3 hours incubation by MALDI-TOF MS. The results showed that the abundance of propionyl-CoA increased and the abundance of 3-OPDC-CoA decreased as a function of the incubation time in the presence of MBP-Ltp2 (Figure 4-36). This result was not observed in the control group where MBP-Ltp2 was not added. Although the amount of propionyl-CoA could not be quantified by MALDI-TOF MS in this experiment, the increase of its abundance reflected that Ltp2 acted as a thiolase in the presence of ChsH1-ChsH2.

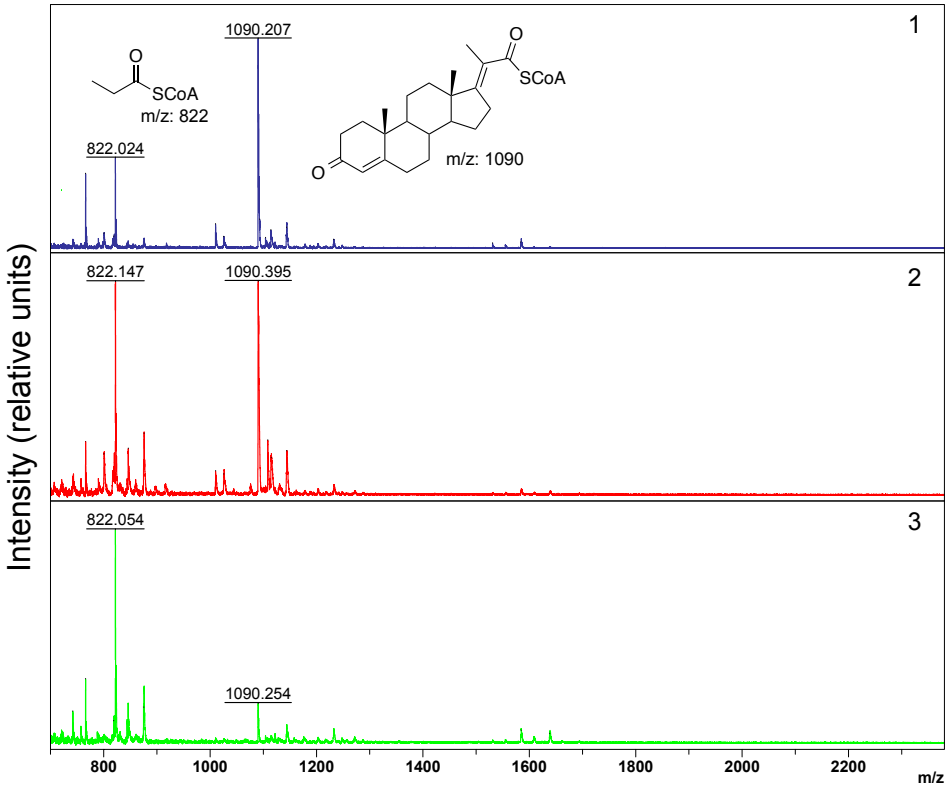
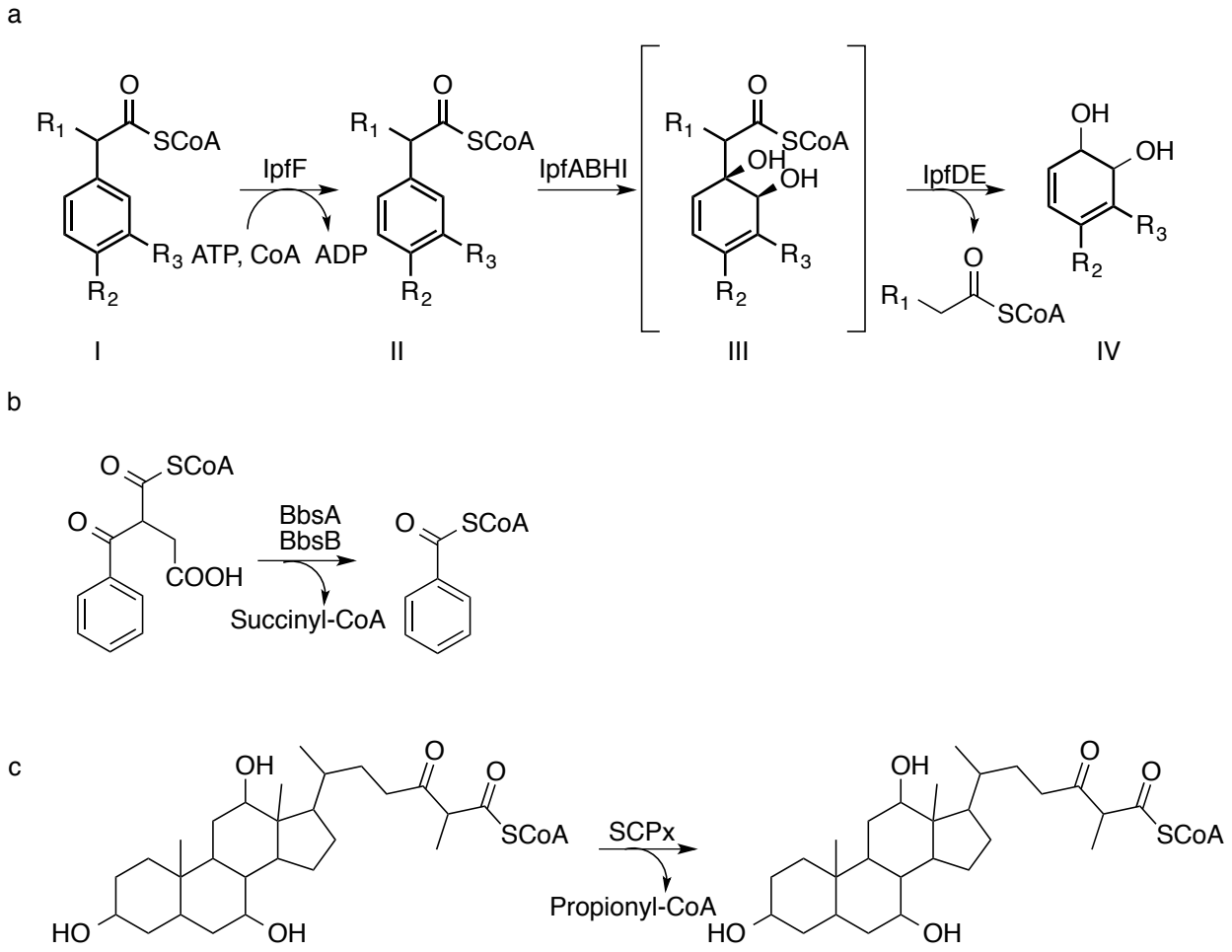


Figure 4-36. The abundance of propionyl-CoA increased when MBP-Ltp2 was added exogenously to the assay of full length ChsH1-ChsH2 with 3-OPDC-CoA as the substrate. (1) 600 nM ChsH1-ChsH2 expressed from *pigr3* construct was incubated with 86 μM 3-OPDC-CoA for 2 hours at 25 °C. (2) Following (1), 600 nM MBP-Ltp2 was added to the assay mixture and incubated for 1 hour. (3) Following (1), 600 nM MBP-Ltp2 was added to the assay mixture and incubated for 3 hours.

Ltp2 contains a highly conserved domain similar to that found in the SCPx family with an e value = $3.81 \times e^{-109}$ from a BLAST search against Non-Redundant Protein Sequence (nr) database (Figure 4-1). The requirement for a DUF35 domain with a SCPx does not only exist in *Mtb*. *Sphingomonas* sp. strain Ibu-2 was isolated from a sewage treatment plant, and shown to use racemic ibuprofen as a sole carbon and energy source.[162] Five *ORFs*, *ipfADBEF*, were hypothesized to be responsible for the complete cleavage of the propionyl side chain of ibuprofen. IpfD is annotated as a SCPx thioase and IpfE is annotated as a DUF35 domain. IpfD is hypothesized to be functional as a SCPx thiolase in the presence of IpfE.[163] Another example is

in the benzylsuccinate pathway. BbsA, which contains a DUF35 motif, is required for the thiolytic decondensation of benzylsuccinate-CoA catalyzed by the SCPx thiolase, BbsB (Scheme 4-1a and 4-1b).[164-166] The mutant mice with the SCPx gene deletion accumulated 24-keto-trihydroxy cholestanoic acid-CoA (24-keto-THCA-CoA), which suggestted that the products of the SCPx gene are responsible for the cleavage of 24-keto-THCA-CoA into cholyl-CoA (Scheme 4-1c).[167]



Scheme 4-1. Sterol carrier proteins (SCPx) catalyze cleavage of α-branched β-keto thioester substrates. (a) The *ipfADBEF* operon is hypothesized to be responsible for the complete cleavage of the propionyl side chain of ibuprofen Compound I and IV were detected by GC/MS in *Sphingomonas* sp. strain Ibu-2[162]. The formation of compound II was detected *in vitro* via IpfF CoA assay. The existence of compound III is hypothesized based on the function annotations of IpfABHI. IpfE, which is annotated as a DUF35 motif containing protein, is hypothesized to be

required for the thiolytic cleavage of ibuprofen-CoA catalyzed by the SCPx thiolase, IpfD. (a) BbsA, which contains a DUF35 motif, is required for the thiolytic decondensation of benzylsuccinate-CoA catalyzed by the SCPx thiolase, BbsB, in the benzylsuccinate pathway. (c) The function of SCPx in mice.

The SCPx thiolase superfamily shares the catalytic site, C/G-H/Y-H/Y triad, from the BLAST search against the Non-Redundant Protein Sequences (nr) Database[168] and from literature reports.[169] As is reported, thiolase superfamily shares the conserved catalytic triad, C-H-C, as the active site. This difference in their protein sequences demonstrates that SCPx thiolase and thiolase evolved differently.[168] The sequence alignment between BbsB from *Aromatoleum aromaticum* (strain EbN1), IfpD from *Sphingomonas* sp. strain Ibu-2, Ltp2 from *Mtb* H37Rv, and homo sapien SCPx showed that they belong to the SCPx thiolase superfamily. (Figure 4-37) We think they may function similarly to take α-branched fatty acid as Homo sapien SCPx thiolase.[167]

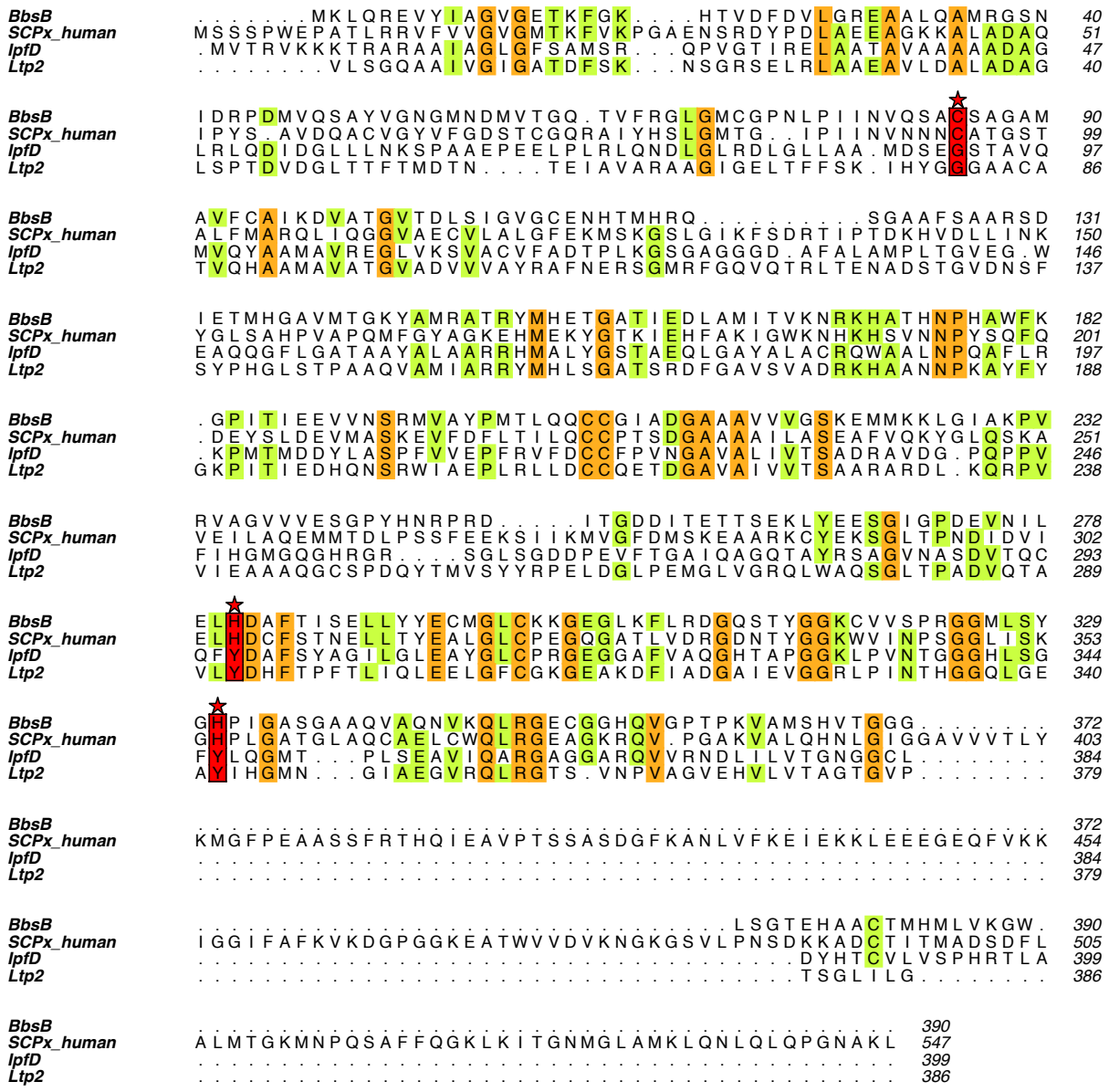


Figure 4-37. Protein sequence alignment of BbsB from *Aromatoleum aromaticum* (strain EbN1), IfpD from *Sphingomonas* sp. strain Ibu-2, Ltp2 from *Mtb* H37Rv and Homo sapien SCPx. Residues are colored based on similarity. The proposed active sites are colored in red and indicated by asterisks.

Conclusion. We successfully assigned the functions of uncharacterized *igr* genes in cholesterol side chain degradation. ChH1-ChsH2 forms a heterotetrameric MaoC-like enoyl-CoA hydratase that catalyzes the hydration of 3-OPDC-CoA to form 17-HOPC-CoA. The structure of ChsH1-ChsH2^{N} is the first report of a heteromeric enoyl-CoA hydratase and the first report of a

hydratase that function in the *Mtb* cholesterol metabolism. We believe that its novel structure and function will be a potential drug target to develop inhibitors to treat TB. Even though the function of Ltp2 has not been directly demonstrated yet, its catalytic role as a sterol carrier protein X thiolase has been tentatively established.

4. Acknowledgements

The gene constructs (*pChsH1*, *pChsH2*, *pigr3*, and *pChsh1ChsH2^{N}*) were made by Dr. Suzanne Thomas. Dr. Suzanne Thomas contributed significantly to the protein expression of ChsH1-ChsH2 and ChsH1-ChsH2^{N}. All the physical characterizations including analytical ultracentrifugation sedimentation (AUC) and UV/LC/MS of ChsH1-ChsH2 and ChsH1-ChsH2^{N} were performed by Dr. Suzanne Thomas; Dr. Kip Guja and Dr. Miguel Garcia-Diaz contributed significantly to all the crystal data collections and structure determinations.

Chapter V. Succinylation regulates the activity of ChsE4-ChsE5 in *Mtb*

5.1 Introduction

Mtb is an intracellular pathogen that is transmitted through aerosol to a new host. During exposure, *Mtb* travels to the alveolus of the lung where it encounters alveolar macrophage. Macrophages constitute the quantitatively largest and the most significant cellular reservoir during its establishment and maintenance of a chronic infection. *Mtb* infected macrophages form a foamy phenotype that is characterized by accumulated lipid bodies (LBs).[30] Microscopic analysis also revealed an accumulation of LBs inside of the pathogen.[170] *Mtb* isolated from the lungs of infected animals and infected macrophages preferentially metabolizes fatty acids rather than carbohydrates. The living environment of *Mtb* is hypoxic and lipid metabolism is the main nutrition source of *Mtb* during its infection.

In order to adapt to environmental changes and nutrition availability, *Mtb* has evolved to control its metabolic fluxes. Understanding the strategies that *Mtb* utilizes to regulate its metabolism in order to cope with changing environments will provide us a new perspective to investigate new TB therapies. Until recently, the only mode of metabolic regulation reported in mycobacteria was thought to be transcriptional regulation of metabolism related genes. A regulatory network based on 50 transcription factors revealed direct interactions between hypoxic response, lipid metabolism and the production of cell wall lipids,[171] allowing us to predict gene expression changes. However, increasing evidence showed that post-translational modifications (PTMs) were applied not only by eukaryotes but also by prokaryotes, including *Mtb*. The well-known PTMs in *Mtb* are glycosylation, lipidation, or phosphorylation, which are important regulators of protein function and compartmentalization. PTMs, like methylation, acetylation and pupylation, are involved in protein stability and function.[172] Furthermore, PTMs could also be the strategies that *Mtb* adopts for metabolism regulations.[173]

Lysine is the most frequently modified residue and it can be subjected to a variety of PTMs, including methylation, acetylation, propionylation, and others. Succinylation, as a newly discovered PTM on lysine,[174] has been proposed to play an even more significant role relative to other PTMs by introducing larger charge and steric alterations. The succinylome analysis in *E.coli*, yeast and mammalian cells demonstrated that succinylation occurs to a larger extent and more dynamically than acetylation.[175] The highly widespread lysine-succinylation in both prokaryotes and eukaryotes indicated that this PTM plays essential roles in diverse cellular processes.[176]

Recently, the succinylome analysis of *Mtb* H37Rv identified 1545 lysine succinylation sites on 626 proteins, and a large portion of those proteins are present in central metabolism pathways.[177] The metabolic network adopted by *Mtb* is a central mediator and defines its virulence. Understanding how *Mtb* regulates its central metabolism by succinylation modification is valuable for understating its pathogenesis and persistence. As a major component of LBs, cholesterol plays an indispensable role throughout the course of TB infection.[178] Therefore, it is crucial to understand how succinylation regulates cholesterol metabolism in *Mtb*.

Among the 626 succinylated proteins identified in *Mtb* H37Rv, 85 of them are proposed to be involved in lipid metabolism, suggesting that lipid metabolism is highly regulated by succinylation. 19 of the 85 succinylated lipid-related genes are regulated by KstR regulons.[177] (Figure 5-1) Previously, we successfully identified and characterized the acyl-CoA dehydrogenases (ACADs) that function in cholesterol side chain degradation.[104] ChsE4-ChsE5 is the only ACAD that is active for taking the 8 carbon side chain cholesterol intermediate, suggesting that its activity controls cholesterol side chain degradation. Understanding the effects of succinylation on the activity of ChsE4-ChsE5 as well as the succinylation regulatory system

will be valuable to elucidate the mechanisms of cholesterol metabolism control from the post translational level in *Mtb*.

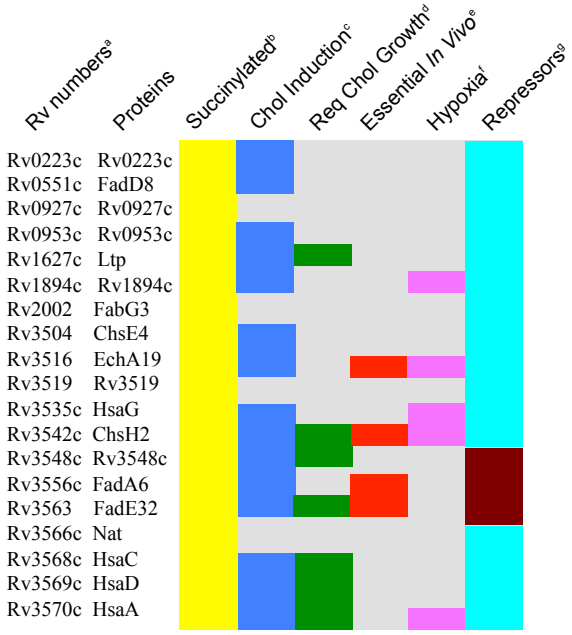


Figure 5-1. Regulation of succinylated enzymes that are cholesterol metabolism-related. There are 19 cholesterol metabolism-related gene products that are succinylated. This heatmap is sorted by Rv numbers in *Mtb* genome. [a]Rv numbers for the indicated proteins. [b]Yellow: succinylation modification.[177] [c]Blue: up-regulated when *Mtb* is grown on cholesterol[179]. [d]Green: required for *Mtb* growth on cholesterol.[82] [e]Red: Essential for *in vivo* growth in mice. [f]Magenta: up-regulated in hypoxic environment. [g]Cyan: regulated by KstR1;[46] Dark red: regulated by KstR2.[96]

As a newly discovered PTM, the succinylation regulatory system is still under investigation. In human, Sirt5 has been demonstrated as a bifunctional enzyme with both desuccinylation and demalonylation activities.[180] CobB is the only sirtuin-like protein in *E. coli*[175] and has been demonstrated as a bifunctional enzyme with equivalent activity of both deacetylation and desuccinylation.[175] Rv1151c (CobB) is the only sirtuin-like protein in the *Mtb* genome, and its deacetylase and desuccinylase activity has been demonstrated with acetyl-CoA synthase (ACS/Rv3667) as the substrate.[175, 181] Thus, the only CobB (Rv1151c) from *Mtb* is the deacetylase as well as the desuccinylase.

So far, no succinylase/succinyl-transferase has been identified in *Mtb*. Meanwhile, out of the 30 proteins[182] annotated as *N*-acetyltransferases in *Mtb*, only 5 of them have been biochemically, functionally or structurally characterized.[183-187] Rv0998 from the *Mtb* genome has the ability to acetylate a series of FadDs and the aforementioned ACS.[186] In addition, Nat (Rv3566c), belongs to the acetyl-transferase superfamily, and has been shown to acetylate a variety of arylamines.[188] However, no data has shown that these acetyl-transferases possess succinyl-transferase activity. Similar to the regulation effect of acetylation modification, succinylation has also been proposed to play a negative role on the enzymatic activity of ACS. The activity of acetyl-CoA synthetase (ACS) is inhibited by succinylation.[177] However, the negative effect of succinylation modification has not been characterized on other enzymes.

Here, we present the succinylation effect on ChsE4-ChsE5, and demonstrate that succinylation is a negative regulator of the activity of ChsE4-ChsE5. This succinylation may alter the ability of *Mtb* to metabolize cholesterol. In addition, we identified that Nat (Rv3566c), previously annotated as an N-acetyltransferase, possesses succinylase activity. By applying mass spectrometry and western blot analysis, we demonstrated that succinylation can happen non-enzymatically but proceed in a faster way in the presence of Nat.

5.2 Experimental methods

Materials, strains, media, and general methods. Total genomic DNA from *Mtb* H37Rv was obtained from the Tuberculosis Research Materials Facility at Colorado State University (Fort Collins, CO). DNA primers were ordered from Eurofins (Huntsville, AL). iProof High-Fidelity DNA Polymerase, used for gene amplification from genomic *Mtb* H37Rv DNA, was purchased from Bio-Rad Laboratories. Trypsin-ultra mass spectrometry grade was purchased from Biolabs (Melville, NY). The pET vector system from Novagen was used for cloning (Madison, WI). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). *E. coli* BL21(DE3) cells were obtained from Bio-Rad. Tryptone and ampicillin were purchased from Fisher Scientific (Pittsburgh, PA). Yeast extract was purchased from Research Products International Co. (Mount Prospect, IL). DNA sequencing was performed at the Stony Brook Sequencing Facility using an Applied Biosystems ABI 3730 (48 cm capillary array) to confirm recombinant gene sequences. Cell disruption was performed using a Constant Systems, Inc. TS Series Benchtop instrument (Kennesaw, Georgia). The following buffers were used: Buffer A (binding): 20 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazole (pH 8.0); Buffer B (elution): 20 mM Tris-HCl, 0.5 M NaCl, 0.5 M imidazole (pH 8.0); Buffer C: 50 mM Tris-HCl, 200 mM NaCl, (pH 8.0). Isolation of hexahistidine tagged recombinant proteins was achieved via immobilized-metal affinity chromatography (IMAC) using Ni-NTA His•Bind Resin (Novagen). Proteins were identified by trypsin digest peptide mass fingerprinting using a Bruker Autoflex II TOF/TOF instrument operating in positive ion mode (Bruker Daltonics Inc., Billerica, MA). FPLC gel filtration chromatography was performed on a dextran Superdex 200 16/60 column from GE Biosciences (GE Healthcare, Piscataway, NJ). Anti-succinyl lysine

antibody was purchased form PTM Biolabs (PTM Biolabs Inc., Chicago, IL). Succinyl-CoA was purchased from Santa Cruz (Santa Cruz Biotechnology, CA).

Gene cloning and plasmid construction. All genes were initially amplified from *Mtb* H37Rv total genomic DNA by PCR. PCR products were doubly digested with NdeI and HindIII restriction enzymes and ligated into a similarly digested pET28b vector. XL1-Blue *E. coli* cells were transformed with the ligated plasmids containing the gene of interest for amplification and isolation of plasmid DNA. DNA sequencing with T7 promoter and terminator primers confirmed that the full gene was present without mutations (Table 5-1).

ChsE4 lysine 238 was mutated to glutamate in *p26N/27*[104] construct using the method of quick change site-directed mutagenesis. Briefly, the following primers with the mutated nucleotide colored in red, are designed to introduce the expected mutation. After thermal cycling amplification, 1 μL of the DpnI restriction enzyme (10 U/μL) was added to digest the parental strands. 2 μL of the DpnI treated DNA product was transferred into XL1-Blue *E.coli* competent cells. The mutations were confirmed by DNA sequencing.

Forward primer: 5'-AAACGCCGGCTGGGAGCTGGTGACCAACC -3'

Reverse primer: 5'- GGTTGGTCACCAGCTCCCAGCCGGCGTTT-3'

Protein expression in *E. coli*. Competent BL21(DE3) *E. coli* cells were transformed with recombinant constructs. Single colonies were selected on LB agar plates supplemented with the 30 μg/mL kanamycin antibiotic. *E. coli* was cultured at 37°C in 2×YT media until the OD_{600} reached 0.6 – 0.8, after which the temperature was lowered to 25 °C. Cellular cultures were allowed to equilibrate, and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM to induce protein overexpression for 20 hours at 25 °C.

IMAC protein purification. Cells were harvested by centrifugation at 5000 g for 15 min at 4 °C and all subsequent steps were conducted at 4 °C. The cells were suspended in Buffer A, and disrupted using a cell disruptor (2 times, 27,000 psi). Cellular debris was removed by centrifugation at 125,000 g for 1 h. The supernatant was loaded onto IMAC His•Bind resin, washed with 10 column volumes of Buffer A, and then eluted with Buffer B. Protein solutions were immediately desalted by dialysis with Buffer C or with a G25 desalting column equilibrated in Buffer C.

Analysis of recombinant proteins. Purified proteins were analyzed by reducing SDS-PAGE and protein identities were confirmed by either in-gel or in-solution trypsin or chymotrypsin digestion and subsequent MALDI-TOF mass spectrometry. For in-gel trypsin digestion, protein bands from SDS-PAGE were excised, washed with H_2O, and dried in CH_3CN. The protein cysteine residues were reduced using 45 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate (NH_4HCO_3) for 45 min at 56 °C, and then alkylated using 55 mM iodoacetamide in 100 mM NH_4HCO_3 for 30 min in the dark. Gel pieces were dried completely by lyophilization, and rehydrated with trypsin solution (200 ng/μL) in 50 mM NH_4HCO_3 for about 45 min on ice. Excess trypsin was washed away, and the gel pieces were resoaked in 50 mM NH_4HCO_3, and allowed to digest for at least 12 h at 37 °C. This solution was combined with peptide fragments, which were extracted thrice using 50 μL of 60% (v/v) CH_3CN in aqueous 0.1% (w/v) TFA. The combined extracts were dried completely and dissolved in 10-15 μL of 0.1% TFA. This solution was mixed with an equal amount of (v/v) of α-cyano-4-hydroxycinnamic acid matrix, and MALDI-TOF data were acquired and analyzed using Flex-Analysis software (Bruker). For in-solution chymotrypsin digestion, ~50 μM of protein was mixed with 2 μL chymotrypsin solution (200 ng/μL), and the digestion was performed at 25 °C for 16 hours at pH 8.0. Digested peptides

were extracted using a C18 zip-tip and eluted with 50% (v/v) CH_3CN in aqueous 0.1% (w/v) TFA. The eluted solution was mixed with an equal amount of (v/v) of α-cyano-4-hydroxycinnamic acid (CHCA) matrix, and MALDI-TOF MS data were acquired and analyzed using Flex-Analysis software (Bruker).

Table 5-1. Expression constructs used in this work.

Construct name	Genes	Restriction sites	Purified enzyme	Source/Reference
pET28b	-	-	-	Novagen
pNat	*Rv3566c*	NdeI/HindIII	Nat	This work
p26N/27	*Rv3504/Rv3505*	NdeI/HindIII	ChsE4-ChsE5	Ref
p26N$_{K238E}$/27	*Rv3504/Rv3505*	NdeI/HindIII	ChsE4$_{K238E}$-ChsE5	This work

Dehydrogenase assay of ChsE4$_{K238E}$-ChsE5. The dehydrogenase activity of ChsE4$_{K238E}$-ChsE5 was tested with 3-OPC-CoA with the artificial electron acceptor ferricenium hexafluorophosphate (500 μM, the extinction coefficient is 4.3 mM^{-1} cm^{-1}) in 100 mM TAPS buffer (pH 8.5) as reported previously[104]. Assays were initiated by the addition of enzyme. Product formation was monitored spectroscopically at 300 nm at 25 °C. Initial velocities were obtained for the first 100-200 seconds. Controls were run without enzyme or without substrate, and showed negligible decrease in absorbance at 300 nm.

Matrix and sample preparation for MALDI-TOF MS. The matrices employed were α-cyano-4-hydroxycinnamic acid (CHCA) and 2, 5-dihydroxylbenzoic acid (2, 5-DHB). A 20 mg/mL of CHCA was prepared in a 1:1 mixture of deionized water and acetonitrile containing 0.1% TFA. A 20 mg/mL of 2, 5-DHB was prepared in a 7:3 mixture of deionized water and acetonitrile containing 0.1% TFA. Proteins after trypsin or chymotrypsin digestion were mixed with CHCA matrix in a 1:1 ratio and then spotted on a MTP 384 target plate and allowed to dry.

Small molecules was mixed with 2, 5-DHB matrix in a 1:1 ratio and then spotted on a MTP 384 target plate and allowed to dry.

Peptide sequence determination by MALDI LIFT-TOF/TOF. $ChsE4_{K238E}$-ChsE5 and $ChsE4_{K238su}$-ChsE5 were run on a SDS-PAGE gel, respectively. Gel bands corresponding to ChsE4 were excised and digested by trypsin. The tryptic digested peptides from $ChsE_{K238E}$ or $ChsE_{K238su}$ were subjected to MALDI-TOF MS analysis. Following MS acquisition, peptides bearing residue 238 were selected as the parent ions and subjected to LIFT TOF/TOF acquisition for MS/MS analysis. Fragmented peaks from the selected parental peak were analyzed by using combined peptide mass fingerprints (PMF) and MS/MS datasets were performed via BioTools with the Rapi De Novo sequencing.

Arylamine *N*-acetyltransferase assay of Nat. In order to determine the *N*-acetyltransferase activity of Nat, isonaizid (INH) was used as the substrate. Detection of acetylation of INH was performed[188] in a total volume of 500 μL. Briefly, an assay containing 800 μM acetyl-CoA, 1.5 mM INH and 150 μM 2-nitrobenzoic acid (DTNB) in 100 mM Tris-HCl buffer (pH 8.0) was incubacted at 25 °C for 5 minutes, then Nat (50 nM) was added to initiate the assay. DTNB acted as the colorimetric agent. DTNB reacts with free CoA generated from hydrolysis of acetyl-CoA by Nat to form a mixed disulfide and TNB.[189] TNB can be detected at 412 nm. The acetyltransferase activity of Nat was monitored spectroscopically at 412 nm at 25 °C. UV-visible spectra were collected as a function of time and stacked together. The product formation was confirmed by MALDI-TOF MS.

Succinylase assay and product determination. The succinylation activity of Nat (50 nM) was assayed with wild type ChsE4-ChsE5 (10 μM) and 0.5 mM succinyl-CoA in 100 mM

Tris-HCl buffer (pH 8.0). The assay was initiated by the addition of 260 nM Nat. Control experiments were conducted without Nat. The assays were incubated at 25 °C for 12 hours. Reaction mixtures were run on a SDS-PAGE gel. Protein bands matching FadE26 were cut using a scalpel and subjected to tryptic digestion. MALDI-TOF mass spectra were obtained for digested peptides and MS/MS analysis were done for peptides bearing Lysine 238.

Western blot analysis. To validate Nat catalyzed succinylation of ChsE4, we performed western blot analysis on Nat treated ChsE4-ChsE5. Briefly, 10 μM ChsE4-ChsE5 was incubated with 0.5 mM or 0.2 mM succinyl-CoA in 100 mM Tris-HCl buffer (pH 8.0) with or without 260 nM Nat at 25 °C. A volume of 10 μL of the reaction mixture would be collected at fixed time points from each reaction. All the collected samples together with WT ChsE4-ChsE4 and Nat were boiled in SDS PAGE loading buffer for 45 s, then subjected to 15% SDS-PAGE gels and transferred to polyvinylidene difluoride (PVDF) membranes at 30 volts and 4 °C for 16 hours. The PVDF membranes were blocked at room temperature for 1 hour in TBST buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl) containing 5% bovine serum albumin (BSA), and then incubated with anti-succinyl lysine antibody (1:1000, in TBST/5%BSA) overnight at 4 °C. After washing five times with TBST buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2.5 mM KCl, 0.05% Tween20), the membrane was treated with horseradish peroxidase (HRP) conjugated goat anti-rabbit antibody (1:2000 dilutions) for 1 hour at room temperature. The membranes were then washed four times with TBST buffer and one time with TBS buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2.5 mM KCl) and visualized with chemiluminescence enhancer solution using a fluorescence scanner.

5.3 Results

ChsE4$_{K238E}$-ChsE5 expression and purification. Succinylation modification has been shown to negatively regulate the activity of acetyl-CoA synthase (ACS).[177] Knowing whether other enzymes also show altered activities after succinylation is important for understanding this regulatory effect. The activity of ChsE4-ChsE5 is crucial for cholesterol metabolism *in vitro*.[104] In order to test whether succinylated ChsE4-ChsE5 displays altered activity, we mutate residue lysine 238 into glutamate from ChsE4 and obtained the mutated protein complex, ChsE4$_{K238E}$-ChsE5. The purified mutant ChsE4$_{K238E}$-ChsE5 was pulled down as a heteromeric complex (Figure 5-2a) with the characteristic FAD absorbance in the UV-visible spectrum (Figure 5-2b). ChsE4$_{K238E}$-ChsE5 was characterized by analytical gel filtration and was found to have the same elution profile as the wild type ChsE4-ChsE5, indicating the ChsE4$_{K238E}$-ChsE5 still forms a stable $\alpha_2\beta_2$ heterotetramer in solution.[104]

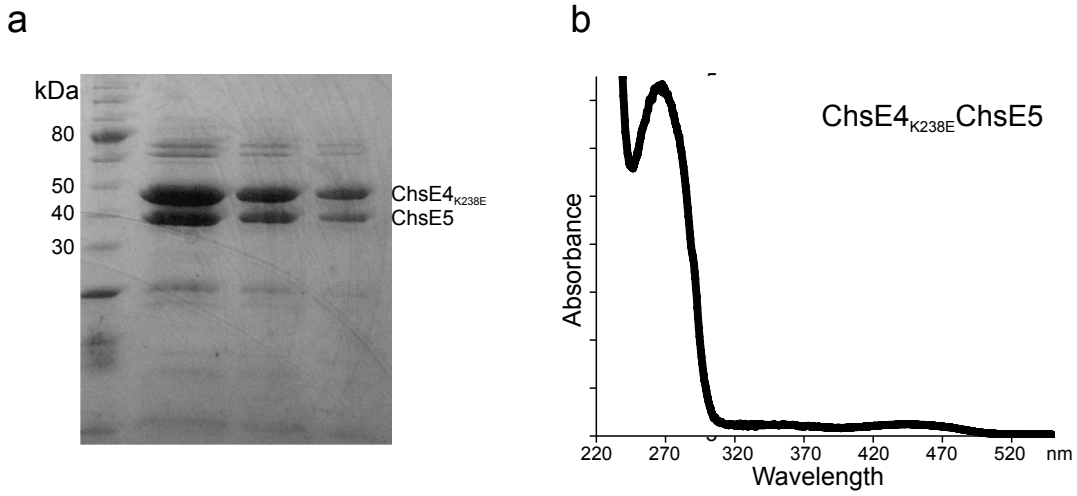


Figure 5-2. Expression and purification of ChsE4$_{K238E}$-ChsE5. (a) SDS-PAGE gel analysis of the purified ChsE4$_{K238E}$-ChsE5. Both the ladder and gel band identities are labeled. (b) The UV-visible spectrum of the purified ChsE4$_{K238E}$-ChsE5.

MS/MS analysis confirmed the K238E mutation in ChsE4. In order to confirm that glutamate was successfully introduced into ChsE4-ChsE5 at residue 238 in ChsE4, MS/MS analysis was performed on the peptide bearing residue 238 following in gel tryptic digestion. The band matching $ChsE4_{K238E}$ was cut from the SDS-PAGE gel and subjected to in gel trypsin digestion. MALDI-TOF MS analysis was first applied to investigate all of the digested peptides. The mass spectra of trypsin digested $ChsE4_{K238E}$ showed 33% coverage of its sequence. After comparing the MS results with the theoretically digested results from SequenceEditor, the peak 2194.438 (m/z) was located. Peak 2194.438 (m/z) matched the mass of the peptide from residue 230 to residue 248 (Figure 5-3).

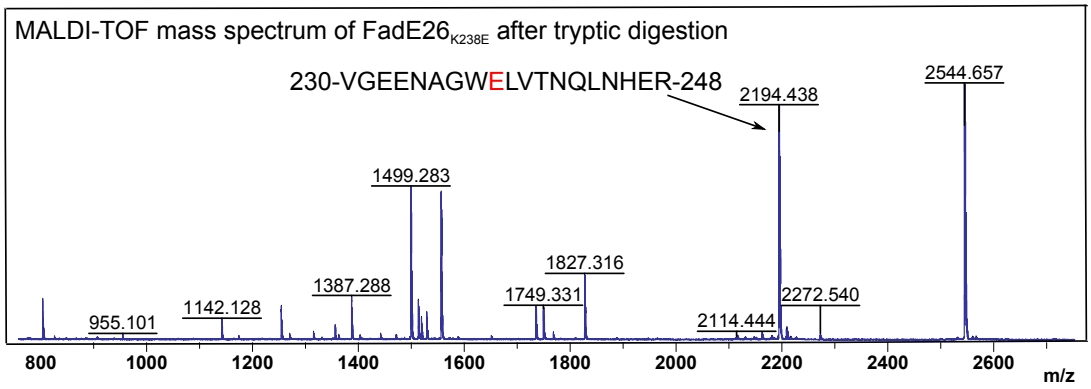


Figure 5-3. Mass spectra analysis of $ChsE4_{K238E}$ after trypsin digestion. The peak 2194.438 matches the mass of peptide from residue 230 to residue 248. The potential sequence of the peptide (230-248) is shown. The mutation is colored in red.

The peak 2194.438 (m/z) was selected and subjected to MALDI TOF/TOF analysis. The results were analyzed using BioTools. A peptide consisting of 19 amino acids, 12 y ions, 9 b ions and 3 a ions were observed (Figure 5-4). The MS/MS fragmentation analysis of this peptide revealed that glutamate successfully replaced the wild type lysine at position 238 in ChsE4 (Figure 5-4).

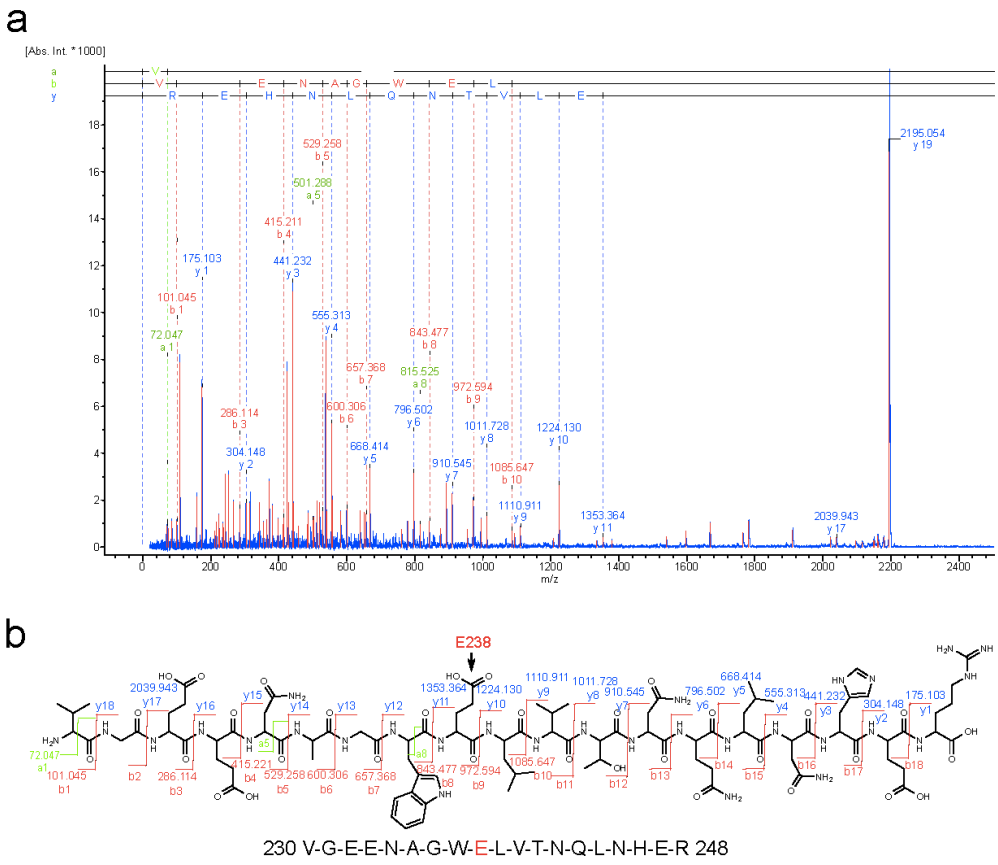


Figure 5-4. MS/MS analysis of the peptide 230-248 confirmed the mutation. (a) The peaks for the y ions are shown in blue. The peaks for the b ions are shown in red and the peaks for the a ions are shown in green. Residues that match the observed peaks were calculated in BioTools using de novo sequencing. The calculated residues are shown on top of the spectrum. (b) The molecular structure of the mutant peptide is shown and all the identified peaks are labeled. Glu_{238} is identified by an arrow.

$ChsE4_{K238E}$-ChsE5 shows decreased dehydrogenase activity. Glutamate is a mimic of succinyl group *in vitro* since both bear the same charge. The activity of $ChsE4_{K238E}$-ChsE5 is a good surrogate for $ChsE4_{succinyl\text{-}K238}$-ChsE5. The activity of ChsE4-ChsE5 has been characterized previously in our lab.[104] The activity of $ChsE4_{K238E}$-ChsE5 was assayed under the same conditions as ChsE4-ChsE5 with 3-OPC-CoA as the substrate. The results showed that the K_M value of $ChsE4_{K238E}$-ChsE5 increases about ten folds relative to ChsE4-ChsE5 (Table 5-2), suggesting $ChsE4_{K238E}$-ChsE5 had much weaker substrate binding ability. In addition, k_{cat} value

of $ChsE4_{K238E}$-ChsE5 decreased (Table 5-2), reflecting that it has lower substate turnover velocity in comparision to WT ChsE4-ChsE5.

From the crystal structure of ChsE4-ChsE5 (PDB code: 4X28), lysine 238 is located at the entrance of the substrate binding pocket (Figure 5-5b). We docked 3-OCS-CoA into the binding pocket of ChsE4 and found that the CoA moiety showed a conserved binding mode.[104] In the docked structure, lysine238 is located in a flexible loop, which is close to the negatively charged CoA moiety. Therefore, we infer that the existence of lysine at position 238 in ChsE4 is to regulate the substrate binding process and substrate binding affinity.

In order to compare the substrate product ratio in the assay of ChsE4-ChsE5 and $ChsE4_{K238E}$-ChsE5, MALDI-TOF mass spectra were taken and compared after 10 minutes incubation with 3-OPC-CoA as the substrate under the maximum velocity condition (Figure 5-6). The substrate turnover ratio of wild type (WT) ChsE4-ChsE5 is about 4 times higher than the mutant $ChsE_{K238E}$-ChsE5 in the first 10 minutes. These results demonstrate the succinylation may negatively regulate the activity of ChsE4-ChsE5.

Table 5-2. Steady-state kinetic parameters for ChsE4-ChsE5 and $ChsE4_{K238E}$-ChsE5 with 3-OPC-CoA[a]

Enzyme	3-OPC-CoA		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
ChsE4-ChsE5	3.3 ± 0.7[b]	1.2 ± 0.04[b]	$(3.6 \pm 0.7) \times 10^{5}$ [b]
$ChsE4_{K238E}$-ChsE5	34 ± 6.6	0.3 ± 0.02	$(8.8 \pm 1.8) \times 10^{3}$

[a] Error bars are the standard deviations of global firs to two independent experiemnts. [b]Data from Ref 102.

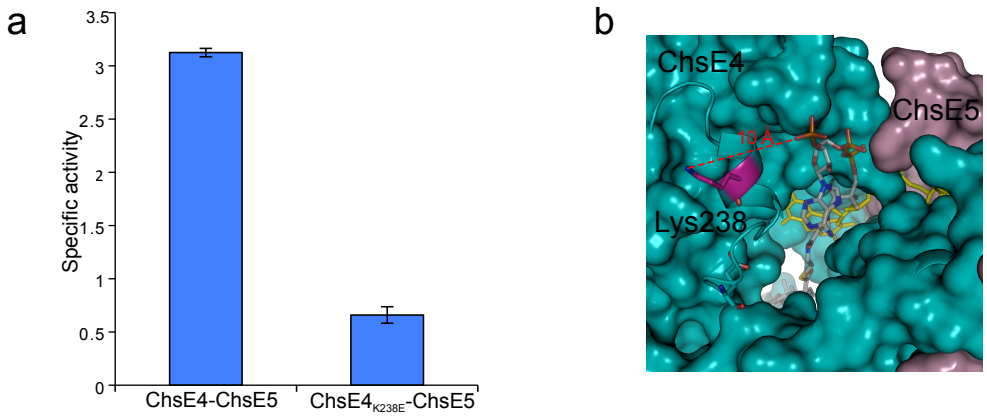


Figure 5-5. ChsE4$_{K238E}$-ChsE5 shows decreased dehydrogenase activity with 3-OPC-CoA as the substrate and the position of K238 in ChsE4-ChsE5 crystal structure. (a) The initial velocity comparison between WT ChsE4-ChsE5 and ChsE4$_{K238E}$-ChsE5. Each assay was monitored in triplicate. The error bars are shown. (b) The relative position of lysine 238, FAD cofactor and the docked substrate (3-OCS-CoA) in the crystal structure of ChsE4-ChsE5. The distance between lysine 238 and a phosphate group from the CoA moiety is shown.

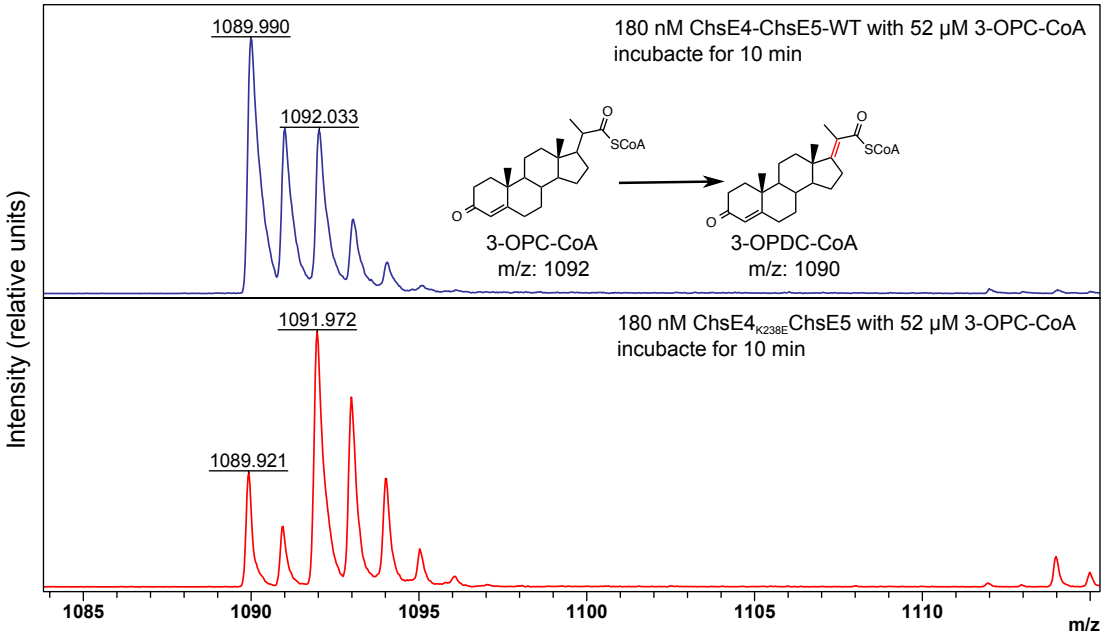


Figure 5-6. The turnover efficiencies were compared after the same incubation time of ChsE4-ChsE5 and ChsE4$_{K238E}$-ChsE5 with 3-OPC-COA as substrate.

Nat expression and purification. The acetyl-transferase activity of Nat has been demonstrated with a variety of arylamines as substrates.[188] Its importance has been highlighted in

the INH resistance *Mtb* strain.[190] The succinylome analysis of both *Mtb* H37Rv and XDR *Mtb* showed that the whole operon (Rv3566c-Rv3570c) where Nat is located was succinylated. No succinylase/succinyl-transferase has been identified yet so far in *Mtb*. Thus, Nat may act as a succinylase and be involved in cholesterol metabolism based on the fact that Nat is regulated by KstR1. In addition, some enzymes involved in cholesterol metabolism in *Mtb* are succinylated. ChsE4-ChsE5, as the first ACAD functioning in the cholesterol side chain degradation,[104] is also succinylated on the lysine 238 in ChsE4.[177] Therefore, we hypothesize that Nat is the succinyl-transferase that regulate cholesterol metabolism related proteins by succinylation.

In order to test whether Nat succinylates ChsE4-ChsE5, we cloned *nat* (Rv3566c), ligated it with pET28b with an N-terminal His_6-tag (Figure 5-7a). *E.coli* BL21 (DE3) cells were transformed with the recombined plasmid for protein expression. Nat was initially pulled down by IMAC and further purified by size exclusion chromatography. SDS-PAGE gel was used to characterize the purified protein (Figure 5-7b). In solution chymotrypsin digestion was applied to confirm the protein identity. The digested peptides were subjected to MALDI-TOF mass spectrometry. The mass spectra analysis of digested Nat showed a 40% sequence coverage (Table 5-3).

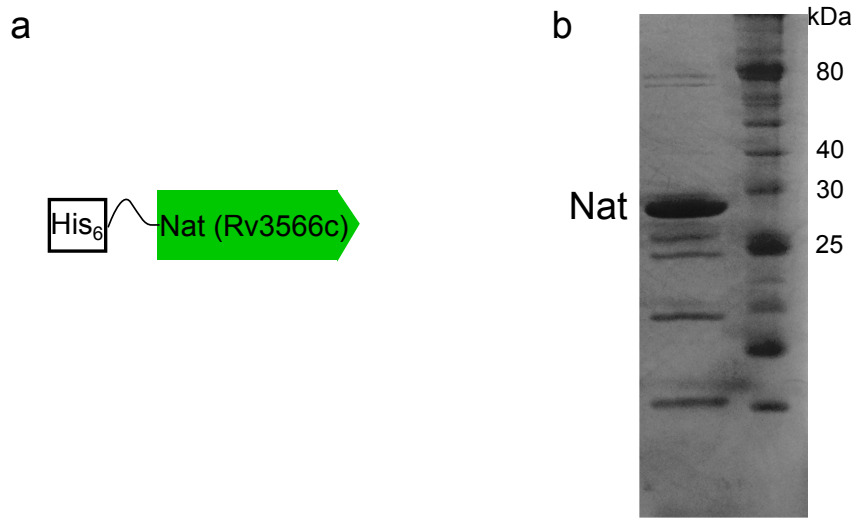


Figure 5-7. The expression and purification of Nat (Rv3566c). (a) The construct for Nat expression. (b) SDS-PAGE gel analysis of the purified Nat.

Table 5-3. Peptides identified from Nat by MALDI-TOF mass spectrometry after chymotrypsin digestion.

Residue	Sequence[a]
1-50	MAL**DLTAYF**DRINYRGATDPTLDV**LQDLV**TVHSRTIPFENLDPLL**GVPVD**
51-100	**DLSPQALAD**KLVLRRRGGYCF**EHNGLMGYVLAEL**GY**RVRRFAARV**VWKLA
101-150	PDAPLPPQTHTLLGVTFPGSGGCYL**VDVGFGGQTPTSPL**RLETGAVQPTT
151-200	HEPYRLEDRVDGF**VLQAMVRDTW**QTLYEFTTQTRPQIDL**KVASWY**ASTHP
201-250	ASKF**VTGLTAAVITDDARW**NLSGRDLAVHRAGGTEKIRLADAAAVV**DTLS**
251-283	**ERFGI**NVADIGERGAL**ETRIDELLARQPGADAP**

[a]Identified peptides are shown in bold-face, underlined. Amino acids are given in single-letter code.

Nat acetylates isoniazid. Nat was annotated as an arylamine N-acetyltransferase and its activity has been shown with a series of acylamines as substrates.[188] In order to determine whether the Nat we purified was active, we tested its acetylation activity with isoniazid (INH) as substrate. Nat catalyzed acetylation of INH results in hydrolysis of acetyl-CoA to give free CoA. DTNB reacts with free CoA in solution to produce TNB, which has a maximum absorbance at λ_{max}=412nm. The acetyltransferase activity of Nat was monitored spectroscopically at 412 nm at 25 °C as a function of time. UV-visible spectra were collected as a function of time (Figure 5-8a).

The results showed that the absorbance increased at 412 nm and the absorbance decreased at 324 nm as a function of time, which reflected the increasing amount of free CoA in solution. The reaction went to completion in 12 hours, suggesting that Nat has low activity when INH was used as the substrate, consistent with previous findings.[191] The product formation was confirmed by MALDI-TOF MS (Figure 5-8b).

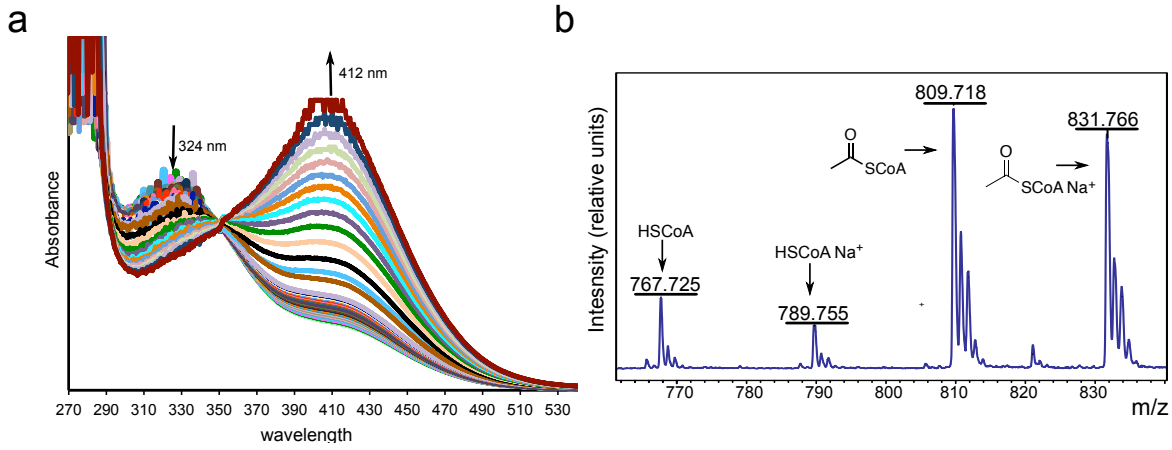


Figure 5-8. Nat acetylates isoniazid *in vitro*. (a) UV-visible spectra analysis of the acetylation ability of Nat (52 nM) with INH (1.5 mM) as the substrate. One spectrum is scanned every 180s in the first 3 hours and the peak completely shifted from 324 nm to 412 nm in 12 hours.. (b) MALDI-TOF MS confirmed the product formation of the acetylation reaction.

Nat succinylates ChsE4-ChsE5 on Lysine 238 of ChsE4. To test if Nat succinylates ChsE4-ChsE5, we incubated Nat with wild type ChsE4-ChsE5 and succinyl-CoA at 25 °C for 16 hours. The reaction mixture was separated by SDS-PAGE and in-gel tryptic digestion was performed on the gel band matching ChsE4. Mass spectrometry analysis of the trypsin digested peptides from ChsE4 showed a 43 % sequence coverage and revealed a 100 Da mass shift relative to the original peptide mass from residue 230 to residue 248 (Figure 5-9). This change in mass is consistent with the succinylated peptides. This peptide (residue number: 230-248) is the one that has been identified in the succinylome analysis in both *Mtb* H37Rv[177] and XDR *Mtb*.[192]

However, due to the low intensity, sequence determination could not be confirmed. It is possible that other succinylation sites also exist, but these sites are not revealed in this experiment. Pre-enrichment process for succinylated peptides is necessary in order to reveal other succinylation sites.

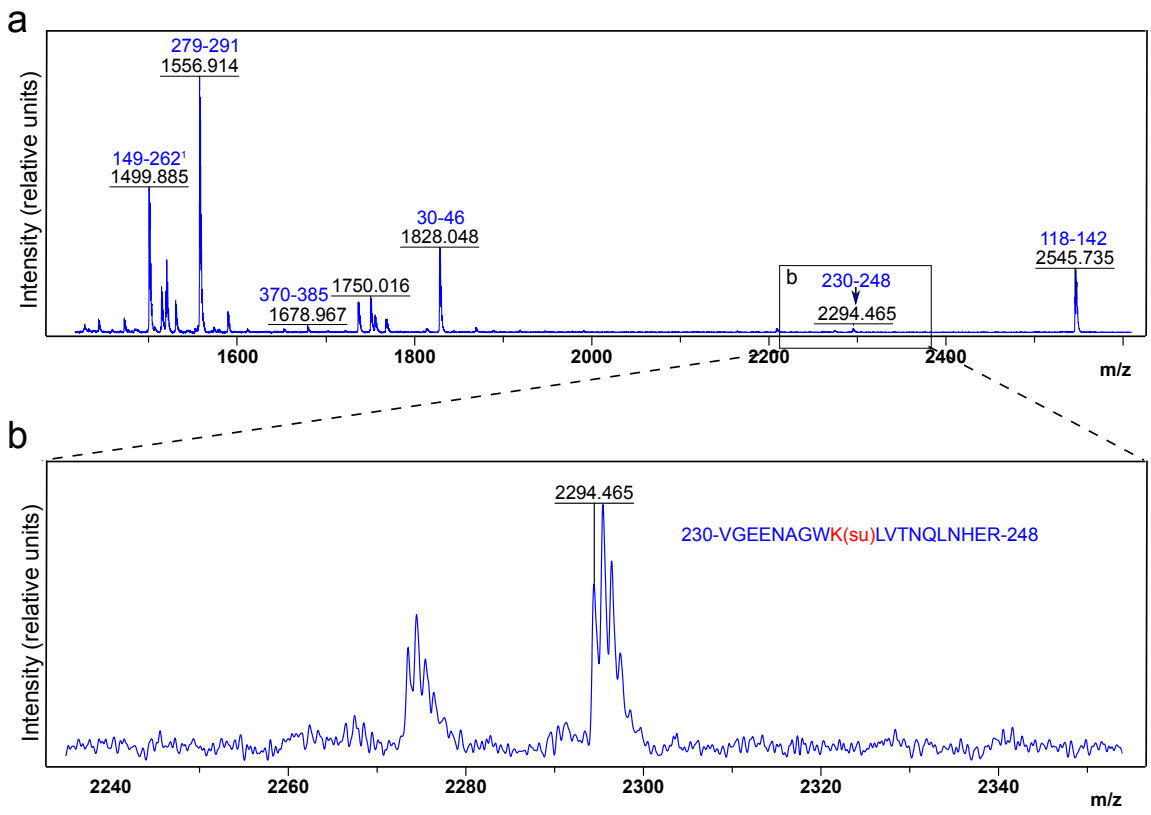


Figure 5-9. The succinylated peptide from ChsE4 identified by MALDI-TOF mass spectrometry. (a) The mass spectra of peptides from ChsE4 after trypsin digestion. The residue numbers are shown on top of the mass to indicate different peptides from ChsE4. (b) The m/z from 2230-2360 is zoomed in and the peptide sequence from 230-248 is shown. The succinylated residue, lysine 238, is colored in red.

Succinylation can happen non-enzymatically and proceed in a faster way in the presence of Nat. To further confirm that Nat can transfer a succinyl group from succinyl-CoA to the lysine residue of ChsE4, we performed western blot analysis of Nat treated ChsE4-ChsE5. To

validate our hypothesis, we set up a control experiment where ChsE4-ChsE5 was incubated in the presence of succinyl-CoA without Nat.

From western blot analysis using anti-succinyllysine antibody, succinyllysine signals were observed without the presence of Nat (Figure 5-10a and 5-10b). This result showed that succinylation can happen nonenzymatically, which is consistent with previous reports.[193] We observed strong succinyllysine signal from WT ChsE4 and the intensity of the bands increased as a function of time in the first 3 hours (Figure 5-10). In the presence of Nat, the succinylation process occurs faster than the nonenzymatic succinylation process (Figure 5-10a). This results shed light on that Nat possesses succinylase activity.

The succinyllysine signal is much weaker when $ChsE4_{K238E}$-ChsE5 was used in the succinylation reaction. This result suggests that lysine 238 may act as a major succinylation site in ChsE4 and its mutation weakens the succinyllysine signal. This is consistent with the mass spectrometry analysis where only the peptide bearing lysine 238 was identified as succinylated among other potential succinylated peptides.

Larger time scale analysis showed the succinyllysine signal from ChsE4 remained unchanged from 1 hour to 16 hours in the presence of Nat when using higher succinyl-CoA concentration (Figure 5-10b). Meanwhile, without Nat the succinyllysine signal still slowly increases and is much weaker in the first 7 hours. This result also suggests that in the presence of Nat, succinylation occurs in higher extent and faster rate. At the same time, we observed that in all the groups, succinyllysine signals for chaperone proteins and ChsE5 existed, and the intensity of the bands increased from 1 hour to 16 hours. This observation is consistent with previous

report that succinylation occurs heavily on chaperone proteins[175, 192] and ChsE5 also bears succinylation sites.

Moreover, ChsE4-ChsE5 and Nat, both purified from *E.coli*, were also subjected to immunoassays using anti-succinyllysine antibody. The results showed that no succinyllysine signals were observed for either of these two proteins. Therefore, neither ChsE4-ChsE5 nor Nat were succinylated after they were purified from *E.coli* BL21 (DE3) and the succinylation occurs only in the presence of succinyl-CoA.

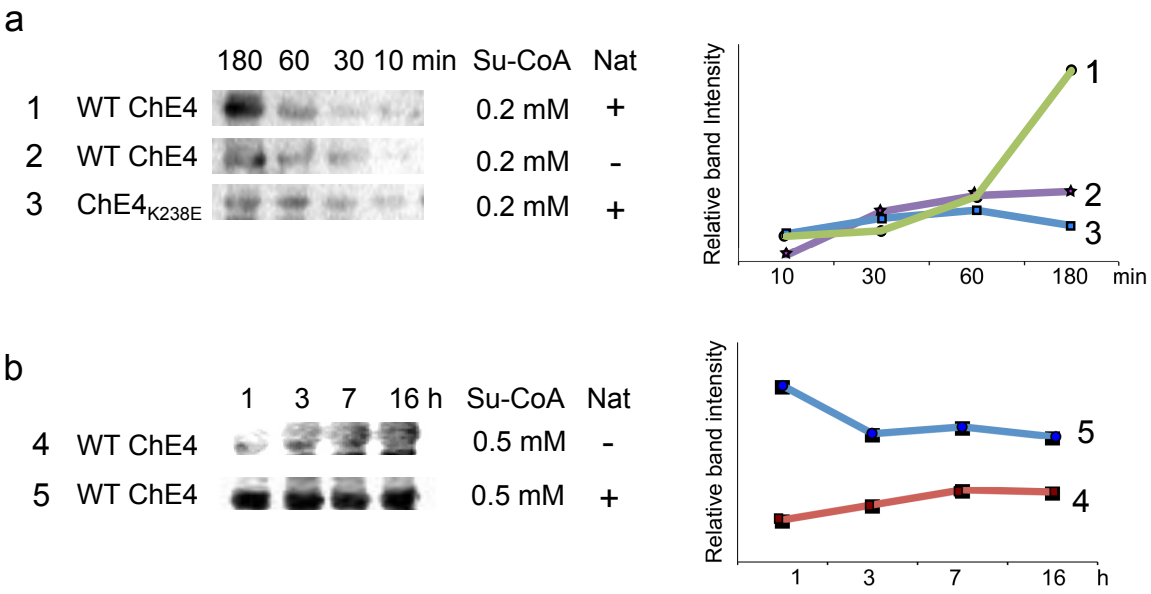


Figure 5-10. Western blot analysis of ChsE4-ChsE5 and ChsE4$_{K238E}$-ChsE5 as a function of time and the integration of the band intensities. (a) With low concentration of succinyl-CoA (0.2 mM), ChsE4-ChsE5 was incubated with Nat (1); ChsE4-ChsE5 was incubated without Nat (2); ChsE4$_{K238E}$-ChsE5 was incubated with Nat (3). Samples were collected after 10, 30, 60, 180 minutes and then subjected to western blot. The band intensities after immunoreaction with anti-succinyllysine antibody were integrated and shown in curve I, curve II, and curve III, respectively. (b) With higher concentration of succinyl-CoA (0.5 mM), ChsE4-ChsE5 was incubacted without Nat (4); ChsE4-ChsE5 was incubated with Nat (5). Samples were collected after 1, 3, 7, 16 hours and then subjected to western blot. The band intensities after immunoreaction with anti-succinyllysine antibody were integrated and shown in curve IV, and curve V, respectively.

5.4 Discussion

As one of *Mtb*'s tricks to hide from the host immune system, *Mtb* enters a non-growing but metabolically active status. This strategy results in a decrease in *Mtb*'s nutritional requirements and provides a long-term state of tolerance of multiple stresses.[26, 27] To accomplish this, a dynamic regulation must exist in *Mtb* to alter its metabolic status in order to promptly adapt to the changing environment.

As the best characterized acetyltransferase, Rv0998 acetylates acetyl-CoA transferase (ACS) with a k_{cat}/K_M of about 10^3 $M^{-1}s^{-1}$. This acetylation inactivates ACS catalysis efficiency. In addition, Rv1151c, a sirtuin-like deacetylase in *Mtb* reactivates acetylated ACS through an NAD^+ dependent manner.[181] This reversible acetylation and deacetylation regulates acetate and propionate metabolism in *M.smeg.*[194] In addition, Rv0998 directly influences fatty acid metabolism by acetylating a series of FadDs in *Mtb.*[186] This pair of enzyme, Rv0998 and Rv1151c, provides an example of how the central metabolism in mycobacteria is regulated by PTM.

Succinylation is a newly identified PTM,[174] and it occurs in a larger extent in both *E.coli*[175] and *Mtb.*[177] Therefore, succinylation may have a more profound regulation of the central metabolism in *Mtb.* Even though an increasing number of succinylated protein substrates have been identified using proteomic techniques, the succinylation regulatory system still remains unknown. Previously, we have identified ChsE4-ChsE5 as a novel $\alpha_2\beta_2$ acyl-CoA dehydrogenase that function in the first cycle of β-oxidation in cholesterol side chain degradation in *Mtb.*[104] Based on our findings, ChsE4-ChsE5 is the only ACAD that is active towards the 8-carbon cholesterol side chain metabolic intermediate,[104] so its activity is important to the

cholesterol metabolism in *Mtb*. In this report, we show for the first time that Nat (Rv3566c) can act as a succinyltransferase, and the activity of ChsE4-ChsE5 is altered by succinylation. This modification occurs on the lysine residue at position 238 in ChsE4, resulting in severely decreased activity of ChsE4-ChsE5. We postulate that reversible succinylation/desuccinylation on cholesterol metabolism-related gene products regulates cholesterol metabolism in *Mtb*. Given that there are only 5 of the 30 predicted acetyltransferases in the *Mtb* genome have been characterized,[182] it is possible that other acetyl-transferases can work as succinyltranferases too. This can be confirmed by generating the H37Rv:Δ*Rv3566c* mutant, and testing the succinylation extent of ChsE4-ChsE5 and other proteins.

By western blot analysis, we found that succinylation could happen nonenzymatically through incubating succinyl-CoA with ChsE4-ChsE5. This result is consistent with the previous observation that succinyl-CoA caused increased succinylation in a concentration dependent manner.[193] However, succinylation occurs in a much faster and more specific way in the presence of Nat.

Stronger succinylation signals were observed when H37Rv was grown in a medium containing succinate as the sole carbon source compared to pyruvate, glycerol or glucose as the sole carbon sources.[177] This result suggested that succinate/succinyl-CoA level directly influences the extent of succinylation. Very recently, it has been shown that as *Mtb* adapts to hypoxia, its TCA cycle would be slowed and remodeled, resulting in an increase level of succinate.[195] The cellular level of succinate is crucial to sustain a variety of biological processes in *Mtb*, such as ATP synthesis, membrane potentials and anaplerosis under hypoxia.[195] Succinate dehydrogenase is the enzyme that converts succinate to fumarate, and its function is crucial for *Mtb* cellular respiration in adaptation to low oxygen conditions.[196] Therefore, maintaining a

stable level of succinate is important for *Mtb* to survive *in vivo*. Given that succinylation occurs in such a large and dynamic manner, reversible succinylation and desuccinylation may be a strategy for *Mtb* to maintain its cellular succinate level (Figure 5-11). We postulate that besides regulating enzyme activity thus influencing central metabolism, succinylation/desuccinylation may be a way to maintain the cellular succinate level in *Mtb*.

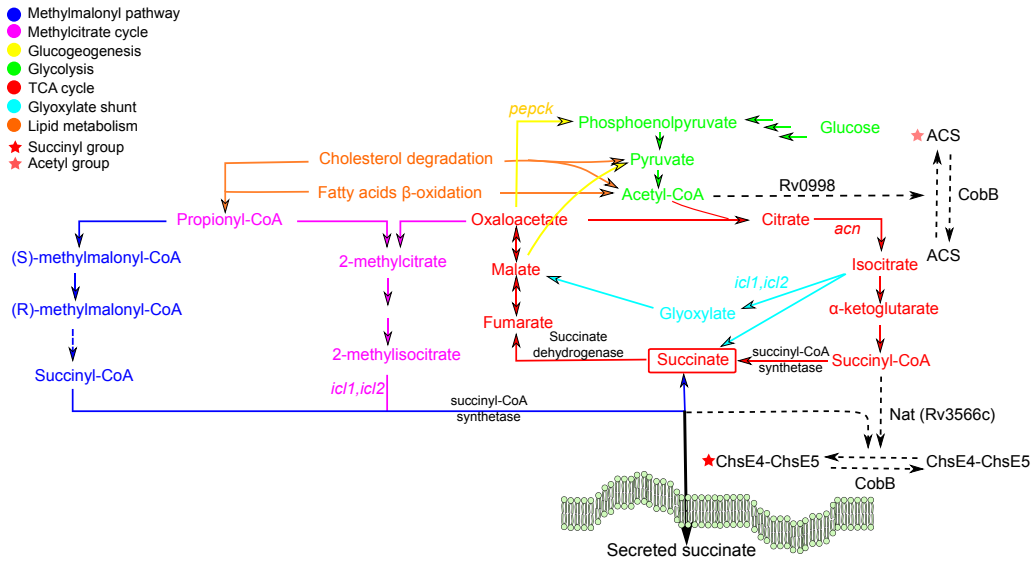


Figure 5-11. The central metabolism network in *Mtb*. Succinate is synthesized by succinyl-CoA synthetase from methylmalonyl pathway and TCA cycle. Succinate is converted to fumarate in TCA cycle. Excess succinate will be secreted out of the cell. Succinyl-CoA can be used in succinylation modification.

The succinylome analysis was conducted on *Mtb* cell lysates after in *vitro* growth and the succinylation on proteins was claimed to be in a low stoichiometry.[193] We believe that as the environment changes, the extent of succinylation will change. The accumulation of succinate in hypoxia makes us infer that enzymes involved in the central metabolism will be succinylated in a higher stoichiometry *in vivo*, resulting in low enzymatic activity and ultimately shutting down cholesterol metabolism. As the granuloma develops, *Mtb* bacteria is gradually exposed to the

airway, where the cholesterol metabolism will be turned on as desuccinylase activity increases in response to an increased level of NAD^+ in an oxygenic environment.

In conclusion, we believe that succinylation is a strategy for dynamically controlling metabolic pathways that enable *Mtb* to quickly adapt to the changing environment via promptly sensing the cellular energy status and flexibly altering reaction rates and directions.

5.5 Acknowledgements

Tianao Yuan helped purified WT ChsE4-ChsE5 that used in the activity comparison between ChsE4-ChsE5 and $ChsE4_{K238E}$-ChsE5. Dr. Peter Tonge's lab kindly provided the isoniazid used in the arylamine *N*-aceyltranferase assay of Nat.

Chapter VI. Conclusions and future perspectives

Understanding how cholesterol metabolism contributes to *Mtb*'s virulence and persistence from the molecular level is difficult due to the incomplete annotation of the genome. Wrong annotations of gene products will occur especially because of the low similarity with other known genomes. This work has made contributions to elucidate gene products with novel structures in cholesterol metabolism structurally and in *Mtb*. In addition, this work for the first time reports that *Mtb* regulates its cholesterol metabolism from post-tranlational level.

We elucidated the FadE proteome in *Mtb* and to assign the functions of cholesterol regulated FadEs. We found that *Mtb* employed novel $\alpha_2\beta_2$ heterotetrameric acyl-CoA dehydrogenases (ACADs) to metabolize cholesterol. In order to unravel the connection between the unique ACAD architecture and its biological function, we searched the whole *Mtb* genome for potential heterotetramers. We identified 6 pairs of heteromeric ACADs and characterized them all. FadE17-FadE18, FadE23-FadE24, ChsE4-ChsE5 (FadE26-FadE27), FadE31-FadE32 and FadE31-FadE33, together with ChsE1-ChsE2 (FadE28-FadE29) are the 6 pairs of ACADs transcribed from *Mtb* genome that form heterotetramers. By blasting all of their protein sequences individually against the Non-Redundant Protein Sequence (nr) Database and analyzing the prevalence of the heteromeric ACADs, we concluded that only organisms, Actinobacteria and Proteobacteria, who can metabolize cholesterol as an energy source, possessed this novel ACAD architecture, suggesting a close connection of those novel ACADs with cholesterol metabolism.

Cholesterol side chain metabolism generates one acetyl-CoA and two propionyl-CoAs, which can be used for energy production and *Mtb* surface lipids synthesis, respectively. In this

thesis we tried to identify and characterize the enzymes involved in the three cycles of β-oxidation that are required to fully metabolize the eight-carbon side chain of cholesterol.

Together with previous work in our lab we have successfully elucidated all the three ACADs in cholesterol side chain degradation. Both ChsE1-ChsE2[56] and ChsE4-ChsE5[104] have been fully characterized structurally and biochemically. ChsE3 is the ACAD functioning in the second cycle of β-oxidation, but its structure has not been revealed yet.

ChsE3 (formerly FadE34, Rv3573c) contains approximately twice the number of amino acids compared to other ACAD chains. It was previously predicted to fold similarly to VLCADs but it is still around 100 amino acids longer than typical VLCADs. Blasting its protein sequence revealed that it had two ACADs domains and each domain was similar to one typical ACAD, suggesting that each ChsE3 single chain may fold into two ACAD domains. A very recent paper demonstrated that only the C-terminal ACAD domain harbored the active residue, Glu, and the two ACAD domains are connected by a short linker.[197] Therefore, it is very possible that one single ChsE3 chain folds into a similar assembly relative to a ChsE4-ChsE5 heterodimer, with FAD cofactor bound at the interface of its two ACAD domains and two ChsE3 chains fold into an $\alpha_2\beta_2$ heterotetramer-like architecture.

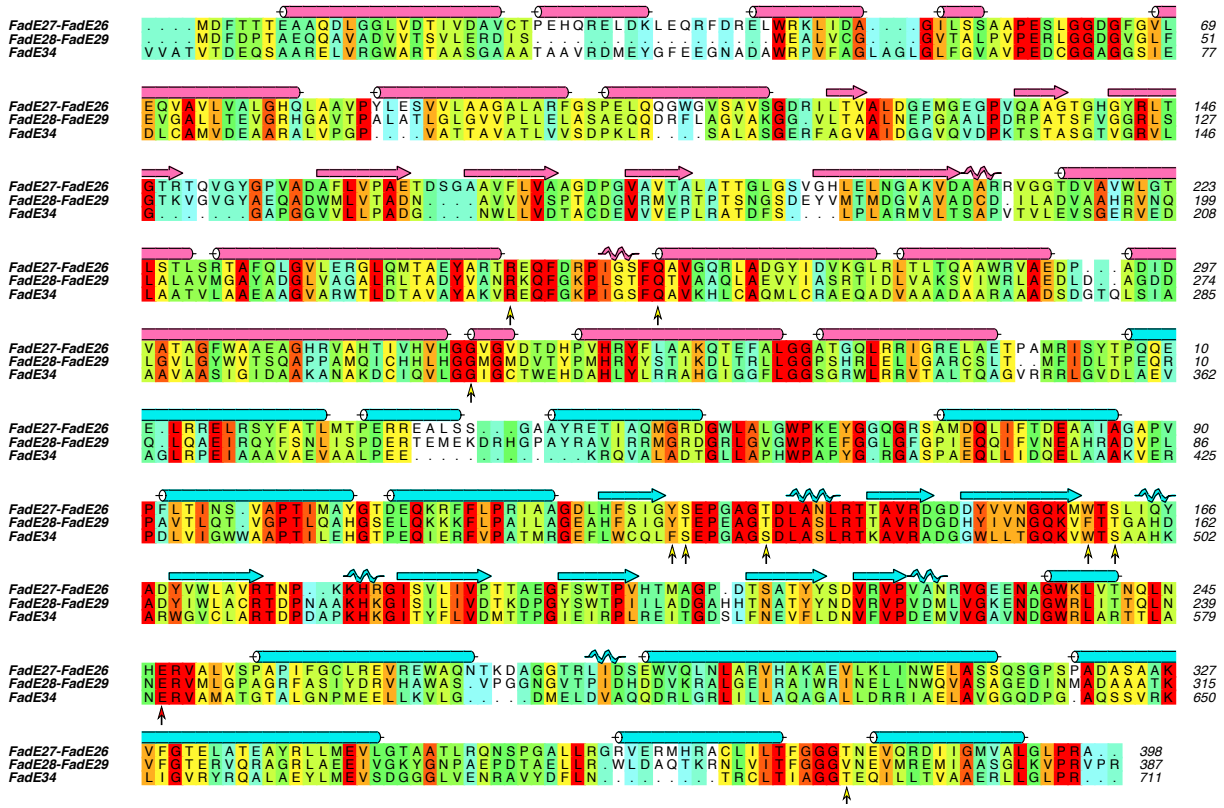


Figure 6-1. Protein sequence alignment between FadE27-FadE26 (ChsE5-ChsE4) heterodimer, FadE34 (ChsE3) single chain, and FadE28-FadE29 (ChsE1-ChsE2) heterodimer. The secondary structure elements are extracted from FadE27-FadE26 and added on top of FadE27-FadE26 protein sequence. The catalytic site is indicated by a red arrow and the FAD binding residues are indicated by yellow arrows.

Protein sequence alignment between ChsE3 single chain, ChsE5-ChsE4 heterodimer, and ChsE1-ChsE2 heterodimer revealed that the C-terminal ACAD domain from ChsE3 is more similar to ChsE4 and ChsE2, which is consistent with the location of the active site. In addition, the majority of the FAD binding residues are in the C-terminal ACAD domain (Figure 6-1). Using the sequence alignment result, a 3D model of ChsE3 was built based on the crystal structure of ChsE5-ChsE4 heterotetramer with two FADs in total (Figure 6-2). From the threaded structure, a linker was employed to connect the two ACAD domains, which were absent in ChsE5-ChsE4. Therefore, it is possible that gene fusion occurred between the two ACAD domains in *chsE3* gene. The subtle structural differences of ChsE3 with ChsE4-ChsE5 and

ChsE1-ChsE2 suggested that the second cycle of β-oxidation is less similar with the other two cycles of cholesterol side chain degradation, which may be consistent with the α-methyl-branched acyl-CoA substrates in the first and third cycle.

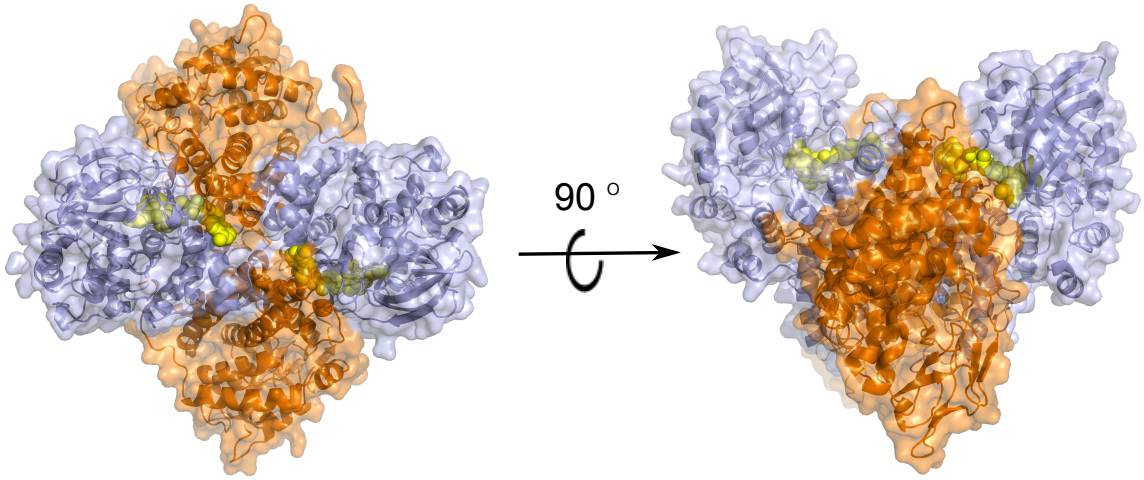
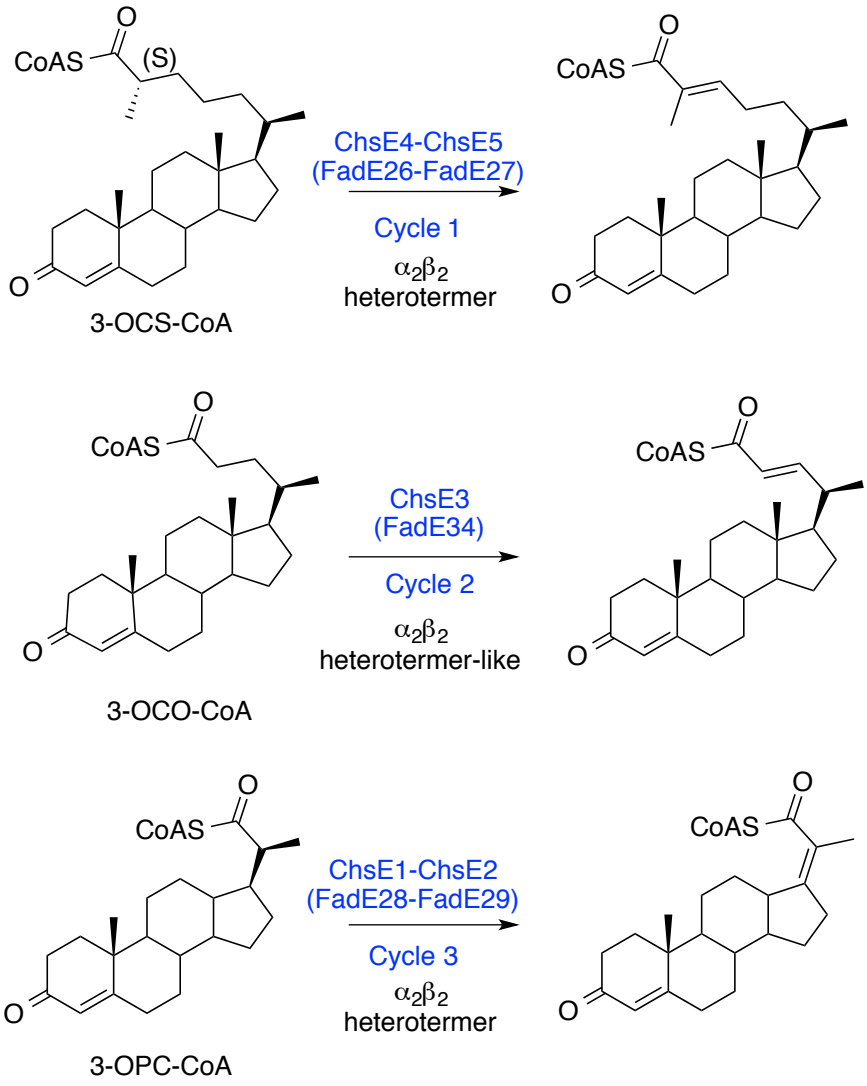


Figure 6-2. The threaded model of ChsE3 based on the crystal structure of ChsE4-ChsE5.

Therefore, all the three ACADs that have been elucidated to function in cholesterol side chain degradation adopt similar $\alpha_2\beta_2$ heterotetrameric assembly (Scheme 6-1). With one FAD binding site per ACAD heterodimer, their substrate binding pockets have been determined to evolutionally bind bulky substrates.



Scheme 6-1. ChsE4-ChsE5, ChsE3, and ChsE1-ChsE2 are the three ACADs that function in cholesterol side chain degradation in *Mtb*.

Recent work by Xie, et al., identified three pathogenicity islands (PAIs) by genome wide screening and visualizing the genomic barcodes of five *Mtb* strains. As one of the three PAIs, MPI-3 (*Rv3504-Rv3514)* was hypothesized not only to be essential in the adaptive evolution of *Mtb* by proving virulent modules but also important in strengthing the latent persistence against the adverse conditions in the host.[198] ChsE4-ChsE5 (R3504-Rv3505) encoded in MPI-3 is the only ACAD that possesses the ability to catalyze the 8-carbon side chain steroid intermediate, 3-

OCS-CoA. Thus, ChsE4-ChsE5 serves as a key target for inhibitor development. Computational screening of inhibitor scaffolds that bind to the ChsE4-ChsE5 structure has identified several potential leads. Initial enzymatic assays *in vitro* have shown promising inhibition activity (By Tianao Yuan).

Recent work from our lab has showed that growing *Mtb* CDC1551: Δ*chsE4* on cholesterol as the sole carbon source did not change the growth rate in comparison to the WT *Mtb* CDC1551. This result is consistent with previously reported that inactivation of the homologous gene of *fadE26* in *Rhodococcus rhodochrous* DSM43269 did not hamper cholesterol metabolism.[110] Interestly, upon growing *Mtb* CDC1551: Δ*chsE4* on Low -density lipoprotein (LDL) (Dr. Xinxin Yang), a peak was identified only in the *Mtb* CDC1551: Δ*chsE4* cell extracts at the retention time from 52.7 minutes to 52.9 minutes (Figure 6-3). Analysis of this peak using high resolution LC-MS showed that the molecular weight of this compound is 721.57 and it is a sodium adduct. We propose that it is a cholesterol ester with 18 carbons and intact side chain (Scheme 6-2a). NMR or MS/MS are needed to demonstrate the exact chemical structure of the accumulated peak. In addition, the *chsE4* complement strain needs to be grown under the same condition to investigate the change of this peak.

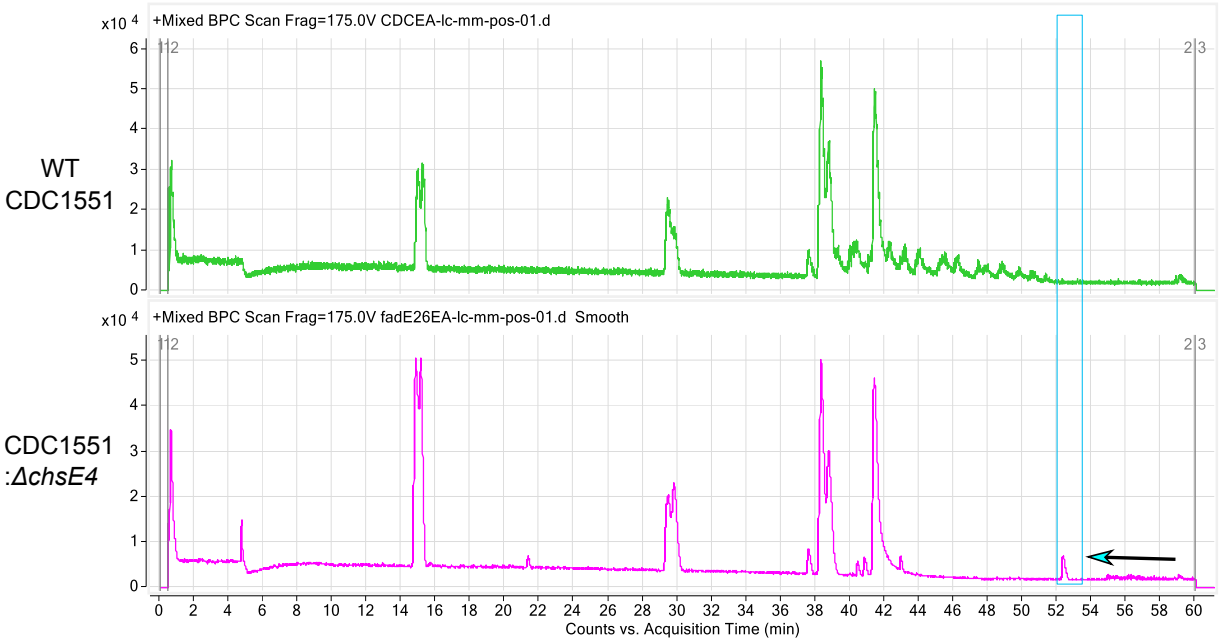
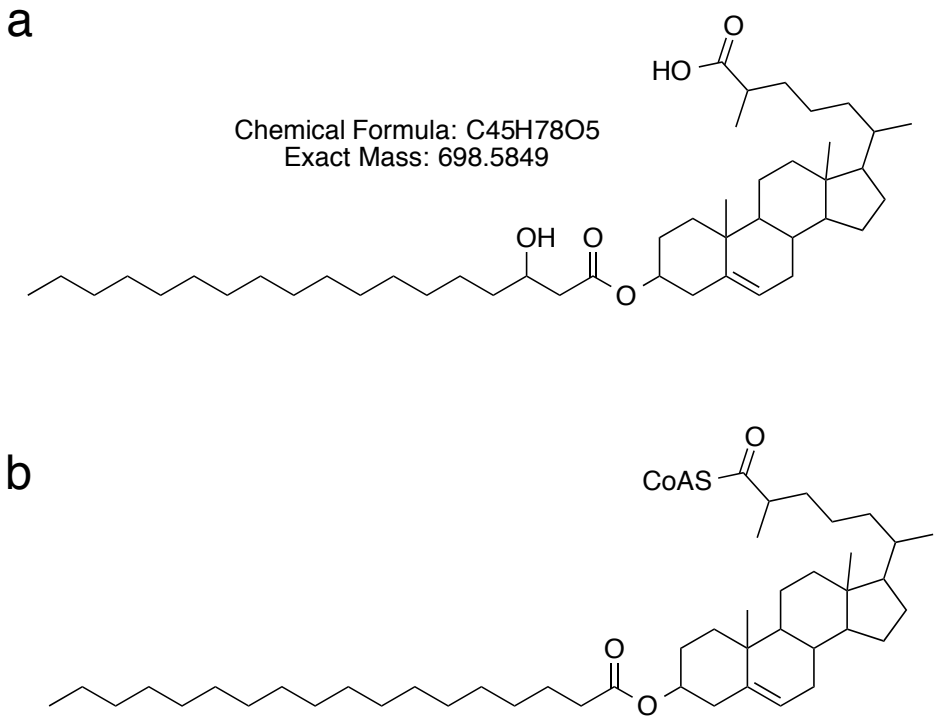


Figure 6-3. Total ion chromatograms of cell extracts from *Mtb* CDC1551 and *Mtb* CDC1551: Δ*chsE4*. The peak that only accumulated in the mutant strain was highlighted in a cyan box and pointed by an arrow.



Scheme 6-2. The proposed chemical structure of the physiological substrate of ChsE4-ChsE5. (a) The proposed chemical structure of the accumulated compound by growing *Mtb* CDC1551:

Δ*chsE4* on LDL. (b) The proposed chemical structure (3-OCS-CoA-stearate) of the physiological substrate of ChsE4-ChsE5.

The metabolite analysis of *Mtb* CDC1551: Δ*chsE4* mutant strain gives us a hint that ChsE4-ChsE5 may take cholesterol ester as its physiological substrate instead of cholesterol (Scheme 6-2b). The synthesis of 3-OCS-CoA-stearate is underway in our lab (By Rui Lu), and the enzymatic activity of ChsE4-ChsE5 will be tested with it as the substrate (Figure 6-2b).

Meanwhile, in order to test whether ChsE4-ChsE5 has the ability to bind cholesterol esters, we docked 3-hexanoic-cholest-4-en-26-oyl-CoA (3-HCS-CoA) into the crystal structure of ChsE4-ChsE5. We found that 3-HCS-CoA fits perfectly into the binding tunnel in ChsE4 and its hexanoic tail occupied the extra space where 3-OCS-CoA could not reach. 3-OCS-CoA and 3-HCS-CoA are almost in the same position (Figure 6-4b) except the extra hexanoic moiety in 3-HCS-CoA. Similar to what we have reported, 3-HCS-CoA is sandwiched in the middle of FAD and Glu247. The polar hydrogen bonding interactions are observed on CoA moiety and the remaining are located in a long, hydrophobic tunnel. This result is consistent with our prediction that ChsE4-ChsE5 may use cholesterol ester as its physiological substrate.

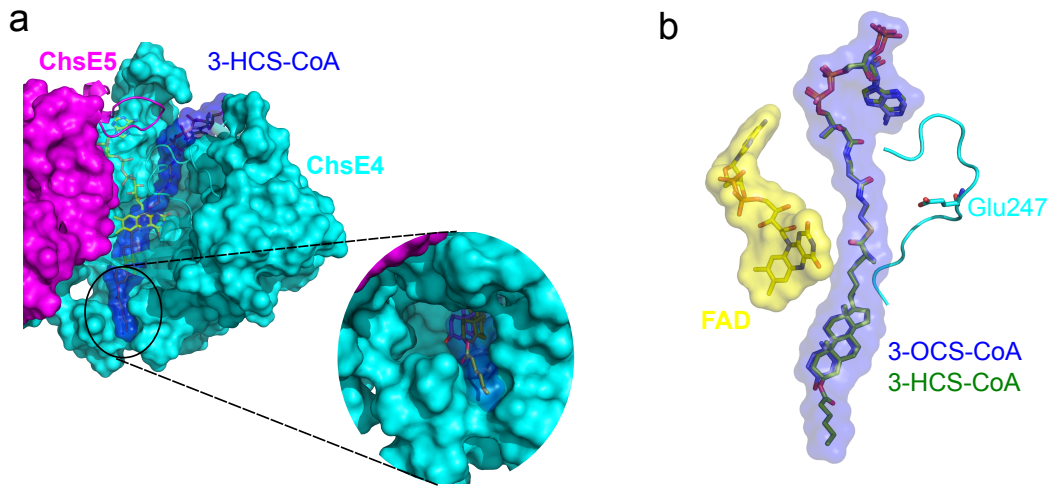
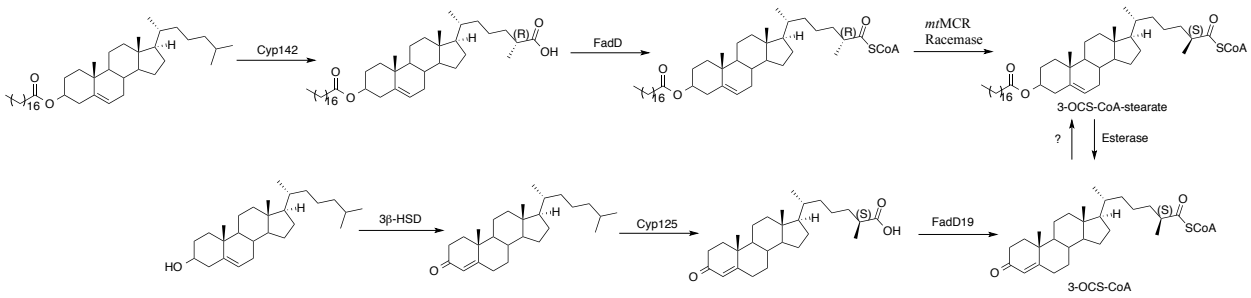


Figure 6-4. (25S)-HCS-CoA is docked into ChsE4-ChsE5 heterodimer. (a) For clarity, half of the binding tunnel is shown by surface representation. The other half of the tunnel is shown by cartoon and transparent (when viewing for the top). The FAD is set as transparent. (b) Orientation and proximity of FAD (yellow), 3-HCS-CoA (green) and the active site base (cyan). 3-OCS-CoA (blue) is also shown for comparison.

Taken together, we posit that 3-OCS-COA-stearate is the physiological substrate of ChsE4-ChsE5 (Scheme 6-2b). The degradation of cholesterol ester may share similar initial steps as cholesterol metabolism. We think that some other ACAD or acyl-CoA oxidase can take 3-OCS-CoA as substrate, and ChsE4-ChsE5 can compensate its activity in *vitro*. If that is the case, the growth rate of *Mtb*: Δ*chsE4* mutant will not attenuated when only cholesterol is available as the carbon source. (Scheme 6-3).



Scheme 6-3. The proposed pathway of cholesterol ester degradation.

In this thesis, we report an $\alpha_2\beta_2$ heterotetrameric enoyl-CoA hydratase, encoded by two adjacent genes, *chsH1* (*Rv3451c*) and *chsH2* (Rv3542c), in the *igr* operon. ChsH1-ChsH2 is also the first hydratase that has been shown to be functional towards a steroid substrate from *Mtb*. ChsH1-ChsH2 belongs to MaoC-like or (*R*)-specific enoyl-CoA hydratase superfamily, which is characterized by its hot-dog folds and highly conserved sequence motif bearing the active site.

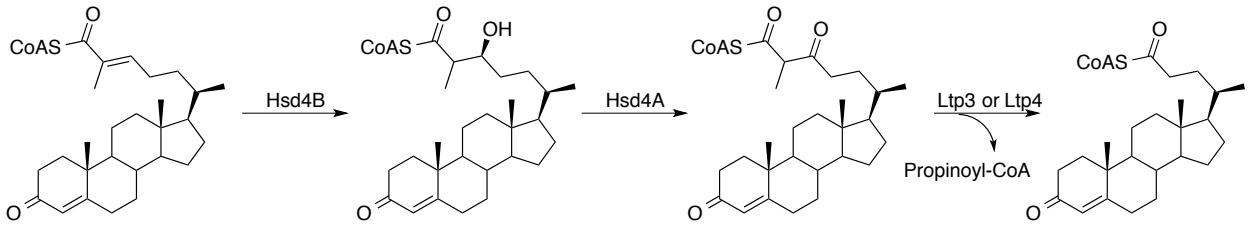
Searching the genome of *Mtb* did not identify other enoyl-CoA hydratase gene organizations similar to *chsH1* and *chsH2*, suggesting that chsH1-ChsH2 may be the only $\alpha_2\beta_2$

heterotetrameric hydratase architecture in *Mtb*. The comparison between ChsH1-ChsH2^N with other characterized MaoC-like enoyl-CoA hydratases revealed a distinctive large binding pocket in ChsH2^N even though their general assemblies are similar. The uniqueness of ChsH2^N rendered itself the ability to take steroid substrate, thus function in cholesterol metabolism. This observation suggests that ChsH1-ChsH2 can be used as a drug target.

Hsd4B (Rv3538) is the second and last KstR1 regulated (*R*)-enoyl-CoA hydratase,[39] therefore may catalyze a similar hydration step as ChsH1-ChsH2. In both the first and third cycle of β-oxidation, the substrates are α-methyl branched acyl-CoAs, which are different compared to the substrate in the second cycle. So it is possible that Hsd4B catalyzes the hydration in the first β-oxidation cycle. Then both the first and the third β-oxidation cycle need a MaoC-like enoyl-CoA hydratase to degrade cholesterol side chain. This proposition is consistent with the observation that (*R*)-enoyl-CoA hydratases are used in peroxisomal β-oxidations for α-branched fatty acid degradation.[151] The stereochemistry characterization of the products generated by ChsH1-ChsH2 has not been done yet, but the highly conserved structure of (*R*)-enoyl-CoA hydratase and the completely different catalysis mechanism between (*R*)-enoyl-CoA hydratase and (*S*)-enoyl-CoA hydratase enable us to believe that the product ChsH1-ChsH2 generated is 17-(*R*)-HOPC-CoA (Scheme 6-4).

FadBs are specific for 3-(*S*)- β-hydroxylacyl-CoAs that are generated by (*S*)-specific-enoyl-CoA hydratases, so they will not be active to take the products generated by Hsd4B. Hsd4A (Rv3502c) is the only 3-(*R*)- β-hydroxylacyl-CoA dehydrogenase regulated by KstR1, which suggests it may be only stereospecific to take 3-(*R*)-β-hydroxylacyl-CoA. Interestingly, in mammals, 3-(*R*)- β-hydroxylacyl-CoA dehydrogenase functions together with (*R*)-enoyl-CoA

hydratases in peroxisomal β-oxidation, forming MFE-2. Therefore, we think that Hsd4A functions after Hsd4B (Rv3538) in the first cycle of β-oxidation. Further biochemical analysis is needed to confirm this hypothesis (Scheme 6-4).



Scheme 6-4.The proposed functions of Hsd4A and Hsd4B and Ltp3/Ltp4 in the first cycle of β-oxidation in cholesterol side chain degradation. Hsd4B catalyzes the hydration step as an (*R*)-enoyl-CoA hydratase. Hsd4B catalyzes the dehydrogenation of 3-(*R*)- β-hydroxylacyl-CoA to form a 3-keto-acyl-CoA. Ltp3 or Ltp4 acts as an SCPx thiolase to catalyze the cleavage of 3-keto-acyl-CoA.

Our data suggests that Ltp2 functions in the last cycle of β-oxidation in cholesterol side chain degradation by converting 17-HOPC-CoA into AD and propionyl-CoA. Based on the fact that SCPx thiolases are specific to α-branched β-keto susbtrates,[167] we hypothesize that another KstR1 regulated Ltp protein will catalyze the cleavage of 3,24-dioxo-cholest-4-ene-26-oyl-CoA in the first cycle of β-oxidation. Ltp3 and Ltp4 are both regulated by KstR1 and a double DUF35 protein located adjacent to them. Therefore, we think that either Ltp3 or Ltp4 acts as the thiolase in the first cycle of β-oxidation (Scheme 6-4). This is consistent with the fact that there are no other *fadA* genes regulated by KstR1 as *fadA5* does. Our initial results suggest the existence of DUF35 is required for Ltp2 to function as an SCPx thiolase in the last cycle of β-oxidation. DUF35 domain may play roles to deliver the generated propionyl-CoA, however, its exact function is yet to be understood.

Regulating gene expression levels to modulate metabolic pathways is a commonly known strategy for *Mtb* to adapt to harsh environment. Similar to what has been discussed in Chapter II, Chapter III and Chapter IV, cholesterol metabolism related genes are upregulated *in vivo*. Their unique protein structures enable them to catalyze the degradation of cholesterol, highlighting their potentials as drug targets. However, alternating the gene expression levels is not the only way that *Mtb* uses to smartly survive and hide from the attack of host immune system.

In the past two years, an increasing amount of evidence has appeared to show that post-translational modification is also a strategy that *Mtb* uses to regulate its metabolism in order to promptly fit a new living condition (Figure 6-5). In this thesis, we demonstrated that succinylation on ChsE4-ChsE5 decreases its dehydrogenase activity. Nat, previously annotated as a N-acetyltransferase, catalyzes the succinylation on ChsE4-ChsE5.

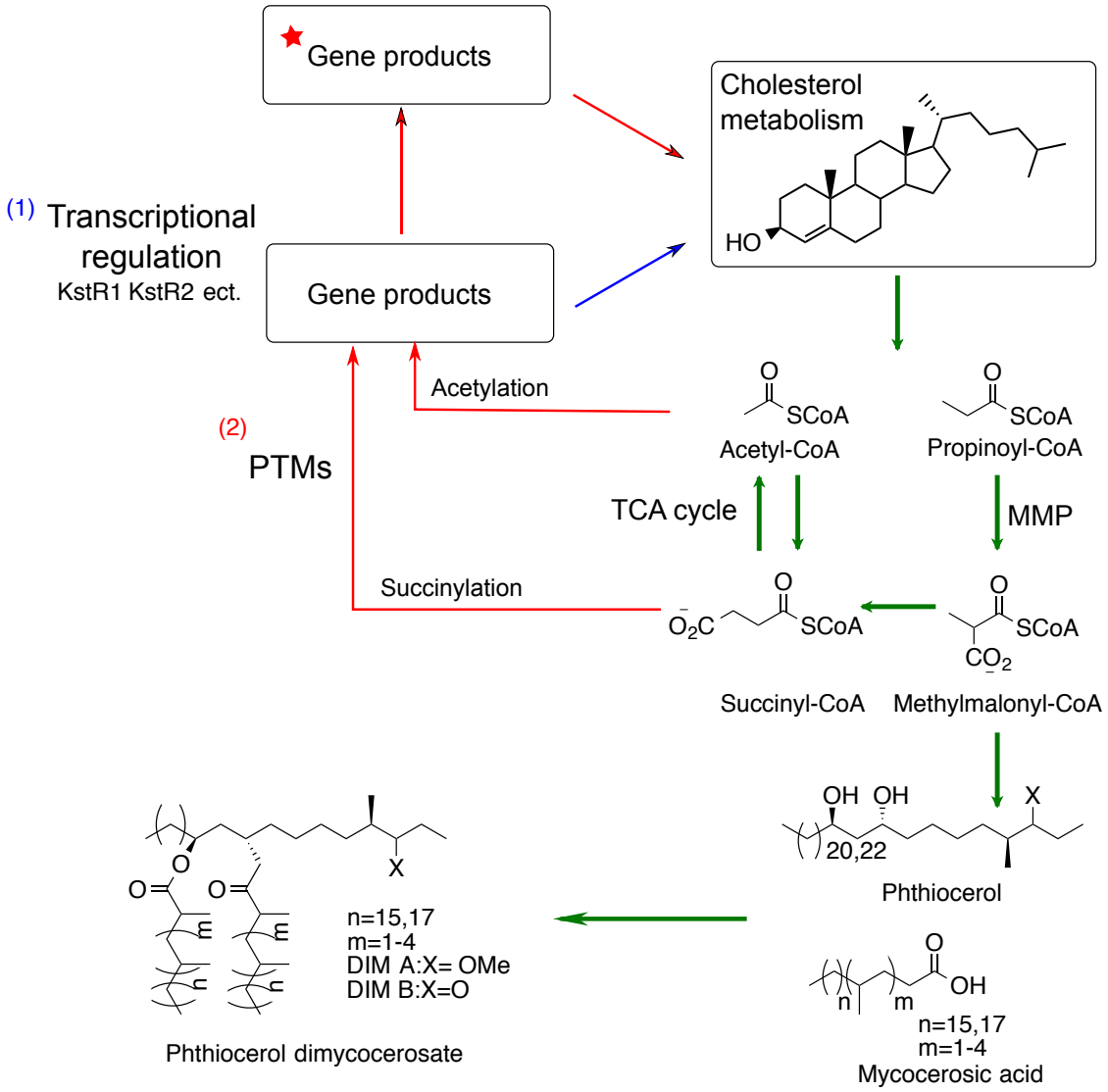


Figure 6-5. Translational regulation and post-translational regulation control cholesterol metabolism in *Mtb*. The translational regulation process is indicated by blue arrows and the post-translational regulation is indicated by red arrows. Cholesterol metabolism and the usage of its products are indicated by green arrows.

This is the first report to characterize a succinyl-modification on a cholesterol regulated enzyme, and its effects on the enzyme activity. However, there are still several questions that need to be addressed. First, we are not sure whether Nat is the only succinylase in *Mtb* genome. In order to answer this question, an H37Rv:Δ*nat* mutant needs to be generated. Western blot

analysis or immunoprecipitation can be used to analyze whether the succinylation levels change or not relative to WT H37Rv. Second, the succinylase activity of Nat is yet to be determined. The nonenzymatical succinylation has to be excluded in order to elucidate the succinylase activity of Nat. The succinylase activity of Nat can be measured by coupling it with α-ketoglutarate dehydrogenase assay. Upon Nat transfers a succinyl group from succinyl-CoA to ChsE4-ChsE5, a free CoA will be generated. α-ketoglutarate dehydrogenase catalyzes the formation of succinyl-CoA in a NAD^+ dependent manner in the presence of succinate by taking free CoA. Thus, the succinylase activity of Nat can be obtained by coupling with α-ketoglutarate dehydrogenase assay with the nonenzymatical succinylation as the control. Third, the substrate specificity of Nat needs to be elucidated. The succinylation ability of Nat on other cholesterol metabolism related enzymes, like ChsH1-ChsH2, ChsE1-ChsE2, ChsE3, FadE31-FadE32, and FadE31-FadE33, needs to be tested.

Succinylation, similar to acetylation, occurs globally. For example, at least 20 succinylated cholesterol metabolism-related proteins have been identified.[177] We are not sure what is the succinylation stoichiometry of each protein and whether lower extent succinylation can cause enzyme activity alteration. These questions also need to be addressed.

In conclusion, cholesterol metabolism is a unique pathway that has not been completely elucidated. It does not only provide *Mtb* nutrients for its survival in the host, but also is crucial for *Mtb*'s virulence and persistence. The enzymes we identified in cholesterol metabolism possess unique structures and no structure analogues exist in the host. This thesis has made contributions to explain *Mtb* cholesterol metabolism from two perspectives, translational

expression regulation and post-translational modification from molecular level, which provides a platform to develop new TB therapies.

VII. References

1. WHO. (2012) Global Tuberculosis Report.

2. WHO. (2014) Global Tuberculosis Report.

3. Sita Lumsden, E. G., and Swoboda, J. A. (1952) Isoniazid in the treatment of pulmonary tuberculosis, *Tubercle 33*, 322-329.

4. Calvori, C., Frontali, L., Leoni, L., and Tecce, G. (1965) Effect of rifamycin on protein synthesis, *Nature 207*, 417-418.

5. Kushner, S., Dalalian, H., Cassell, R. T., Sanjurjo, J. L., McKenzie, D., and Subbarow, Y. (1948) Experimental chemotherapy of tuberculosis. I. substituted nicotinamides *J. Org. Chem. 13*, 834-836.

6. Takayama, K., Armstrong, E. L., Kunugi, K. A., and Kilburn, J. O. (1979) Inhibition by ethambutol of mycolic acid transfer into the cell wall of *Mycobacterium smegmatis*, *Antimicrob. Agents Chemother. 16*, 240-242.

7. Garrod, L. P. (1950) The nature of the action of streptomycin on tubercle bacilli, *Am. Rev. Tuberc. 62*, 582-585.

8. board, G. t. c. a. (Retrieved 2012-08-18) Ethionamide, *TB online*.

9. Pestka, S. (1974) The use of inhibitors in studies of protein synthesis, In *Methods in enzymology* (Kivie Moldave, L. G., Ed.), pp 261-282, Academic Press.

10. Heifets, L., and Lindholm-Levy, P. (1989) Comparison of bactericidal activities of streptomycin, amikacin, kanamycin, and capreomycin against *Mycobacterium avium* and *Mycobacterium tuberculosis*, *Antimicrob. Agents Chemother. 33*, 1298-1301.

11. Fu, K. P., Lafredo, S. C., Foleno, B., Isaacson, D. M., Barrett, J. F., Tobia, A. J., and Rosenthale, M. E. (1992) *In vitro* and *in vivo* antibacterial activities of levofloxacin (l-ofloxacin), an optically active ofloxacin, *Antimicrob. Agents Chemother. 36*, 860-866.

12. 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.

13. Sato, K., Matsuura, Y., Inoue, M., Une, T., Osada, Y., Ogawa, H., and Mitsuhashi, S. (1982) *In vitro* and *in vivo* activity of DL-8280, a new oxazine derivative, *Antimicrob. Agents Chemother. 22*, 548-553.

14. Lambert, M. P., and Neuhaus, F. C. (1972) Mechanism of d-cycloserine action: alanine racemase from *Escherichia coli* W, *J. Bacteriol. 110*, 978-987.

15. Galietti, F., Giorgis, G. E., Oliaro, A., Boaro, D., Ardizzi, A., Barberis, S., and Massaglia, G. M. (1991) Tolerability to terizidone (TZ) in the treatment of pulmonary tuberculosis in dialyzed patients, *Minerva Med. 82*, 477-481.

16. Fajardo, T. T., Guinto, R. S., Cellona, R. V., Abalos, R. M., Dela Cruz, E. C., and Gelber, R. H. (2006) A clinical trial of ethionamide and prothionamide for treatment of lepromatous leprosy, *Am. J. Trop. Med. Hyg. 74*, 457-461.

17. Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H. W. H., 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.

18. (2007) New FDA orphan drugs: AVI-4658, SQ109, ATIR, *Medscape Medical News*.

19. 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.

20. Singh, R., Manjunatha, U., Boshoff, H. I. M., 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. (2008) PA-824 kills nonreplicating *Mycobacterium tuberculosis* by intracellular NO release, *Science 322*, 1392-1395.

21. Balasubramanian, V., Solapure, S., Iyer, H., Ghosh, A., Sharma, S., Kaur, P., Deepthi, R., Subbulakshmi, V., Ramya, V., Ramachandran, V., Balganesh, M., Wright, L., Melnick, D., Butler, S. L., and Sambandamurthy, V. K. (2014) Bactericidal activity and mechanism of action of AZD5847, a novel oxazolidinone for treatment of Tuberculosis, *Antimicrob. Agents Chemother. 58*, 495-502.

22. WHO. (2009) Treatment of tuberculosis: guidelines., *Geneva, Switzerland: WHO, 2009*.

23. WHO. (2011) Guidelines for the programmatic management of drug-resistant tuberculosis, *Geneva, Switzerland WHO, 2011*.

24. Cox, E., and Laessig, K. (2014) FDA approval of bedaquiline — the benefit–risk balance for drug-resistant tuberculosis, *N. Engl. J. Med. 371*, 689-691.

25. Russell, D. G. (2001) *Mycobacterium tuberculosis*: here today, and here tomorrow, *Nat. Rev. Mol. Cell Biol. 2*, 569-586.

26. Kussell, E., and Leibler, S. (2005) Phenotypic diversity, population growth, and information in fluctuating environments, *Science 309*, 2075-2078.

27. Kussell, E. L., Kishony, R., Balaban, N. Q., and Leibler, S. (2005) Bacterial persistence: a model of survival in changing environments, *Genetics 169*, 1807-1814.

28. Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeón, L., Abdallah, A. M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., van der Laan, T., Kant, A., Bossers-de Vries, R., Willemsen, P., Bitter, W., van Soolingen, D., Brosch, R., van der Wel, N., and Peters, P. J. (2012) ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria, *Cell. Microbiol. 14*, 1287-1298.

29. Russell, D. G. (2007) Who puts the tubercle in tuberculosis?, *Nat. Rev. Microbiol. 5*, 39-47.

30. Russell, D. G., Barry, C. E., and Flynn, J. L. (2010) Tuberculosis: What we don't know can, and does, hurt us, *Science 328*, 852-856.

31. 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.

32. Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffé, 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.

33. Deb, C., Daniel, J., Sirakova, T. D., Abomoelak, B., Dubey, V. S., and Kolattukudy, P. E. (2006) A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *Mycobacterium tuberculosis*, *J. Biol. Chem. 281*, 3866-3875.

34. Segal, W., and Bloch, H. (1956) Biochemical differentiation of *Mycobacterium tuberculosis* growth *in vivo* and *in vitro J. Bacteriol. 72*, 132-141.

35. Cole, S. T., Brosch, R., Parkhil, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry 3rd, C. E., Tekaia, 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.

36. 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.

37. 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, *Nat. Med. 11*, 638-644.

38. Marrero, J., Rhee, K. Y., Schnappinger, D., Pethe, K., and Ehrt, S. (2010) Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium*

tuberculosis to establish and maintain infection, *Proc. Natl. Acad. Sci. U. S. A. 107*, 9819-9824.

39. 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.

40. Pandey, A. K., and Sassetti, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol, *Proc. Natl. Acad. Sci. U. S. A. 105*, 4376-4380.

41. Wipperman, M. F., Sampson, N. S., and Thomas, S. T. (2014) Pathogen roid rage: cholesterol utilization by *Mycobacterium tuberculosis*, *Crit. Rev. Biochem. Mol. Biol. 49*, 269-293.

42. 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.

43. 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.

44. Rathor, N., Chandolia, A., Saini, N. K., Sinha, R., Pathak, R., Garima, K., Singh, S., Varma-Basil, M., and Bose, M. (2013) An insight into the regulation of *mce4* operon of *Mycobacterium tuberculosis*, *Tuberculosis 93*, 389-397.

45. Kendall, S. L., Burgess, P., Balhana, R., Withers, M., ten Bokum, A., Lott, J. S., Gao, C., Uhia-Castro, I., and Stoker, N. G. (2010) Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2, *Microbiology 156*, 1362-1371.

46. Kendall, S. L., Withers, M., Soffair, C. N., Moreland, N. J., Gurcha, S., Sidders, B., Frita, R., Ten Bokum, A., Besra, G. S., Lott, J. S., and Stoker, N. G. (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*, *Mol. Microbiol. 65*, 684-699.

47. Casabon, I., Zhu, S. H., Otani, H., Liu, J., Mohn, W. W., and Eltis, L. D. (2013) Regulation of the KstR2 regulon of *Mycobacterium tuberculosis* by a cholesterol catabolite, *Mol. Microbiol. 89*, 1201-1212.

48. 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.

49. Chang, J. C., Miner, M. D., Pandey, A. K., Gill, W. P., Harik, N. S., Sassetti, C. M., and Sherman, D. R. (2009) *igr* genes and *Mycobacterium tuberculosis* cholesterol metabolism, *J. Bacteriol. 191*, 5232-5239.

50. 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.

51. Nesbitt, N. M., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N. S., and Dubnau, E. (2010) A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol, *Infect. Immun. 78*, 275-282.

52. Schaefer, Christin M., Lu, R., Nesbitt, Natasha M., Schiebel, J., Sampson, Nicole S., and Kisker, C. (2015) FadA5 a thiolase from *Mycobacterium tuberculosis*: A steroid-binding pocket reveals the potential for drug development against Tuberculosis, *Structure 23*, 21-33.

53. 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.

54. Capyk, J. K., Kalscheuer, R., Stewart, G. R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W. R., Jr., 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.

55. Casabon, I., Swain, K., Crowe, A. M., Eltis, L. D., and Mohn, W. W. (2014) Actinobacterial acyl coenzyme A synthetases involved in steroid side-chain catabolism, *J Bacteriol 196*, 579-587.

56. Thomas, S. T., and Sampson, N. S. (2013) *Mycobacterium tuberculosis* utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain, *Biochemistry 52*, 2895-2904.

57. Knol, J., Bodewits, K., Hessels, G. I., Dijkhuizen, L., and Van der Geize, R. (2008) 3-Keto-5 α -steroid Δ1-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.

58. Brzezinska, M., Szulc, I., Brzostek, A., Klink, M., Kielbik, M., Sulowska, Z., Pawelczyk, J., and Dziadek, J. (2013) The role of 3-ketosteroid 1(2)-dehydrogenase in the pathogenicity of *Mycobacterium tuberculosis*, *BMC Microbiol. 13*, 43.

59. 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.

60. Capyk, J. K., Casabon, I., Gruninger, R., Strynadka, N. C., and Eltis, L. D. (2011) Activity of 3-ketosteroid 9 -hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*, *J. Biol. Chem. 286*, 40717-40724.

61. Dresen, C., Lin, L. Y.-C., 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.

62. Yam, K. C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.-X., Snieckus, V., Ly, L. H., Converse, P. J., Jacobs, 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.

63. 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.

64. Lack, N., Lowe, E. D., Liu, J., Eltis, L. D., Noble, M. E. M., 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.

65. 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.

66. Carere, J., McKenna, S. E., Kimber, M. S., and Seah, S. Y. K. (2013) Characterization of an aldolase–dehydrogenase complex from the cholesterol degradation pathway of *Mycobacterium tuberculosis*, *Biochemistry 52*, 3502-3511.

67. van der Geize, R., Grommen, A. W. F., Hessels, G. I., Jacobs, A. A. C., 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.

68. Wipperman, M. F., Yang, M., Thomas, S. T., and Sampson, N. S. (2013) Shrinking the FadE proteome of *Mycobacterium tuberculosis*: Insights into cholesterol metabolism through identification of an $\alpha_2\beta_2$ heterotetrameric acyl coenzyme A dehydrogenase family, *J. Bacteriol. 195*, 4331-4341.

69. Rhee, K. Y., Carvalho, L. P. S. d., 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.

70. Jain, M., Petzold, C. J., Schelle, M. W., Leavell, M. D., Mougous, J. D., Bertozzi, C. R., Leary, J. A., and Cox, J. S. (2007) Lipidomics reveals control of *Mycobacterium*

tuberculosis virulence lipids via metabolic coupling, *Proc. Natl. Acad. Sci. U.S.A. 104*, 5133-5138.

71. Eoh, H., and Rhee, K. Y. (2014) Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of *Mycobacterium tuberculosis* on fatty acids, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4976-4981.

72. Takayama, K., Wang, C., and Besra, G. S. (2005) Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*, *Clin. Microbiol. Rev. 18*, 81-101.

73. Carmena, R., Duriez, P., and Fruchart, J.-C. (2004) Atherogenic lipoprotein particles in atherosclerosis, *Circulation 109*, III-2-III-7.

74. Gregersen, N., Bross, P., and Andresen, B. S. (2004) Genetic defects in fatty acid β-oxidation and acyl-CoA dehydrogenases, *Eur. J. Biochem. 271*, 470-482.

75. Bahnson, B. J., Anderson, V. E., and Petsko, G. A. (2002) Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion, *Biochemistry 41*, 2621-2629.

76. Mannaerts, G., Van Veldhoven, P., and Casteels, M. (2000) Peroxisomal lipid degradation via β-and α-oxidation in mammals, *Cell Biochem. Biophys. 32*, 73-87.

77. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, 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.

78. Chen, L., and Vitkup, D. (2007) Distribution of orphan metabolic activities, *Trends Biotechnol. 25*, 343-348.

79. VanderVen, B. C., Fahey, R. J., Lee, W., Liu, Y., Abramovitch, R. B., Memmott, C., Crowe, A. M., Eltis, L. D., Perola, E., Deininger, D. D., Wang, T., Locher, C. P., and Russell, D. G. (2015) Novel inhibitors of cholesterol degradation in *Mycobacterium tuberculosis* reveal how the bacterium's metabolism Is constrained by the intracellular environment, *PLoS Pathog. 11*, e1004679.

80. 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.

81. Slayden, R. A., Jackson, M., Zucker, J., Ramirez, M. V., Dawson, C. C., Crew, R., Sampson, N. S., Thomas, S. T., Jamshidi, N., Sisk, P., Caspi, R., Crick, D. C., McNeil, M. R., Pavelka, M. S., Niederweis, M., Siroy, A., Dona, V., McFadden, J., Boshoff, H., and Lew, J. M. (2013) Updating and curating metabolic pathways of TB, *Tuberculosis 93*, 47-59.

82. Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J., and Sassetti, C. M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism, *PLoS Pathog. 7*.

83. Ghisla, S., and Thorpe, C. (2004) Acyl-CoA dehydrogenases. A mechanistic overview, *Eur. J. Biochem. 271*, 494-508.

84. Kim, J. J., and Miura, R. (2004) Acyl-CoA dehydrogenases and acyl-CoA oxidases. Structural basis for mechanistic similarities and differences, *Eur. J. Biochem. 271*, 483-493.

85. 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 A resolution: structural basis for substrate specificity, *Biochemistry 36*, 8455-8464.

86. Djordjevic, S., Dong, Y., Paschke, R., Frerman, F. E., Strauss, A. W., and Kim, J. J. (1994) Identification of the catalytic base in long chain acyl-CoA dehydrogenase, *Biochemistry 33*, 4258-4264.

87. 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. O. (2010) *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one, *Mol. Microbiol. 77*, 730-742.

88. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal, *Nucleic Acids Res. 40*, W597-603.

89. Lehman, T. C., Hale De Fau - Bhala, A., Bhala A Fau - Thorpe, C., and Thorpe, C. (1990) An acyl-coenzyme A dehydrogenase assay utilizing the ferricenium ion, *Anal. Biochem.*, 280-284.

90. Fontan, P. A., Aris, V., Alvarez, M. E., Ghanny, S., Cheng, J., Soteropoulos, P., Trevani, A., Pine, R., and Smith, I. (2008) *Mycobacterium tuberculosis* sigma factor E regulon modulates the host inflammatory response, *J. Infect. Dis. 198*, 877-885.

91. de la Paz Santangelo, M., Klepp, L., Nunez-Garcia, J., Blanco, F. C., Soria, M., Garcia-Pelayo, M. C., Bianco, M. V., Cataldi, A. A., Golby, P., Jackson, M., Gordon, S. V., and Bigi, F. (2009) Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in Mycobacterium tuberculosis, *Microbiology 155*, 2245-2255.

92. Parish, T., Smith, D. A., Roberts, G., Betts, J., and Stoker, N. G. (2003) The senX3-regX3 two-component regulatory system of Mycobacterium tuberculosis is required for virulence, *Microbiology 149*, 1423-1435.

93. Weintraub, H., Vincent, F., Baulieu, E. E., and Alfsen, A. (1977) Interaction of steroids with Pseudomonas testosteroni 3-oxo-steroid Δ4-Δ5-isomerase, *Biochemistry 16*, 5045-5053.

94. Arfmann, H., Timmis, K. N., and Wittich, R. (1997) Mineralization of 4-chlorodibenzofuran by a consortium consisting of *Sphingomonas sp*. Strain RW1 and *Burkholderia sp*. Strain JWS, *Appl. Environ. Microbiol. 63*, 3458-3462.

95. Horinouchi, M., Hayashi, T., Yamamoto, T., and Kudo, T. (2003) A new bacterial steroid degradation gene cluster in *Comamonas testosteroni* TA441 which consists of aromatic-compound degradation genes for seco-steroids and 3-ketosteroid dehydrogenase genes, *Appl. Environ. Microbiol. 69*, 4421-4430.

96. Kendall, S. L., Burgess, P., Balhana, R., Withers, M., Ten Bokum, A., Lott, J. S., Gao, C., Uhia-Castro, I., and Stoker, N. G. (2010) Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2, *Microbiology 156*, 1362-1371.

97. Casabon, I., Crowe, A. M., Liu, J., and Eltis, L. D. (2013) FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria, *Mol. Microbiol. 87*, 269-283.

98. Wheeler, P. R., Brosch, R., Coldham, N. G., Inwald, J. K., Hewinson, R. G., and Gordon, S. V. (2008) Functional analysis of a clonal deletion in an epidemic strain of *Mycobacterium bovis* reveals a role in lipid metabolism, *Microbiology 154*, 3731-3742.

99. Ilina, E. N., Shitikov, E. A., Ikryannikova, L. N., Alekseev, D. G., Kamashev, D. E., Malakhova, M. V., Parfenova, T. V., Afanas'ev, M. V., Ischenko, D. S., Bazaleev, N. A., Smirnova, T. G., Larionova, E. E., Chernousova, L. N., Beletsky, A. V., Mardanov, A. V., Ravin, N. V., Skryabin, K. G., and Govorun, V. M. (2013) Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia, *PloS One 8*, e56577.

100. Santangelo, M. d. l. P., Klepp, L., Nuñez-García, J., Blanco, F. C., Soria, M., García-Pelayo, M. d. C., Bianco, M. V., Cataldi, A. A., Golby, P., Jackson, M., Gordon, S. V., and Bigi, F. (2009) Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*, *Microbiology 155*, 2245-2255.

101. Ansong, C., Ortega, C., Payne, S. H., Haft, D. H., Chauvigne-Hines, L. M., Lewis, M. P., Ollodart, A. R., Purvine, S. O., Shukla, A. K., Fortuin, S., Smith, R. D., Adkins, J. N., Grundner, C., and Wright, A. T. (2013) Identification of widespread adenosine nucleotide binding in *Mycobacterium tuberculosis*, *Chem. Biol. 20*, 123-133.

102. Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M., and Murray, M. B. (2009) Tuberculosis drug resistance mutation database, *PLoS Med. 6*, e1000002.

103. Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization, *Proc. Natl. Acad. Sci. U. S. A. 96*, 12833-12838.

104. Yang, M., Lu, R., Guja, K. E., Wipperman, M. F., St. Clair, J. R., Bonds, A. C., Garcia-Diaz, M., and Sampson, N. S. (2015) Unraveling Cholesterol Catabolism in *Mycobacterium tuberculosis*: ChsE4-ChsE5 $\alpha_2\beta_2$ Acyl-CoA Dehydrogenase Initiates β-Oxidation of 3-Oxo-cholest-4-en-26-oyl CoA, *ACS Infect. Dis. 1*, 110-125.

105. 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.

106. Sih, C. J., Tai, H.-H., Tsong, Y. Y., Lee, S. S., and Coombe, R. G. (1968) Mechanisms of steroid oxidation by microorgansism. XIV. Pathway of cholesterol side-chain degradation, *Biochemistry 7*, 808-818.

107. Reddy, J. K., and Hashimoto, T. (2001) Peroxisomal β-oxidation and peroxisome proliferator-activated receptor α: an adaptive metabolic system, *Annu. Rev. Nutr. 21*, 193-230.

108. Yang, X., Dubnau, E., Smith, I., and Sampson, N. S. (2007) Rv1106c from *Mycobacterium tuberculosis* is a 3β-hydroxysteroid dehydrogenase, *Biochemistry 46*, 9058-9067.

109. 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.

110. Wilbrink, M. H., Petrusma, M., Dijkhuizen, L., and van der Geize, R. (2011) FadD19 of *Rhodococcus rhodochrous* DSM43269, a steroid-coenzyme A ligase essential for degradation of C-24 branched sterol side chains, *Appl. Environ. Microbiol. 77*, 4455-4464.

111. Yang, M., Guja, K. E., Thomas, S. T., Garcia-Diaz, M., and Sampson, N. S. (2014) A Distinct MaoC-like enoyl-CoA hydratase architecture mediates cholesterol catabolism in *Mycobacterium tuberculosis*, *ACS Chem. Biol. 9*, 2632-2645.

112. Li, W., Ge, F., Zhang, Q., Ren, Y., Yuan, J., He, J., Li, W., Chen, G., Zhang, G., Zhuang, Y., and Xu, L. (2014) Identification of gene expression profiles in the actinomycete *Gordonia neofelifaecis* grown with different steroids, *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 57*, 345-353.

113. Tserng, K. Y., Jin, S. J., and Hoppel, C. L. (1991) Spiropentaneacetic acid as a specific inhibitor of medium-chain acyl-CoA dehydrogenase, *Biochemistry 30*, 10755-10760.

114. Slomp, J., G, and Johnson, J. L. (1958) Ozonolysis. II. 1 The effect of pyridine on the ozonolysis of 4, 22-stigmastadien-3-one, *J. Am. Chem. Soc. 80*, 915-921.

115. Kabsch, W. (2010) XDS, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125-132.

116. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 69*, 1204-1214.

117. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 293-302.

118. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1772-1779.

119. Dauter, Z., Dauter, M., and Dodson, E. (2002) Jolly SAD, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 494-506.

120. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, In *Macromolecular Crystallography Protocols* (Doublié, S., Ed.), pp 215-230, Humana Press.

121. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 52*, 30-42.

122. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 1002-1011.

123. Cowtan, K. (2010) Recent developments in classical density modification, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 470-478.

124. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 486-501.

125. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 355-367.

126. Zwart, P., Afonine, P., Grosse-Kunstleve, R., Hung, L.-W., Ioerger, T., McCoy, A., McKee, E., Moriarty, N., Read, R., Sacchettini, J., Sauter, N., Storoni, L., Terwilliger, T., and Adams, P. (2008) Automated Structure Solution with the PHENIX Suite, In *Structural Proteomics* (Kobe, B., Guss, M., and Huber, T., Eds.), pp 419-435, Humana Press.

127. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 12-21.

128. Skinner, J. M., Cowan, M., Buono, R., Nolan, W., Bosshard, H., Robinson, H. H., Heroux, A., Soares, A. S., Schneider, D. K., and Sweet, R. M. (2006) Integrated software for macromolecular crystallography synchrotron beamlines II: revision, robots and a database, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 1340-1347.

129. Orville, A. M., Buono, R., Cowan, M., Héroux, A., Shea-McCarthy, G., Schneider, D. K., Skinner, J. M., Skinner, M. J., Stoner-Ma, D., and Sweet, R. M. (2011) Correlated single-crystal electronic absorption spectroscopy and X-ray crystallography at NSLS beamline X26-C, *J. Synchrotron Radiat. 18*, 358-366.

130. Schmitz, W., and Conzelmann, E. (1997) Stereochemistry of peroxisomal and mitochondrial β-oxidation of α-methylacyl-CoAs, *Eur. J. Biochem. 244*, 434-440.

131. Pascual, C., Meier, J., and Simon, W. (1966) Regel zur Abschatzung der Chemischen Verschiebung von Protonen an einer Doppelbindung, *Helv. Chim. Acta 49*, 164-168.

132. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality, *Science 336*, 1030-1033.

133. Kim, J. J., Wang, M., and Paschke, R. (1993) Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate, *Proc. Natl. Acad. Sci. U. S. A. 90*, 7523-7527.

134. Kozlikova, B., Sebestova, E., Sustr, V., Brezovsky, J., Strnad, O., Daniel, L., Bednar, D., Pavelka, A., Manak, M., Bezdeka, M., Benes, P., Kotry, M., Gora, A., Damborsky, J., and Sochor, J. (2014) CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures, *Bioinformatics 30*, 2684-2685.

135. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*, *Nat. Chem. Biol. 5*, 593-599.

136. Agnihotri, G., and Liu, H.-w. (2003) Enoyl-CoA hydratase. reaction, mechanism, and inhibition., *Bioorg. Med. Chem. 11*, 9-20.

137. Kim, Y.-G., and Maas, S. (2001) Multiple site-directed mutagenesis *in vitro*, In *Vitro Mutagenesis Protocols* (Braman, J., Ed.), pp 29-36, Springer.

138. 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.

139. Tserng, K. Y., Jin, S. J., and Hoppel, C. L. (1991) Spiropentaneacetic acid as a specific inhibitor of medium-chain acyl-CoA dehydrogenase., *Biochemistry 30*, 10755-10760.

140. Sheldrick, G. (2008) A short history of SHELX, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. 64*, 112-122.

141. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7, *Nat. Protocols 3*, 1171-1179.

142. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126-2132.

143. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1948-1954.

144. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res. 35*, W375-W383.

145. Allaire, M., and Yang, L. (2011) Biomolecular solution X-ray scattering at the National Synchrotron Light Source, *J. Synchrotron Radiat. 18*, 41-44.

146. 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.

147. Jiang, L. L., Kurosawa, T., Sato, M., Suzuki, Y., and Hashimoto, T. (1997) Physiological role of D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein, *J. Biochem. 121*, 506-513.

148. Russell, D. W. (2003) The enzymes, regulation, and genetics of bile acid synthesis, *Annu. Rev. Biochem. 72*, 137-174.

149. Dillon, S. C., and Bateman, A. (2004) The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases., *BMC Bioinf. 5*, 109.

150. Koski, M. K., Haapalainen, A. M., Hiltunen, J. K., and Glumoff, T. (2004) A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2, *J. Biol. Chem. 279*, 24666-24672.

151. 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.

152. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38*, W545-W549.

153. 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.

154. Doi, Y., Kitamura, S., and Abe, H. (1995) Microbial synthesis and characterization of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate), *Macromolecules 28*, 4822-4828.

155. 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. Sect. F Struct. Biol. Cryst. Commun. 66*, 1160-1166.

156. Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M. A., Wilson, I. A., Godzik, A., and Wooley, J. (2010) TOPSAN: use of a collaborative environment for annotating, analyzing and disseminating data on JCSG and PSI structures., *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66*, 1143-1147.

157. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., and Yeates, T. O. (1999) Detecting Protein Function and Protein-Protein Interactions from Genome Sequences, *Science 285*, 751-753.

158. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) Protein interaction maps for complete genomes based on gene fusion events., *Nature 402*, 86-90.

159. Sacco, E., Legendre, V., Laval, F., Zerbib, D., Montrozier, H., Eynard, N., Guilhot, C., Daffé, M., and Quémard, A. (2007) Rv3389C from *Mycobacterium tuberculosis*, a member of the (*R*)-specific hydratase/dehydratase family, *Biochim. Biophys. Acta 1774*, 303-311.

160. Castell, A., Johansson, P., Unge, T., Jones, T. A., and Bäckbro, K. (2005) Rv0216, a conserved hypothetical protein from *Mycobacterium tuberculosis* that is essential for bacterial survival during infection, has a double hotdog fold, *Protein Sci. 14*, 1850-1862.

161. Johansson, P., Castell, A., Jones, T. A., and Bäckbro, K. (2006) Structure and function of Rv0130, a conserved hypothetical protein from *Mycobacterium tuberculosis*, *Protein Sci. 15*, 2300-2309.

162. Murdoch, R. W., and Hay, A. G. (2005) Formation of catechols via removal of acid side chains from ibuprofen and related aromatic acids, *Appl. Environ. Microbiol. 71*, 6121-6125.

163. Murdoch, R. W., and Hay, A. G. (2013) Genetic and chemical characterization of ibuprofen degradation by *Sphingomonas* Ibu-2, *Microbiology 159*, 621-632.

164. Kühner, S., Wöhlbrand, L., Fritz, I., Wruck, W., Hultschig, C., Hufnagel, P., Kube, M., Reinhardt, R., and Rabus, R. (2005) Substrate-dependent regulation of anaerobic degradation pathways for toluene and ethylbenzene in a denitrifying bacterium, strain EbN1, *J. Bacteriol. 187*, 1493-1503.

165. Kube, M., Heider, J., Amann, J., Hufnagel, P., Kühner, S., Beck, A., Reinhardt, R., and Rabus, R. (2004) Genes involved in the anaerobic degradation of toluene in a denitrifying bacterium, strain EbN1, *Arch. Microbiol. 181*, 182-194.

166. Leuthner, B., and Heider, J. (2000) Anaerobic toluene catabolism of thauera aromatica: the bbs operon codes for enzymes of β -oxidation of the intermediate benzylsuccinate, *J. Bacteriol. 182*, 272-277.

167. Kannenberg, F., Ellinghaus, P., Assmann, G., and Seedorf, U. (1999) Aberrant oxidation of the cholesterol side chain in bile acid synthesis of sterol carrier protein-2/sterol carrier protein-x knockout mice, *J. Biol. Chem. 274*, 35455-35460.

168. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015) CDD: NCBI's conserved domain database, *Nucleic Acids Res. 43*, D222-D226.

169. Guo, X.-R., Zheng, S.-C., Liu, L., and Feng, Q.-L. (2009) The sterol carrier protein 2/3-oxoacyl-CoA thiolase (SCPx) is involved in cholesterol uptake in the midgut of Spodoptera litura: gene cloning, expression, localization and functional analyses, *BMC Mol. Biol. 10*, 102-102.

170. Ehrt, S., and Schnappinger, D. (2007) *Mycobacterium tuberculosis* virulence: lipids inside and out, *Nat. Med. 13*, 284-285.

171. Galagan, J. E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., Dolganov, G., Glotova, I., Abeel, T., Mahwinney, C., Kennedy, A. D., Allard, R., Brabant, W., Krueger, A., Jaini, S., Honda, B., Yu, W.-H., Hickey, M. J., Zucker, J., Garay, C., Weiner, B., Sisk, P., Stolte, C., Winkler, J. K., Van de Peer, Y., Iazzetti, P., Camacho, D., Dreyfuss, J., Liu, Y., Dorhoi, A., Mollenkopf, H.-J., Drogaris, P., Lamontagne, J., Zhou, Y., Piquenot, J., Park, S. T., Raman, S., Kaufmann, S. H. E., Mohney, R. P., Chelsky, D., Moody, D. B., Sherman, D. R., and Schoolnik, G. K. (2013) The *Mycobacterium tuberculosis* regulatory network and hypoxia, *Nature 499*, 178-183.

172. Van Els, C. A. C. M., Corbiere, V., Smits, K., van Gaans-van den Brink, J. A. M., Poelen, M. C. M., Mascart, F., Meiring, H. D., and Locht, C. (2014) Towards understanding the essence of post-translational modifications for the *Mycobacterium tuberculosis* immunoproteome, *Front. Immunol. 5*, 1-10.

173. Yu, B., Kim, J., Moon, J., Ryu, S., and Pan, J.-G. (2008) The diversity of lysine-acetylated proteins in *Escherichia coli*, *J. Microbiol. Biotechnol. 18*, 1529-1536.

174. Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y., and Zhao, Y. (2011) Identification of lysine succinylation as a new post-translational modification, *Nat. Chem. Biol. 7*, 58-63.

175. Colak, G., Xie, Z., Zhu, A. Y., Dai, L., Lu, Z., Zhang, Y., Wan, X., Chen, Y., Cha, Y. H., Lin, H., Zhao, Y., and Tan, M. (2013) Identification of lysine succinylation substrates and

the succinylation regulatory enzyme CobB in *Escherichia coli*, *Mol. Cell. Proteomics 12*, 3509-3520.

176. Weinert, Brian T., Schölz, C., Wagner, Sebastian A., Iesmantavicius, V., Su, D., Daniel, Jeremy A., and Choudhary, C. (2013) Lysine succinylation ss a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation, *Cell Rep. 4*, 842-851.

177. Yang, M., Wang, Y., Chen, Y., Cheng, Z., Gu, J., Deng, J., Bi, L., Chen, C., Mo, R., Wang, X., and Ge, F. (2015) Succinylome analysis reveals the involvement of lysine succinylation in metabolism in pathogenic *Mycobacterium tuberculosis* H37Rv, *Mol. Cell. Proteomics 14*, 796-811.

178. Ouellet, H., Johnston, J.B., Ortiz de Montellano, P.R. (2011) Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*, *Trends Microbiol. 19*, 530 - 539.

179. Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., Dolganov, G., Efron, B., Butcher, P. D., Nathan, C., and Schoolnik, G. K. (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment, *J. Exp. Med. 198*, 693-704.

180. Du, J., Zhou, Y., Su, X., Yu, J. J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J. H., Choi, B. H., He, B., Chen, W., Zhang, S., Cerione, R. A., Auwerx, J., Hao, Q., and Lin, H. (2011) Sirt5 Is a NAD-dependent protein lysine demalonylase and desuccinylase, *Science 334*, 806-809.

181. Xu, H., Hegde, S. S., and Blanchard, J. S. (2011) Reversible acetylation and inactivation of *Mycobacterium tuberculosis* acetyl-CoA synthetase is dependent on cAMP, *Biochemistry 50*, 5883-5892.

182. Consortium, T. U. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt), *Nucleic Acids Res. 40*, D71-D75.

183. Card, G. L., Peterson, N. A., Smith, C. A., Rupp, B., Schick, B. M., and Baker, E. N. (2005) The crystal structure of Rv1347c, a putative antibiotic resistance protein from *Mycobacterium tuberculosis*, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis, *J. Biol. Chem. 280*, 13978-13986.

184. Errey, J. C., and Blanchard, J. S. (2005) Functional characterization of a novel ArgA from *Mycobacterium tuberculosis*, *J. Bacteriol. 187*, 3039-3044.

185. Hegde, S. S., Javid-Majd, F., and Blanchard, J. S. (2001) Overexpression and mechanistic analysis of chromosomally encoded aminoglycoside 2′ -N-acetyltransferase (AAC(2′)-Ic) from *Mycobacterium tuberculosis*, *J. Biol. Chem 276*, 45876-45881.

186. Nambi, S., Gupta, K., Bhattacharyya, M., Ramakrishnan, P., Ravikumar, V., Siddiqui, N., Thomas, A. T., and Visweswariah, S. S. (2013) Cyclic AMP-dependent protein lysine

acylation in Mycobacteria regulates fatty acid and propionate metabolism, *J. Biol. Chem. 288*, 14114-14124.

187. Vetting, M. W., Roderick, S. L., Yu, M., and Blanchard, J. S. (2003) Crystal structure of mycothiol synthase (Rv0819) from *Mycobacterium tuberculosis* shows structural homology to the GNAT family of N-acetyltransferases, *Protein Science 12*, 1954-1959.

188. Brooke, E. W., Davies, S. G., Mulvaney, A. W., Pompeo, F., Sim, E., and Vickers, R. J. (2003) An approach to identifying novel substrates of bacterial arylamine N-acetyltransferases, *Bioorg. Med. Chem. 11*, 1227-1234.

189. Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) Reassessment of Ellman's reagent, In *Methods Enzymol.* (C.H.W. Hirs, S. N. T., Ed.), pp 49-60, Academic Press.

190. Bhakta, S., Besra, G. S., Upton, A. M., Parish, T., Sholto-Douglas-Vernon, C., Gibson, K. J. C., Knutton, S., Gordon, S., daSilva, R. P., Anderton, M. C., and Sim, E. (2004) Arylamine N-acetyltransferase is required for synthesis of mycolic acids and complex lipids in *Mycobacterium bovis* BCG and represents a novel drug target, *J. Exp. Med. 199*, 1191-1199.

191. Suzuki, H., Ohnishi, Y., and Horinouchi, S. (2007) Arylamine *N*-acetyltransferase responsible for acetylation of 2-aminophenols in *streptomyces griseus*, *J. Bacteriol. 189*, 2155-2159.

192. Xie, L., Liu, W., Li, Q., Chen, S., Xu, M., Huang, Q., Zeng, J., Zhou, M., and Xie, J. (2014) First Succinyl-Proteome Profiling of Extensively Drug-Resistant Mycobacterium tuberculosis Revealed Involvement of Succinylation in Cellular Physiology, *J. Proteome Res. 14*, 107-119.

193. Weinert, Brian T., Schölz, C., Wagner, Sebastian A., Iesmantavicius, V., Su, D., Daniel, Jeremy A., and Choudhary, C. (2013) Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation, *Cell Rep. 4*, 842-851.

194. Hayden, J. D., Brown, L. R., Gunawardena, H. P., Perkowski, E. F., Chen, X., and Braunstein, M. (2013) Reversible acetylation regulates acetate and propionate metabolism in *Mycobacterium smegmatis*, *Microbiology 159*, 1986-1999.

195. Eoh, H., and Rhee, K. Y. (2013) Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A. 110*, 6554-6559.

196. Hartman, T., Weinrick, B., Vilchèze, C., Berney, M., Tufariello, J., Cook, G. M., and Jacobs, W. R., Jr. (2014) Succinate dehydrogenase is the regulator of respiration in *Mycobacterium tuberculosis*, *PLoS Pathog. 10*, e1004510.

197. Ruprecht, A., Maddox, J., Stirling, A. J., Visaggio, N., and Seah, S. Y. K. (2015) Characterization of novel Acyl-CoA Dehydrogenases Involved in Bacterial Steroid Degradation, *J. Bacteriol. 197*, 1360-1367.

198. Xie, J., Zhou, F., Xu, G., Mai, G., Hu, J., Wang, G., and Li, F. (2014) Genome-wide screening of pathogenicity islands in *Mycobacterium tuberculosis* based on the genomic barcode visualization, *Mol. Biol. Rep. 41*, 5883-5889.