

MYCOBACTERIUM TUBERCULOSIS METHYLCITRATE CYCLE REGULATION  
AND  
STUDY OF NON-COVALENT NATURAL PRODUCT INHIBITORS OF  
PROTEASOME

A Dissertation

by

SU TANG

Submitted to the Office of Graduate and Professional Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James Sacchettini
Committee Members,	Mary Bryk
	Stephen Safe
	Steve Maxwell
Head of Department,	Dorothy Shippen

December 2018

Major Subject: Biochemistry

Copyright 2018 Su Tang

## ABSTRACT

The methylcitrate cycle and glyoxylate bypass are two critical pathways of the central carbon metabolism of the human pathogen *Mycobacterium tuberculosis* (*Mtb*). Transcription of the key enzymes in these pathways is controlled by two paralogous local regulators, MtPrpR and MtRamB. No structure is available for these regulators or their homologs and the regulatory mechanisms are not well understood. In this study, we present the crystal structures of different forms of MtPrpR including two domains of unknown function. Our studies showed that MtPrpR is an iron-sulfur protein that binds to coenzyme A (CoA) and its derivatives. A single amino acid polymorphism in the CoA binding pocket enables the protein homologs to distinguish among different CoA derivatives. Our study suggested the methylcitrate cycle and glyoxylate bypass are regulated through a combination of feedforward and feedback mechanisms based on the CoA pool composition.

The ubiquitin-proteasome pathway plays a pivotal role in cellular protein breakdown. Dysregulation of proteolysis leads to various human diseases including cancer. Inhibiting the proteasome is a promising approach in cancer treatment. Several inhibitors targeting the 20S proteasome have been tested on humans and approved by the FDA as antitumor drugs. In this study, we characterized a fungal derived natural product compound with a  $[M+H]^+$   $m/z$  value of 721.33 (designated Mz721). Mz721 preferentially targets the trypsin- and chymotrypsin-like sites of the yeast 20S proteasome. The  $IC_{50}$  value for the trypsin-like site in particular is below five nanomolar.

The crystal structure of the yeast 20S proteasome in complex with Mz721 revealed that the inhibitor is structurally similar to the previously reported TMC-95 series natural products. It binds to all active sites in a non-covalent mode. Being cytotoxic against cancer cells, Mz721 is much less toxic than bortezomib, a first-in-class agent for multiple myeloma treatment, on the human dermal fibroblasts. As a result, Mz721 can act as a lead molecule to develop drugs that target the trypsin-like site of the 20S proteasome.

## DEDICATION

To my family

## ACKNOWLEDGEMENTS

I would like to thank my committee chair and advisor, Prof. Sacchettini, for providing us with the state-of-the-art instruments so my multi-disciplinary projects could be conducted. I would like to thank my committee member, Prof. Bryk, for teaching me yeast genetics when I rotated in her laboratory and for her encouragement throughout my Ph.D. career. I would also like to thank Prof. Safe and Prof. Maxwell, for their guidance and support during the course of this research.

Thanks also go to my friends and colleagues for sharing their ideas, knowledge and experiences. Special thanks go to Dr. Dwight Baker for sharing with me his rich experience in the field of natural products, and to Prof. Thomas Meek for his mentoring on covalent inhibitors and enzymology.

Finally, but most importantly, I thank my mother and father for their support and encouragement and to my wife for her patience and love.

## CONTRIBUTORS AND FUNDING SOURCES

### **Contributors**

This work was supervised by a dissertation committee consisting of Prof. Sacchettini, Prof. Bryk and Prof. Safe of the Department of the Biochemistry and Biophysics, and Prof. Maxwell of the College of Medicine.

In Chapter II, the transcriptional experiments were conducted by Nathan Hicks in Prof. Sarah Fortune's laboratory at Harvard T.H. Chan School of Public Health. Some of the cloning and the iron-sulfur cluster identification work were conducted with the assistance of Dr. Yu-Shan Cheng, then a member of the laboratory. The high-resolution mass spectrometry experiments in both Chapter II and III were performed by Andres Silva in the laboratory. In Chapter III, the fungal materials are assets of Dow AgroSciences LLC. The yeast strain was provided by Prof. Bryk. The fungal culture was conducted by Dr. Dwight Baker in the laboratory.

All other work conducted for the dissertation was completed by the student independently.

### **Funding Sources**

The work in Chapter II was made possible in part by National Institutes of Health (NIH) under Grant Number P01AI095208. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

## NOMENCLATURE

1,4-PDC-CoA	3-oxo-1,4-pregnadiene-20-carboxyl-CoA
17-OHPC-CoA	17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA
2-MC	2-methylcitrate
2-MIC	2-methylisocitrate
3,4-DHSA	3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione
3-HSA	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione
3OChA	3-oxo-4-cholestenoic acid
3-OCO-CoA	3-oxocholest-4-en-24-oyl-CoA
3-OCS-CoA	3-oxocholest-4-en-26-oyl-CoA
3-OPC-CoA	3-oxo-4-pregnene-20-carboxyl-CoA
3-OPDC-CoA	3-OPC-CoA to 3-oxo-4,17-pregnadiene-20-carboxyl-CoA
4,9-DSHA	4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid
9-OHAD	9-hydroxy-4-androstene-3,17-dione
9-OHADD	9-hydroxy-1,4-androstadiene-3,17-dione
ABC	ATP-binding cassette
AD	4-androstendione
ADD	1,4-androstadiene-3,17-dione
Atc	anhydrous tetracycline
CoA	Coenzyme A
CT-L	chymotrypsin-like

Da	dalton
DHP	2,3-dihydroxyl-biphenyl
EMSA	electrophoretic mobility shift assay
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FBPase	fructose-1,6-bisphosphatase
GAF	cGMP phosphodiesterase/adenylyl cyclase/Eh1A
HDF	human dermal fibroblast
HHd	2-hydroxy-hexa-2,4-dienoic acid
HIP	3 $\alpha$ -H-4 $\alpha$ (3'-propanoate)-7 $\beta$ -methylhexahydro-1,5-indanedione
HOH	4-hydroxy-2-oxohexanoate
HOP	4-hydroxy-2-oxopentanoate
HOPDA	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid
HPLC	high performance liquid chromatography
HSD	3 $\beta$ -hydroxy-steroid dehydrogenase
HTH	helix-turn-helix
IC <sub>50</sub>	half maximal inhibitory concentration
Icl	isocitrate lyase
igr	intracellular growth
INF- $\gamma$	interferon- $\gamma$
kb	kilobase
kDa	kilodaltons



KG	$\alpha$ -ketoglutarate
LCMS	liquid chromatography mass spectrometry
MAS	mycocerosic acid synthase
MBL	methyl-branched lipid
MCC	methylcitrate cycle
MCD	methylcitrate dehydratase
mce	mammalian cell entry
MCL	methylcitrate lyase
MCM	methylmalonyl-CoA mutase
MCS	methylcitrate synthase
MM	multiple myeloma
MMCE	methylmalonyl-CoA epimerase
MMP	methylmalonyl pathway
MS	mass spectrometry
mtb	Mycobacterium tuberculosis
NAD	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NBD	nucleotide-binding domain
PAS	Per-Arnt-Sim
PCC	propionyl-CoA carboxylase
PDIM	phthiocerol dimycocerosate
PEG	polyethylene glycol

PGPH	post-glutamyl-peptide hydrolase
PKS	polyketide synthase
PrpC	methylcitrate synthase
PrpD	methylcitrate dehydratase
PrpR	propionate regulator
RamB	regulator of acetate metabolism B
SAM	S-adenosylmethionine
SAR	structure-activity relationship
SEC	size exclusion chromatography
SL-1	sulfolipid-1
SNP	single nucleotide polymorphism
TDM	trehalose dimycolate
T-L	trypsin-like
TMM	trehalose monomycolate
TraSH	transposon site hybridization

## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION .....	iv
ACKNOWLEDGEMENTS .....	v
CONTRIBUTORS AND FUNDING SOURCES.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS .....	xi
LIST OF FIGURES.....	xv
LIST OF TABLES .....	xviii
CHAPTER I INTRODUCTION: CHOLESTEROL METABOLISM IN <i>MYCOBACTERIUM TUBERCULOSIS</i> .....	1
I.1 Cholesterol utilization by microorganisms.....	1
I.1.1 Role of cholesterol in microorganism infection .....	1
I.1.2 Microorganisms using cholesterol as a carbon source .....	2
I.1.3 The in vivo carbon source conundrum of <i>Mtb</i> .....	3
I.2 Cholesterol uptake by <i>Mtb</i> .....	4
I.3 Regulation of cholesterol uptake and utilization.....	12
I.3.1 The <i>kstR</i> regulon .....	12
I.3.2 The <i>kstR2</i> regulon .....	15
I.4 Cholesterol degradation pathway in <i>Mtb</i> .....	16
I.4.1 $\beta$ -oxidation of cholesterol aliphatic side-chain.....	16
I.4.1.1 HSD oxidizes the 3-OH and isomerizes cholesterol.....	16
I.4.1.2 Cytochrome P450 enzymes activate the aliphatic side chain .....	17
I.4.1.3 $\beta$ -oxidation of the side chain.....	19
I.4.2 Degradation of the A/B-Ring .....	23
I.4.2.1 <i>KstD</i> catalyzes the 1,2-desaturation.....	23
I.4.2.2 <i>KSH</i> catalyzes the 9 $\alpha$ -hydroxylation .....	24
I.4.2.3 <i>HsaA-HsaB</i> catalyzes the 4-hydroxylation.....	25
I.4.2.4 <i>HsaC</i> opens A-ring.....	26
I.4.2.5 <i>HsaD</i> cleaves HHD off 4,9-DSHA .....	30

I.4.2.6 HsaEFG degrades HHD .....	30
I.4.3 Degradation of the C/D-Ring .....	33
I.4.3.1 FadD3 initiates the degradation of C/D-ring .....	33
I.4.3.2 Degradation of the C/D-ring yields acetyl-, propionyl- and succinyl-CoA .....	34
I.5 Metabolism of the unique product propionyl-CoA in <i>Mtb</i> .....	36
I.5.1 Methylmalonyl pathway (MMP) .....	36
I.5.1.1 <i>Mtb</i> has a functional MMP .....	36
I.5.1.2 Vitamin B <sub>12</sub> acquisition in <i>Mtb</i> .....	38
I.5.1.3 <i>Mtb</i> may circumvent B <sub>12</sub> -dependent metabolic steps .....	40
I.5.2 Methyl-branched lipid synthesis .....	41
I.5.2.1 <i>Mtb</i> cell wall and the lipid-rich outer membrane .....	41
I.5.2.2 Stereochemistry in MBL biosynthesis and virulence of <i>Mtb</i> .....	44
I.5.3 The methylcitrate cycle .....	47
I.5.3.1 Enzymes of MCC in general and in <i>Mtb</i> .....	47
I.5.3.2 Toxicity of propionyl-CoA and related metabolic intermediates .....	48
I.5.3.3 Role of MCC <i>in vivo</i> .....	51

CHAPTER II SHORT-CHAIN FATTY ACYL COENZYME A DERIVATIVES REGULATE MYCOBACTERIUM TUBERCULOSIS METHYLCITRATE CYCLE AND GLYOXYLATE SHUNT .....

II.1 Introduction .....	53
II.2 Materials and Methods .....	57
II.2.1 Bacterial strains, media, and growth conditions .....	57
II.2.2 Cloning, protein expression, and purification .....	57
II.2.3 Crystallization .....	63
II.2.4 Data collection and structure determination .....	64
II.2.5 Iron-sulfur cluster identification and redox characterisation .....	66
II.2.6 Coenzyme A derivative isolation and identification .....	67
II.2.7 Protein sequence alignment and analysis .....	68
II.2.8 Transcription level quantification .....	69
II.2.9 Recognition DNA preparation and DNA-binding assay .....	70
II.3 Results .....	70
II.3.1 Identification of an iron-sulfur cluster in MtPrpR .....	70
II.3.2 Crystal structure of MtPrpR <sub>81-486</sub> .....	74
II.3.3 Identification of Coenzyme A in MtPrpR .....	82
II.3.4 The iron-sulfur cluster is critical for CoA binding and transcription activation .....	88
II.3.5 CoA derivatives control the MCC pathway activation .....	91
II.3.6 MtRamB may recognise succinyl-CoA .....	94
II.3.7 PrpR and RamB cross-regulate icl1 but not MCC .....	97

II.3.8 MtPrpR may require propionyl-CoA to specifically bind to the promoter region of prp operon .....	99
II.4 Discussion .....	103
II.4.1 Regulatory model of MCC and glyoxylate shunt by MtPrpR and MtRamB .....	103
II.4.2 The role of iron-sulfur cluster .....	107
II.4.3 The protein folding of IssR in other proteins .....	108
 CHAPTER III NATURAL PRODUCT MZ721 IS A POTENT PROTEASOME INHIBITOR.....	 109
III.1 Introduction.....	109
III.1.1 The Ubiquitin-Proteasome Pathway .....	109
III.1.2 Structure and Function of Proteasome .....	111
III.1.3 UPS Functions in Tumorigenesis.....	113
III.1.4 Proteasome Inhibitors .....	115
III.2 Materials and methods.....	124
III.2.1 Proteasome inhibitors and fungal strain acquisition .....	124
III.2.2 Taxonomy, strain identification .....	125
III.2.3 Fermentation .....	128
III.2.4 Natural product extraction and purification .....	128
III.2.5 HPLC and mass spectrometry.....	129
III.2.6 Yeast and human cell line .....	129
III.2.7 Yeast 20S proteasome Cloning, Protein Purification and Crystallography .....	130
III.2.8 X-ray diffraction data collection and processing .....	131
III.2.9 Yeast Proteasome Bioactivity Assay .....	131
III.2.10 Cell Culture and Treatments .....	132
III.3 Results.....	133
III.3.1 Identification and taxonomy study of the fungal strain F01804 .....	133
III.3.2 Fermentation and extraction of the novel secondary metabolites.....	137
III.3.3 Mz721 preferentially inhibits the Trypsin-like activity of the yeast 20S proteasome .....	145
III.3.4 Mz721 binds to the active sites of the 20S proteasome .....	148
III.3.5 Mz721 showed cytotoxicity on several cancer cell lines.....	158
III.4 Discussion.....	161
 CHAPTER IV CONCLUSIONS AND FUTURE DIRECTIONS .....	 164
IV.1 MtPrpR and the MCC transcriptional regulation .....	164
IV.1.1 Propionyl-CoA bound MtPrpR structure.....	164
IV.1.2 MtPrpR tetramer mystery .....	165
IV.1.3 Signaling molecules that may induce CoA derivative exchange.....	166
IV.2 Mz721 purification, identification and characterization .....	166

IV.2.1 Genome sequencing and mining of the biosynthetic pathways of Mz721 and analogs .....	167
IV.2.2 Mz721 extraction and purification.....	169
IV.2.3 Mz721 accurate structure identification .....	170
REFERENCES .....	171

## LIST OF FIGURES

	Page
Figure 1. Structure and numbering of cholesterol.....	6
Figure 2. The signal molecules of <i>kstR</i> and <i>kstR2</i> regulons.....	15
Figure 3. Cholesterol aliphatic side chain activation .....	19
Figure 4. Degradation of aliphatic side-chain .....	23
Figure 5. Activation of A/B-ring by desaturation and hydroxylation.....	25
Figure 6. Degradation of A/B-ring (Part 1).....	28
Figure 7. Suggested catalytic cycle for extradiol dioxygenase, adapted from (85) with permission.....	29
Figure 8. Degradation of A/B-ring (Part 2).....	32
Figure 9. Degradation of C/D-ring, adapted and modified from (71).....	35
Figure 10. Overview of the propionyl-CoA assimilation pathways.....	36
Figure 11. Vitamin B12 biosynthesis.....	39
Figure 12. Schematic of <i>Mtb</i> cell wall and biosyntheses of MBLs, a portion of the figure is adapted from (113) with permission.....	43
Figure 13. Consensus enzyme catalytic mechanisms of the methylcitrate cycle.....	48
Figure 14. Genome organization of <i>ramB-icl1</i> regulon and the <i>prpR-prpDC</i> regulon. ....	55
Figure 15. Domain organization of MtPrpR. ....	56
Figure 16. Expression and purification of MtPrpR constructs.....	71
Figure 17. Characterization of the iron-sulfur cluster of MtPrpR <sub>81-486</sub> .....	73
Figure 18. Structure of MtPrpR <sub>81-486</sub> and analytical SEC of MtPrpR <sub>81-486</sub> in solution .....	76
Figure 19. Mass spectrometry of the His-tagged MtPrpR <sub>81-486</sub> crystals.....	77

Figure 20. Iron-sulfur cluster binding mode of MtPrpR <sub>81-486</sub> .....	78
Figure 21. Sequence alignment of MtPrpR and homologs .....	80
Figure 22. Structure comparison between MtPrpR <sub>81-486</sub> and <i>D. deserti</i> IrrE .....	81
Figure 23. Identification of CoA in MtPrpR <sub>81-486</sub> .....	83
Figure 24. Structure and ligand binding mode comparisons between canonical GAF and MtPrpR GAF-like domains.....	85
Figure 25. Interactions between MtPrpR <sub>81-486</sub> and CoA .....	86
Figure 26. Relationship between iron-sulfur cluster and CoA binding.....	87
Figure 27. Transcription levels of <i>prpD</i> and <i>prpR</i> in MtPrpR variants under acetate or propionate conditions .....	89
Figure 28. Structural characterization of MtPrpR <sub>155-440</sub> .....	90
Figure 29. MtPrpR <sub>81-486</sub> with propionyl-CoA computationally modeled into the CoA binding cavity.....	92
Figure 30. Study of the role of Phe155 in CoA derivative binding .....	94
Figure 31. Characterization of MtPrpR <sub>81-486</sub> _F155H as a mimic of MtRamB.....	96
Figure 32. Cross-regulation of MCC and glyoxylate shunt by <i>prpR</i> and <i>ramB</i> .....	98
Figure 33. Binding of MtPrpR_V9 to different forms of 30 bp DNA .....	101
Figure 34. Binding of wildtype and F155H variant MtPrpR_V9 to the 260 bp promoter region of the <i>prp</i> operon.....	102
Figure 35. Binding of MtPrpR_V9 to physiologically irrelevant DNA fragments of the <i>Mtb embC</i> gene.....	102
Figure 36. The effects of 2-methylcitrate and cAMP on the DNA affinity of MtPrpR_V9 .....	103
Figure 37. Schematic of MtPrpR/MtRamB regulation mediated by short chain fatty acyl CoA molecules.....	105
Figure 38. Schematic of the ubiquitin-proteasome system .....	111
Figure 39. Mechanism of proteolysis catalyzed by proteasome Thr 1.....	113



Figure 40. Major categories of covalent inhibitors and the corresponding adducts .....	116
Figure 41. Examples of the covalent proteasome inhibitors .....	117
Figure 42. Structures of TMC-95A—D, adapted from (269) with permission.....	120
Figure 43. Taxonomy study of fungal strain F01804 (Part 1).....	135
Figure 44. Taxonomy study of fungal strain F01804 (Part 2).....	136
Figure 45. Taxonomy study of fungal strain F01804 (Part 3).....	137
Figure 46. HPLC-MS of the Mz721 standard fraction.....	139
Figure 47. HPLC-MS of extraction with different solvents.....	140
Figure 48. Inhibition profile of the preparative HPLC fractions.....	142
Figure 49. LC-MS of Fraction 33 in Fig. 48 .....	143
Figure 50. LC-MS of Fraction 42 in Fig. 48 .....	144
Figure 51. Yeast 20S proteasome inhibition test of Mz721 .....	146
Figure 52. Yeast 20S proteasome inhibition profile of Mz705-mix .....	148
Figure 53. T-L site bound by Mz721 .....	151
Figure 54. The comparison of TMC-95A with Mz721 .....	152
Figure 55. Mz721 forms a host of hydrogen bonds with the proteasome T-L active site .....	154
Figure 56. Binding mode of Mz721 in CT-L and PGPH sites.....	156
Figure 57. Electron density of Mz705-mix overlaid with Mz721 coordinates at the T-L site.....	157
Figure 58. Effects of Mz721 on tumor cell lines and human dermal fibroblasts.....	159
Figure 59. Sequence alignment of selected subunits between the yeast and human proteasome .....	160
Figure 60. Retro-biosynthesis analysis of Mz721 .....	169

## LIST OF TABLES

	Page
Table 1. Summary of cholesterol utilization related genes .....	7
Table 2. Primers used in this study .....	59
Table 3. Crystallographic data collection and refinement statistics .....	75
Table 4. Structure-activity relationship of TMC-95A based on previous literatures .....	121
Table 5. Structure-activity relationship of a backbone modified TMC-95A analog .....	123
Table 6. Fungal 28S universal primers used in this research .....	127
Table 7. Inhibitory activities of Mz721 and the TMC-95 series against CT-L, T-L and PGPH active sites of the 20S proteasome .....	147
Table 8. Crystallography data collection and refinement statistics .....	149

## CHAPTER I

### INTRODUCTION: CHOLESTEROL METABOLISM IN *MYCOBACTERIUM*

#### *TUBERCULOSIS*

### **I.1 Cholesterol utilization by microorganisms**

#### *I.1.1 Role of cholesterol in microorganism infection*

Composed of a rigid tetracyclic hydrocarbon core and an eight-carbon side-chain, cholesterol is very hydrophobic in general (Fig. 1). The solubility of cholesterol in water is only  $1.8 \pm 0.8 \mu\text{g/mL}$ , and the molecule tends to self-associate (1). But the  $3\beta$ -hydroxyl group makes cholesterol slightly amphiphilic, and, thereby, a crucial component of the cell membrane. More than 90% of cellular cholesterol is located in the plasma membrane (2), and cholesterol accounts for 10 to 45 molar percentage (mol%) of the plasma membrane (3). The heterogeneous distribution of the plasma membrane results in cholesterol-rich subdomains known as “lipid rafts”, which are actively involved in multiple cellular signaling processes. However, lipid rafts are also utilized by some intracellular microorganisms, including *Escherichia coli*, *Chlamydia trachomatis*, *Helicobacter pylori*, *Plasmodium falciparum* (2). The internalization of these microorganisms involves attachment to the lipid rafts, although cholesterol *per se* may not be absolutely required (2).

Nevertheless, cholesterol is essential in mycobacteria internalization by macrophages, a process known as the phagocytosis. Cholesterol is accumulated at the mycobacterial entry site of macrophages. Depletion of the plasma membrane cholesterol

inhibits the uptake of *Mycobacterium tuberculosis* (*Mtb*) and *M. bovis* Bacille Calmette-Guérin (BCG) by macrophages (4). In addition to phagocytosis, cholesterol is also required for *Mtb* persistence. For a common intracellular pathogen in a macrophage, the fusion of a lysosome to a pathogen-containing phagosome can lead to the digestion of the pathogen and disinfection. However, *Mtb* can recruit a membrane attached protein TACO (tryptophane aspartate-containing coat protein) to the surface of a phagosome, which prevents the fusion between a lysosome and a phagosome (5). The attachment of TACO to the phagosome membrane is also cholesterol-dependent (4).

#### I.1.2 *Microorganisms using cholesterol as a carbon source*

Cholesterol utilization by microorganisms was first documented in 1913 when Söhngen showed that soil microorganisms could grow on cholesterol medium (6). The degradation of cholesterol by actinomycetes, including multiple species of fast-growing mycobacteria, was reported in the 1940s by Tak (7) and Turfitt (8, 9). Several side chain-degraded and ring-cleaved metabolic intermediates were detected in *Proactinomyces erythropilis*, *Pseudomonas spp.*, and *Nocardia restrictus* (10-13).

*M. smegmatis*, a fast-growing non-pathogenic saprophytic actinobacterium, is also able to degrade cholesterol (14). Slow-growing mycobacteria such as *Mtb* and *M. bovis* BCG were observed to uptake and accumulate cholesterol, but the degradation intermediates were claimed undetectable (14). Therefore, for a long time, *Mtb* had been considered not being able to use cholesterol as a carbon source.

### I.1.3 *The in vivo carbon source conundrum of Mtb*

After being inhaled through aerosol droplets, *Mtb* will pass through the human airway and reach the alveoli where the bacilli get phagocytosed by macrophages. *Mtb* rapidly proliferate in the macrophages until the onset of acquired immune response. The infected macrophages are surrounded by numerous immune cells including monocytes and lymphocytes. Together they form a tissue known as a granuloma (15). Further development of cellular immunity leads to the necrosis of *Mtb* infected macrophages and the formation of a caseum center of a granuloma where *Mtb* can persist for decades with little or no replication. The caseum center is hypoxic, low in pH, iron limited, nutrient depleted, but rich in lipids. (16-21). *Mtb* has successfully evolved many features in order to adapt this habitat. A significant feature of this adaptation is a carbon source switch. (22-25).

First observed by Segal and Bloch in 1956, *Mtb* freshly isolated from infected lungs behaved quite differently from that grown on culture media (22). The former had a much lower metabolic activity as measured by its weak ability to reduce the redox indicator tetrazolium chloride. Moreover, the respiratory rate of the former increased only in response to a few substrates including fatty acids but not glucose. These observations led to a hypothesis that *Mtb* shifted its carbon source from carbohydrate-based to lipid-based nutrients during infection. The hypothesis was supported by *Mtb* whole genome sequencing and preliminary annotation in 1998. Its genome encodes more than 250 lipid metabolic genes (26). Fatty acids were initially regarded as the sole lipid carbon source *in vivo*. This was because the glyoxylate shunt, which assimilates the fatty

acid degradation product acetyl-CoA (27, 28), was demonstrated essential for *Mtb* growth *in vivo*. The glyoxylate shunt enzyme isocitrate lyase 1 (Icl1) was not only required for bacterial persistence (23, 27, 29) but also upregulated during infection (30). Although very profound, this discovery could not explain why *Mtb* needs so many lipid metabolic genes. Moreover, many of these genes did not have functionally characterized orthologs.

During a study of *Rhodococcus jostii*, a cluster of 51 genes was linked to cholesterol utilization (31) (see Section I.2). Homologs of the genes were also found clustered in the *Mtb* genome (31). The gene cluster was renamed “Cho region” for cholesterol utilization. This breakthrough shed light on a large number of the previously uncharacterized lipid metabolic genes and paved the way for the subsequent studies on cholesterol metabolism by *Mtb*.

## **I.2 Cholesterol uptake by *Mtb***

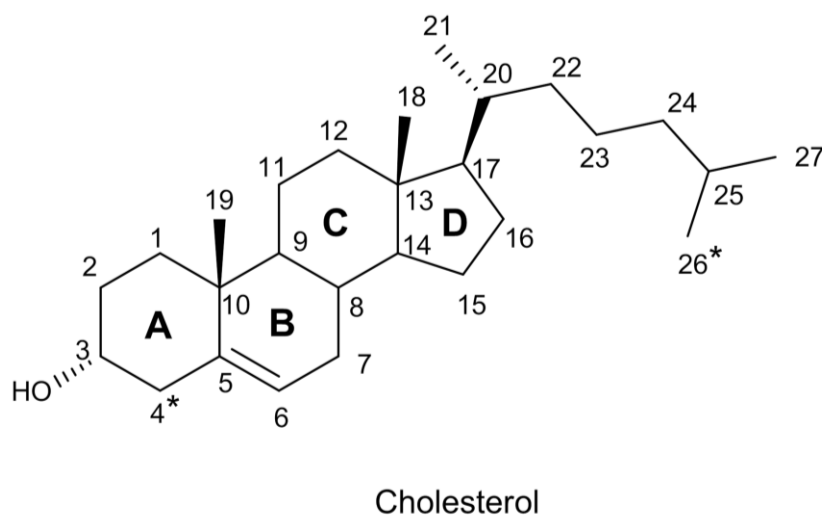
The discovery of the cholesterol transport system was an outcome of genome-wide identification of *in vivo* essential genes in *Mtb*. Sassetti et al. generated a transposon insertion mutant library of *Mtb* (32). By comparing the *in vitro* and *in vivo* growth of the mutant library (32), they identified genes that were only required for *Mtb* survival during infection in mice (29). Apart from a panel of functionally characterized genes, they discovered two gene operons (*rv0166-0178*, and *rv3494c-3501c*) that “have no obvious homologs outside of mycobacteria and closely related species” (29). These gene operons had been previously demonstrated to allow *Mtb* or even nonpathogenic *E.*

*coli*, when carrying these genes, to enter mammalian cells. Thereby, the gene operons were renamed the mammalian cell entry (*mce*) family (33). The two gene operons, designated *mce1* and *mce4* (32), were required during different stages of infection. Specifically, *mce1* was used during initial infection, while *mce4* was essential for maintaining a chronic infection (29, 34).

A bioinformatics study suggested that the *mce* operons encoded multisubunit transport systems. The YrbEA and YrbEB subunits in both Mce systems showed sequence homology to the transmembrane permease subunits of ATP-binding cassette (ABC) transporters (29, 35). However, both *mce* operons lack a nucleotide-binding domain (NBD) subunit which is an essential component of an ABC transporter. Genetic interaction mapping suggested *rv0655* outside the *mce* operon as the putative NBD subunit of both *mce* operons, and the gene was renamed *mceG* (35). Double deletion of *mceG* and *mce4* (or *mceG* and *mce1*) had less than an additive effect because they participated in the same pathway. MceG is homologous to the NBD subunits of ABC transporters (29, 35). Besides *mceG*, a host of putative lipid metabolic genes also appeared in the *mce* genetic interaction map (29, 35), indicating that Mce systems might be lipid transporters.

A breakthrough occurred in 2006. Differential gene expression in response to carbon sources was carried out on a steroid-degrading soil actinomycete *R. jostii* RHA1 strain. Van der Geize et al. (31) showed that 51 genes within a 235-kb region of the rhodococcal genome were specifically expressed under cholesterol exposure. Mutagenesis and biochemical assays confirmed that many of these genes participated in

cholesterol catabolism, specifically, the degradation of the side chain and the A/B-ring (Fig. 1). Twenty-eight out of 51 genes were functionally annotated. Moreover, these 51 genes were mapped onto a cluster of 83 genes in the *Mtb* genome (*rv3492c-rv3574*), which is the aforementioned Cho region (many of the gene functions will be covered in Section I.4). The *mce4* operon, which contains eight genes, belongs to the Cho region. This significant discovery elucidated the function of the *mce4* operon and additionally suggested that *Mtb* might catabolize cholesterol. In agreement with this discovery, many of these genes in *Mtb* had been previously shown to be essential (36) and upregulated (30) during the bacterial growth in INF- $\gamma$  activated macrophages (A summary of cholesterol utilization related genes are listed in Table 1. The data are exclusively from those unbiased high throughput experiments without a particular target of interest).



**Figure 1. Structure and numbering of cholesterol**  
 Cholesterol contains A/B/C/D rings and an aliphatic side-chain including C21-C27 carbons. The  $^{14}\text{C}$ -isotope-labeled carbons are indicated with asterisks.



**Table 1. Summary of cholesterol utilization related genes**

<b>Locus</b>	<b>Gene name</b>	<b>Annotation</b>	<b>Essential for growth on cholesterol</b>	<b>Essential for growth in macrophage</b>	<b>Essential for growth in mice</b>	<b>Essential in vitro (glycerol)</b>	<b>Upregulated in macrophage</b>	<b>Upregulated in cholesterol</b>
<i>rv0467</i>	<i>icl1</i>	isocitrate lyase					y	n/a
<i>rv0655</i>	<i>mceG/mkl</i>	NBD of Mce transporters	y	y	y			n/a
<i>rv1106c</i>	<i>3<math>\beta</math>-hsd</i>	3 $\beta$ -hydroxysteroid dehydrogenase						n/a
<i>rv1130</i>	<i>prpD</i>	methylcitrate dehydratase	y				y	n/a
<i>rv1131</i>	<i>prpC</i>	methylcitrate synthase	y				y	n/a
<i>rv1143</i>	<i>mcr</i>	methyl acyl-CoA racemase						n/a
<i>rv1322A</i>		putative Methylmalonyl-CoA epimerase						n/a
<i>rv1410c</i>		P55 efflux pump		y				n/a
<i>rv1492</i>	<i>mutA</i>	methylmalonyl-CoA mutase $\beta$ -subunit						n/a
<i>rv1493</i>	<i>mutB</i>	methylmalonyl-CoA mutase $\alpha$ -subunit						n/a
<i>rv3280</i>	<i>accD5</i>	acyl-CoA carboxylase $\beta$						n/a
<i>rv3281</i>	<i>accE5</i>	acyl-CoA carboxylase $\epsilon$				y		n/a
<i>rv3285</i>	<i>accA3</i>	acyl-CoA carboxylase $\alpha$				y		n/a
<i>rv3409</i>	<i>choD</i>	cholesterol oxidase						n/a
<i>rv3494c</i>	<i>mceF</i>	Mce4	y					
<i>rv3495c</i>	<i>lprN</i>	Mce4	y		y			
<i>rv3496c</i>	<i>mce4D</i>	Mce4	y					
<i>rv3497c</i>	<i>mce4C</i>	Mce4	y		y			
<i>rv3498c</i>	<i>mce4B</i>	Mce4	y					
<i>rv3499c</i>	<i>mce4A</i>	Mce4	y		y			

**Table 1. Continued**

<b>Locus</b>	<b>Gene name</b>	<b>Annotation</b>	<b>essential for growth on cholesterol</b>	<b>Essential for growth in macrophage</b>	<b>Essential for growth in mice</b>	<b>Essential in vitro (glycerol)</b>	<b>Upregulated in macrophage</b>	<b>Upregulated in cholesterol</b>
<i>rv3500c</i>	<i>yrbe4B</i>	Mce4	y					
<i>rv3501c</i>	<i>yrbe4A</i>	Mce4	y		y			
<i>rv3502c</i>	<i>hsd4A</i>	dehydrogenase	y		y			y
<i>rv3503c</i>	<i>fdxD</i>	ferredoxin						y
<i>rv3504</i>	<i>chsE4</i>	acyl-CoA dehydrogenase						y
<i>rv3505</i>	<i>chsE5</i>	acyl-CoA dehydrogenase					y	y
<i>rv3506</i>	<i>fadD17</i>	acyl-CoA synthetase						
<i>rv3513c</i>	<i>fadD18</i>	probable acyl-CoA ligase						
<i>rv3515c</i>	<i>fadD19</i>	acyl-CoA synthetase	y				y	y
<i>rv3516</i>	<i>echA19</i>	Enoyl-CoA hydratase					y	y
<i>rv3518c</i>	<i>cyp142</i>	cytochrome P450						y
<i>rv3520c</i>								y
<i>rv3521</i>								y
<i>rv3522</i>	<i>ltp4</i>	ketoacyl-CoA thiolase						y
<i>rv3523</i>	<i>ltp3</i>	ketoacyl-CoA thiolase			y			y
<i>rv3526</i>	<i>kshA</i>	monooxygenase	y				y	y
<i>rv3534c</i>	<i>hsaF</i>	aldolase	y		y		y	y
<i>rv3535c</i>	<i>hsaG</i>	aldehyde dehydrogenase					y	y
<i>rv3536c</i>	<i>hsaE</i>	HHD hydratase	y				y	y
<i>rv3537</i>	<i>kstD</i>	dehydrogenase	y	y			y	y
<i>rv3538</i>	<i>hsd4B</i>	2-enoyl acyl-CoA hydratase				y	y	y

**Table 1. Continued**

<b>Locus</b>	<b>Gene name</b>	<b>Annotation</b>	<b>essential for growth on cholesterol</b>	<b>Essential for growth in macrophage</b>	<b>Essential for growth in mice</b>	<b>Essential in vitro (glycerol)</b>	<b>Upregulated in macrophage</b>	<b>Upregulated in cholesterol</b>
<i>rv3540c</i>	<i>ltp2</i>	$\beta$ -hydroxyl-CoA aldolase	y		y		y	y
<i>rv3541c</i>	<i>chsH1</i>	acyl-CoA hydratase		y	y		y	y
<i>rv3542c</i>	<i>chsH2</i>	acyl-CoA hydratase	y	y	y		y	y
<i>rv3543c</i>	<i>chsE2</i>	acyl-CoA dehydrogenase	y				y	y
<i>rv3544c</i>	<i>chsE1</i>	acyl-CoA dehydrogenase	y	y	y		y	y
<i>rv3545c</i>	<i>cyp125</i>	cytochrome P450	y		y		y	y
<i>rv3546</i>	<i>fadA5</i>	ketoacyl-CoA thiolase	y				y	y
<i>rv3548c</i>			y					y
<i>rv3549c</i>			y					y
<i>rv3550</i>	<i>echA20</i>	enoy-CoA hydratase						y
<i>rv3551</i>	<i>ipdA</i>	hydrolase	y	y	y			y
<i>rv3552</i>	<i>ipdB</i>	hydrolase		y			y	y
<i>rv3553</i>	<i>ipdC</i>	acyl-CoA reductase	y					y
<i>rv3556c</i>	<i>fadA6</i>	ketoacyl-CoA thiolase		y	y		y	y
<i>rv3557c</i>	<i>kstR2</i>	Tet-R type regulator					y	y
<i>rv3559c</i>	<i>ipdF</i>	ketoacyl-CoA reductase	y			y		y
<i>rv3560c</i>	<i>fadE30</i>	acyl-CoA dehydrogenase	y	y	y			y
<i>rv3561</i>	<i>fadD3</i>	acyl-CoA synthetase	y				y	y
<i>rv3562</i>	<i>fadE31</i>	acyl-CoA dehydrogenase					y	y
<i>rv3563</i>	<i>fadE32</i>	acyl-CoA dehydrogenase	y	y	y			y
<i>rv3564</i>	<i>fadE33</i>	acyl-CoA dehydrogenase	y			y		y

**Table 1. Continued**

<b>Locus</b>	<b>Gene name</b>	<b>Annotation</b>	<b>essential for growth on cholesterol</b>	<b>Essential for growth in macrophage</b>	<b>Essential for growth in mice</b>	<b>Essential in vitro (glycerol)</b>	<b>Upregulated in macrophage</b>	<b>Upregulated in cholesterol</b>
<i>rv3565</i>	<i>aspB</i>							y
<i>rv3567c</i>	<i>hsaB</i>	monooxygenase						y
<i>rv3568c</i>	<i>hsaC</i>	DHSA dioxygenase	y					y
<i>rv3569c</i>	<i>hsaD</i>	DSHA hydrolase	y	y			y	y
<i>rv3570c</i>	<i>hsaA</i>	monooxygenase	y	y			y	y
<i>rv3571</i>	<i>kshB</i>	monooxygenase	y				y	y
<i>rv3573c</i>	<i>chsE3</i>	acyl-CoA dehydrogenase	y					y
<i>rv3574</i>	<i>kstR</i>	Tet-R type regulator			y		y	y
<i>rv3723</i>	<i>lucA</i>	lipid transport coordinator		y	y			n/a

Essentiality for growth on cholesterol data are from (37)

Essentiality for growth in macrophage data are from (36)

Essentiality for growth in mice data are from (29)

Essentiality for growth in glycerol data are from (29)

Upregulation in macrophage data are from (30)

Upregulation in cholesterol data are from (38), n/a: not assessed.

Soon after the discovery of the Cho region, *Mtb* was confirmed to be able to use cholesterol (34, 39), and the identity of the *mce4* operon as a cholesterol transporter was also confirmed (34, 40). Tracking radioisotope-labeled cholesterol suggested the fate of two carbon atoms, C4 and C26 (Fig. 1, labeled with asterisks), that were catabolized into CO<sub>2</sub> or assimilated into mycobacterial lipids, respectively (see Section I.4). As mentioned before, *mce4* was required for *Mtb* persistence, while *mce1* was crucial for initial infection. In addition, the  $\Delta mce4$  strain could grow normally on the resting macrophages but was highly attenuated in INF- $\gamma$  stimulated macrophages (34). These indicated that the bacilli switch their nutrient source during different stages of infection, and the switch is probably an adaptation to the immune response of the host (34). *Mce1* operon was later demonstrated to encode a multisubunit fatty acid transporter (41), suggesting the crucial role of fatty acids during early infection.

In the aforementioned *mce* operon genetic interaction study, there were a group of genes besides *mceG* that showed positive genetic interactions with *mce4* operon. While most of these genes participate in the lipid degradation, two of them were demonstrated to be part of the transport system. *Rv1410c* encodes a transmembrane protein known as the P55 multidrug efflux pump. Besides conferring drug resistance in *Mtb*, *Rv1410c* was suggested to be part of the Mce4 transport system (42). Deleting P55 encoding gene in *M. bovis* BCG strain caused severe growth defect on cholesterol containing media or *in vivo*. Another gene associated with the Mce4 transport system is *rv3723*. This gene was renamed lipid uptake coordinator A (*lucA*) (43). In a synthetic lethal screening, *lucA* mutants were able to rescue the growth inhibition of  $\Delta icl1$  *Mtb* in

cholesterol-containing medium, because blocking cholesterol import could stop the production and accumulation of toxic catabolic intermediates that required detoxification by Icl1 (41) (see Section I.4.3 for Icl1 detoxification).  $\Delta lucA$  *Mtb* imported 70% less cholesterol than did the wildtype strain (41). LucA was also shown to interact directly with the subunits of Mce1 and Mce4, indicating it is part of the Mce lipid transport systems (41).

To date, all of the reported cholesterol import processes have been associated with the Mce4 transport system. However,  $\Delta mce4$  *Mtb* mutants still retained a low level of cholesterol uptake and metabolism, suggesting the existence of other complementary, although less efficient, cholesterol import approaches (34). These approaches remain to be identified.

### **I.3 Regulation of cholesterol uptake and utilization**

#### *I.3.1 The kstR regulon*

As mentioned above, the Cho region containing the *mce4* operon could be upregulated on cholesterol exposure or during macrophage infection. The upregulation is mediated by a TetR-type transcriptional regulator KstR which is also encoded in the *R. jostii* or *Mtb* Cho region (31). In fact, years before the discovery of the Cho region, Kendall et al. had been aware of the importance of *kstR* when they compared the results of three earlier experiments (29, 30, 44, 45). They noticed that among all the *Mtb* genes, *kstR* was the only gene upregulated during macrophage infection, induced by immunity,

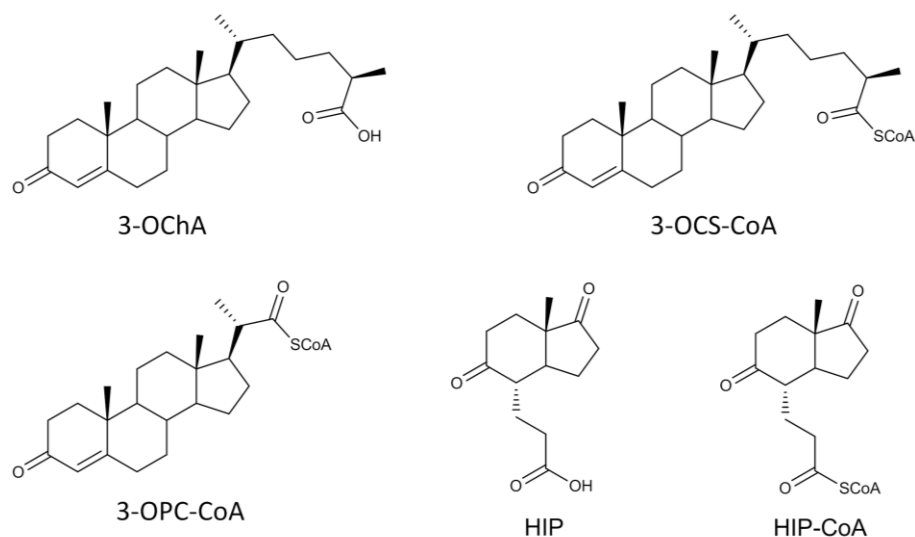
and essential in mice infection. This indicated the pivotal role of *kstR* for *Mtb* pathogenicity (44).

A follow-up study by Kendall et al. (46) showed that *M. smegmatis* KstR repressed the expression of 83 genes, namely the *kstR* regulon. Deletion of *kstR* elevated the transcription of these genes by one to two orders of magnitude, including the *mce4* operon (46) (the relationship between *kstR* and cholesterol was not known when their experiments commenced). Notably, the *kstR* regulon in *M. smegmatis* is the counterpart of the *Mtb* Cho region.

The recognition DNA sequence of KstR was predicted to contain a 14-bp palindrome TnnAACnnGTTnnA (n denotes A/T/C/G) which is located upstream of the gene operons in the *kstR* regulon (46). The dissociation constant ( $K_d$ ) between *Mtb* KstR and a 28-bp DNA segment harboring the predicted sequence was below 10 nM (47). The binding site was substantiated by DNase I footprinting of the promoter region of *M. smegmatis* *msmeg\_5228* gene (48). This gene encodes 3 $\beta$ -hydroxy-steroid dehydrogenase (HSD) in cholesterol degradation (49) (see Section I.4.1 for enzyme function). *Msmeg\_5228* gene is outside the Cho region, but its transcription was activated by a hundred-fold in *kstR* deleted *M. smegmatis* (48). Interestingly, the transcription of *Mtb hsd* (*rv1106c*) was not elevated in cholesterol medium or during macrophage infection due to a lack of the palindromic recognition DNA sequence in the promoter region (48, 50, 51). It is likely that *Mtb hsd* co-expresses with the *xseAB* exonuclease in the same operon, and the expression is constitutive. *Mtb* may have gained some benefit during evolution when HSD is constantly expressed (48).

In order to identify the ligand of KstR, Garcia-Fernandez et al. tested a panel of cholesterol catabolic intermediates and several fatty acids using a combination of thermostability assay, electromobility shift assay (EMSA), and *in vitro* transcription assay (52). They showed that an early cholesterol degradation intermediate 3-oxo-4-cholestenoic acid (3OChA, Fig. 2) was the ligand of KstR. However, cholesterol *per se* did not show any effect possibly due to its poor solubility in aqueous solution (47). This result was challenged by Ho et al. (47), who showed that 3OChA did not affect the interaction between KstR and the DNA probe. Instead, two CoA-ester intermediates, 3-oxocholest-4-en-26-oyl-CoA (3-OCS-CoA, Fig. 2) and 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA), were able to release DNA from KstR (compounds in the original literature were termed 3-OChA-CoA and 4-BNC-CoA, but the nomenclature will be consistent throughout this review). The  $K_d$  values for each effector against KstR were around 60 and 280 nM, respectively. Interestingly, in the crystal structure of KstR in complex with 3OCS-CoA or 3-OPC-CoA, the CoA moiety beyond the sulfur atom in either structure was invisible, indicating that CoA was not directly involved in the binding but might improve the solubility of the effector (47). Both 3OCS-CoA and 3-OPC-CoA were able to induce a conformational change of KstR (47).





**Figure 2. The signal molecules of *kstR* and *kstR2* regulons**

### 1.3.2 The *kstR2* regulon

Kendall et al. later found a 15-gene sub-region (*rv3548c-rv3553*, *rv3556-rv3557*, *rv3559-rv3565*) inside the *kstR* regulon that was regulated by a second TetR-type of regulator KstR2 (Rv3557) (53). Those 15 genes in the *kstR2* regulon distributed in three operons each containing a 14-bp palindrome AnCAAGnnCTTGnT in the promoter region. DNA probe containing the palindrome could be specifically shifted by KstR2 in EMSA (53). Similar to the *kstR* regulon, deleting *M. smegmatis kstR2* de-repressed the transcription of those 15 gene orthologs in *M. smegmatis*. Interestingly, the regulation by KstR and KstR2 were independent (53).

Casabon et al. observed that *kstR2* regulon could be activated by carbon sources including cholesterol or 3 $\alpha$ -H-4 $\alpha$ -(3'-propanoate)-7 $\beta$ -methylhexahydro-1,5-indanedione (HIP, Fig. 2) (54). The latter is a catabolic intermediate of the cholesterol

C/D-ring degradation (see Section I.4.3). In addition, deletion of *fadD3* (*rv3561*) inside the *kstR2* regulon abolished the regulon activation. The authors further demonstrated that HIP-CoA (Fig. 2), the product of FadD3, was the effector of the KstR2 (54, 55). This naturally led them to the prediction that the role of the *kstR2* regulon was to catabolize the C/D-ring of cholesterol, and that *Mtb* was able to degrade cholesterol completely (48).

#### **I. 4 Cholesterol degradation pathway in *Mtb***

Cholesterol degradation has been observed to occur concurrently from both the aliphatic side-chain and the A-ring in some cholesterol-degrading bacteria. But the exact order of the two is not clear and may be organism various (56, 57). *Mtb* and *M. bovis* BCG cannot process the A/B-ring when the side-chain is intact (58, 59). Cholesterol catabolism will be reviewed in the order of side-chain, the A/B-ring and, lastly, the C/D-ring degradation.

##### *I.4.1 $\beta$ -oxidation of cholesterol aliphatic side-chain*

###### **I.4.1.1 HSD oxidizes the 3-OH and isomerizes cholesterol**

HSD, an NAD<sup>+</sup>-dependent dehydrogenase/isomerase, is largely considered to catalyze the initial step in cholesterol degradation (48, 49, 59). HSD converts the 3 $\beta$ -hydroxy to 3-oxo on the A-ring and then isomerizes the de-saturated position from  $\Delta^5$  to  $\Delta^4$  (50) The resultant product is an enone-containing molecule cholest-4-en-3-one (Fig. 3). This step may not be absolutely required before side-chain degradation, because *hsd*-

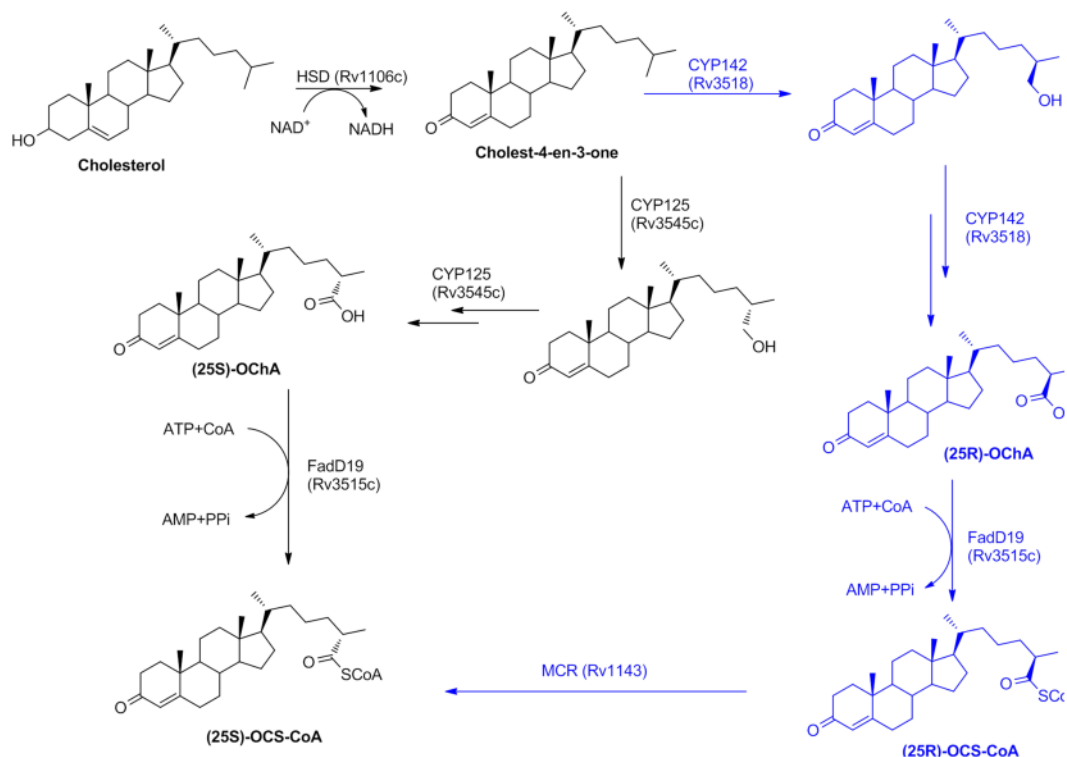
deleted *Mtb* laboratory strain H37Rv could survive on cholesterol medium. As mentioned in I.3.1, *hsd* is outside the *Mtb* Cho region, and its transcription is not regulated by KstR due to the lack of recognition DNA in the promoter region (46, 48, 53).

Besides the NAD<sup>+</sup>/NADP<sup>+</sup>-dependent dehydrogenases, some cholesterol-degrading microorganisms use a FAD-dependent cholesterol oxidase (ChOx) to catalyze the same transformation via a different mechanism. *Mtb* genome encodes a ChOx type of enzyme, ChoD (Rv3409c) (51), (60). However, ChoD appeared not being able to catalyze the transformation of cholesterol to cholest-4-en-3-one, nor was it essential for cholesterol utilization (49). Moreover, *hsd* was required for *Mtb* clinical strain CDC1551 to grow on the medium using cholesterol as the sole carbon source (61), indicating the non-redundant function of HSD in *Mtb* CDC1551. The *in vivo* essentiality of *hsd* was paradoxical. *Hsd* was not essential for *Mtb* to infect mice macrophages or guinea pigs. This might be because during the experimental conditions, *Mtb* had access to multiple carbon sources (50).

#### I.4.1.2 Cytochrome P450 enzymes activate the aliphatic side chain

In order to subject the aliphatic side chain into  $\beta$ -oxidation cycles, the terminal methyl groups must be first oxidized to carboxylic acid form. An operon containing six genes (*rv3540c-rv3545c*) was previously shown by Chang et al. to be essential for *Mtb* growth in macrophage and in mice, and the operon was renamed the intracellular growth (*igr*) operon (62). The *igr* operon contains a cytochrome P450 enzyme CYP125

(Rv3545c). CYP125 catalyzes sequential oxidation on the C26 of cholest-4-en-3-one (58, 59, 63-65) (Fig. 3). It is worth mentioning that C25 of cholesterol is a prochiral carbon which is turned into a chiral center after asymmetric modification on either C26 or C27 terminal methyl group. CYP125 stereospecifically generates the (25S)- form. Another cytochrome P450 enzyme CYP142 (Rv3518), on the other hand, produces the (25R)- form (66) (Fig. 3). These two enzymes successively oxidize cholest-4-en-3-one all the way to carboxylic acid form, each maintaining their own stereospecificity. FadD19 (Rv3515c), a fatty acyl CoA ligase, can take both chiral forms as substrates and ligate them with a coenzyme A molecule while still maintaining the chiral configuration of C25 (67, 68). However, only the (25S)- form of the resultant product 3-OCS-CoA is the substrate of  $\beta$ -oxidation enzymes (69). *Mtb* genome encodes an  $\alpha$ -methyl acyl CoA racemase (MCR, Rv1143) which interconverts the (25R)- and (25S)-3-OCS-CoA (69). FadD19 and MCR together make CYP125 and CYP142 partially redundant in *Mtb* H37Rv strain, and either CYP is nonessential in the presence of the other. The clinical *Mtb* strain CDC1551, however, does not encode CYP142 in the genome, which makes CYP125 indispensable (66).



**Figure 3. Cholesterol aliphatic side chain activation**  
**Pathway drawn in blue indicates the intermediates with (25R)- configuration.**

#### I.4.1.3 $\beta$ -oxidation of the side chain

Each cycle of lipid  $\beta$ -oxidation typically involves an acyl-CoA dehydrogenase (FadE), an (S)-enoyl-CoA hydratase (EchA), a  $\beta$ -hydroxy-acyl-CoA dehydrogenase (FadB), and a  $\beta$ -ketoacyl-CoA thiolase (FadA). However, *Mtb* genome, including the Cho region, is highly redundant in each type of these enzymes. As a result, a specific reaction may be catalyzed by multiple enzymes or enzyme complexes. On the other

hand, some of these enzymes may catalyze multiple reactions. It is difficult to annotate the gene functions precisely.

ChsE4-ChsE5 (formerly FadE26-FadE27, Rv3504-Rv3505) dehydrogenase complex has been shown to catalyze the first round of  $\beta$ -oxidation of 3-OCS-CoA (70) (Fig. 4). ChsE4-ChsE5 only recognizes the (2S)- form of 3-OCS-CoA. Crystal structure of ChsE4-ChsE5 complex revealed that the two proteins were structurally similar to each other and formed an  $\alpha_2\beta_2$ -type of heterotetramer (70). Notably, ChsE4-ChsE5 was the first structure of  $\alpha_2\beta_2$ -type of acyl-CoA dehydrogenase. Unlike the typical homotetrameric acyl-CoA dehydrogenases where each polypeptide chain binds to one FAD cofactor, ChsE4-ChsE5 heterotetramer only binds to two FAD molecules. Each FAD was located on the interface of an  $\alpha\beta$ -heterodimer (70). ChsE4-ChsE5 complex was shown to be very versatile. It exhibited high specific activity not only for 3-OCS-CoA, but also for the CoA-esters of the second and third rounds of  $\beta$ -oxidation. This potentially explains why the downstream acyl-CoA dehydrogenases ChsE3 and ChsE1-ChsE2 are not absolutely essential (70).

The *echA*, *fadB* and *fadA* genes involved in the first round  $\beta$ -oxidation are not clear. Since there are only two *echA* genes, *echA19* and *echA20*, in the Cho region, and *echA20* is involved in C/D-ring degradation (71), it is possible that EchA19 catalyzes the hydration of multiple enoyl-CoA intermediates. Another possibility is the involvement of EchAs outside the Cho region (51). There is no *fadB* gene in the Cho region. This step may be catalyzed by genes elsewhere in the genome (51). Similar to *echA*, the Cho region encodes two *fadA* genes, *fadA5* and *fadA6*. The latter participates in the C/D-ring

degradation (71, 72), which indicates the potential versatility of *fadA5* in the side chain degradation. FadA5 was shown to be required to generate the side-chain-removed product 4-androstene-3,17-dione (AD) (38), but the exact substrate specificity of FadA5 has not been investigated.

At the end of the first round of  $\beta$ -oxidation, a propionyl-CoA molecule containing the C25 to C27 of cholesterol is released. As mentioned in Section I.2.1, the radioactivity of  $^{14}\text{C}$ 26-labeled cholesterol was detected in mycobacterial lipids (34). This was because propionyl-CoA could be converted to methylmalonyl-CoA and incorporated into the methyl-branched lipids (MBLs) (see Section I.4.2).

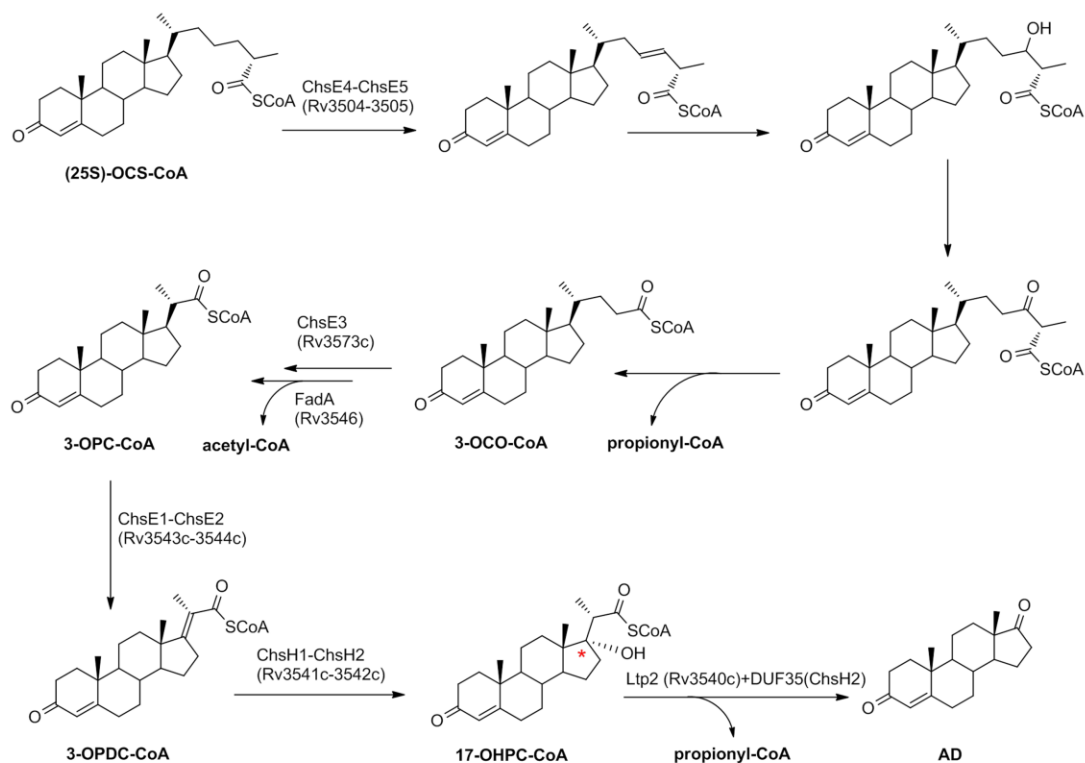
The second round of  $\beta$ -oxidation may be catalyzed by an acyl-CoA dehydrogenase ChsE3 (formerly FadE34, Rv3573c) (70). Unlike ChsE4-ChsE5, ChsE3 showed substrate specificity for 3-OCO-CoA (70). At the end of the second round  $\beta$ -oxidation, an acetyl-CoA molecule containing C23 and C24 of cholesterol may be released by FadA5 (38) (Fig. 4).

The third round of  $\beta$ -oxidation is non-cannonical, and it is well-studied. The genes involved are clustered in the aforementioned *igr* operon. ChsE1-ChsE2 (formerly FadE28-FadE29, Rv3543c-Rv3544c) (73, 74) specifically catalyze the conversion of 3-OPC-CoA into 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA,  $k_{\text{cat}}/K_{\text{m}} = 2.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ) (74) ChsE1-ChsE2 complex is also a heterotetramer with two FAD molecules bound (74). The  $\beta$ -hydroxylation at C17 of 3-OPDC-CoA is catalyzed by a MaoC-like (R)-enoyl-CoA hydratase complex ChsH1-ChsH2 (Rv3541c-Rv3542c). This reaction generates 17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA (17-OHPC-CoA)

(73, 75). Notably, after  $\beta$ -hydroxylation, C17 becomes a tertiary carbon with no hydrogen atom attached (Fig. 4, labeled with an asterisk). Dehydrogenation cannot occur on this carbon to produce  $\beta$ -ketoacyl-CoA as a substrate of a thiolase. As a result, an aldolase Ltp2 (lipid transfer protein, Rv3540c) is employed to cleave a propionyl-CoA directly from 17-OHPC-CoA (75, 76). The retro-aldol cleavage by Ltp2 required the presence of ChsH1-ChsH2 (76). Notably, ChsH2 contains a C-terminal domain of unknown function DUF35 that could be pulled down by Ltp2 (75, 76). The aldolase activity could be reproduced simply by Ltp2-DUF35 complex against 17-OHPC-CoA. The specific activity for this reaction was quite high with a  $k_{\text{cat}}/K_m$  value of around  $2.4 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$  (76).

After three rounds of  $\beta$ -oxidation, the aliphatic side-chain of cholesterol is catabolized into two molecules of propionyl-CoA and one molecule of acetyl-CoA (Fig. 4). The remaining compound AD contains only the A/B and C/D rings.





**Figure 4. Degradation of aliphatic side-chain C17 in 17-OHPC-CoA is labeled with an asterisk.**

#### I.4.2 Degradation of the A/B-Ring

##### I.4.2.1 KstD catalyzes the 1,2-desaturation

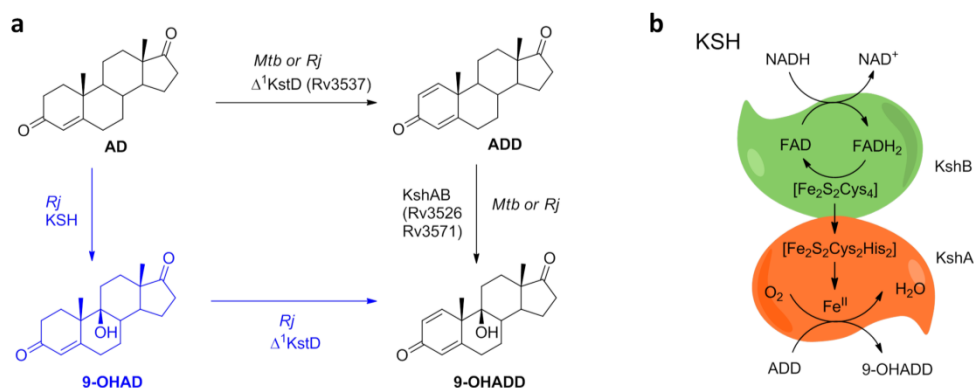
In order to open the A/B-ring, multiple oxidation reactions have to occur on the rings. The first step is to desaturate the C1-C2 single bond of AD to generate 1,4-androstadiene-3,17-dione (ADD) (Fig. 5a). This reaction is catalyzed by a FAD-dependent ketosteroid dehydrogenase  $\Delta^1$  KstD (Rv3537). *In vitro* enzyme assay of *Mtb* KstD showed that it recognized a narrow range of substrates which need to be side-chain completely degraded (77). Disrupting *kstD* resulted in the accumulation of a downstream

intermediate 9-hydroxy-1,4-androstadiene-3,17-dione (9-OHADD) in *Mtb* (39). This indicates that the downstream hydroxylase KSH can share the substrate of KstD. In fact, the order of 1,2-desaturation and 9-hydroxylation in *R.jostii* is not critical because *rhodococci* have three KstD enzymes and they can accept a wide range of steroid substrates (77, 78). However, *Mtb* genome only encodes one  $\Delta^1$  KstD, and it did not show *in vitro* activity against 9-OHAD (77), which means desaturation has to occur prior to hydroxylation. This raises the query of how this order is controlled in *Mtb* (see below).

#### I.4.2.2 KSH catalyzes the 9 $\alpha$ -hydroxylation

The hydroxylation at C9 of ADD is catalyzed by a two-component enzyme 3-ketosteroid 9 $\alpha$ -hydroxylase (64, 79) (Fig. 5a). The  $\alpha$ -subunit KshA (Rv3526) is an oxygenase containing a Rieske-type [Fe<sub>2</sub>S<sub>2</sub>Cys<sub>2</sub>His<sub>2</sub>] cluster and a non-heme mononuclear Fe<sup>2+</sup> center. The  $\beta$ -subunit KshB (Rv3571) is a reductase containing a plant-type [Fe<sub>2</sub>S<sub>2</sub>Cys<sub>4</sub>] cluster and a FAD cofactor. During the catalysis, the electron from an NADH in the solution is first transferred to FAD in the KshB reductase subunit (Fig. 5b). Relayed by the iron-sulfur clusters in KshB and KshA, the electron reaches the mononuclear ferrous center of KshA oxygenase subunit. The C9 hydroxylation is accomplished by consuming an oxygen molecule as the terminal electron acceptor (64, 79). The reported specific activity of *Mtb* KshA-KshB was only twice for ADD than for AD (79). However, when scrutinizing the global fitting curve of the steady-state kinetics data for AD, one would notice that the standard deviation of the replicates was large, and

the fitting was not accurate. In addition, the maximum velocity for ADD was approximately ten-fold higher than for AD (79). These data indicate that KSH in *Mtb* prefers  $\Delta^{1,4}$  dual desaturated substrates, and explains the previous question of how *Mtb* ensures that KstD is upstream of KSH.



**Figure 5. Activation of A/B-ring by desaturation and hydroxylation**  
**a, The catalytic order of *Mtb* KstD and KSH is strict (black), but that of *R. jostii* homologs is flexible (both black and blue). b, Schematic of the catalytic mechanism of KSH.**

#### I.4.2.3 HsaA-HsaB catalyzes the 4-hydroxylation

The  $\Delta^1$  de-saturation and 9 $\alpha$ -hydroxylation will lead to the spontaneous open of the B-ring and aromatization of the A-ring (57) to produce 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (3-HSA). In order to open the A-ring, another hydroxyl group is introduced at C4- position (Fig. 6a, colored in blue) to generate a catechol-containing derivative 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (3,4-DHSA) (14, 31, 80). This step in *Mtb* is catalyzed by HsaA-HsaB (Rv3570c

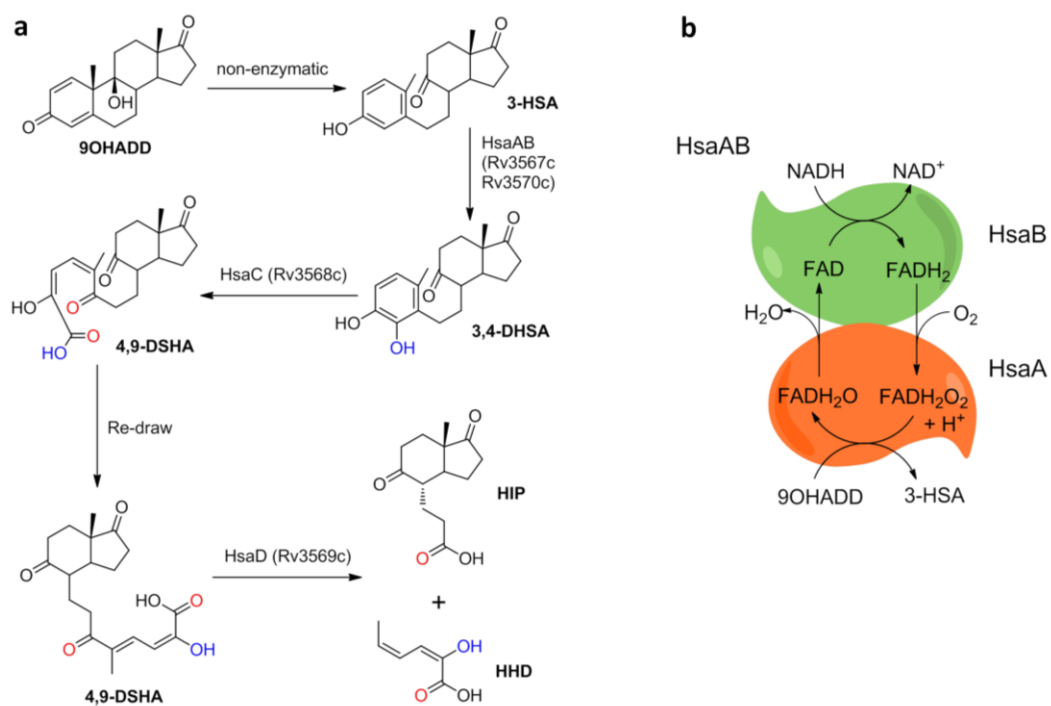
and Rv3567c) (81) which are encoded in the *hsa* operon containing *hsaABCD* genes. HsaA-HsaB enzymes belong to the two-component flavin-dependent monooxygenases (TC-FDMs) (81) where the oxidized and reduced flavin is shuttled between the reductase and oxygenase subunits during catalysis (82) (Fig. 6b). Specifically, HsaB reductase subunit reduces a flavin using an NADH. The reduced flavin is transferred to HsaA oxygenase subunit, where the hydroxylation occurs on the C4 position of 3,4-DHSA by consuming an oxygen molecule. The oxidized flavin is then shuttled back to HsaB for the next round of catalysis. Different monooxygenases have the preference to use either FAD or FMN as the flavin cofactor (83). *Mtb* HsaA-HsaB showed similar specific activity when using FAD ( $k_{\text{cat}}/K_m = 770 \text{ M}^{-1} \text{ s}^{-1}$ ) or FMN ( $k_{\text{cat}}/K_m = 1000 \text{ M}^{-1} \text{ s}^{-1}$ ) as a cofactor, although both the  $k_{\text{cat}}$  and  $K_m$  were higher for FMN (81).

#### I.4.2.4 HsaC opens A-ring

The *meta*-cleavage between C4 and C5 of 3,4-DHSA is catalyzed by HsaC (Rv3568c), an iron-dependent extradiol dioxygenase. The reaction generates 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid (4,9-DSHA). Unlike the intradiol dioxygenases which cleave the aromatic bond between two hydroxylated carbons, an extradiol dioxygenase cleaves an aromatic ring between a hydroxylated carbon and an adjacent non-hydroxylated carbon (84, 85). The *meta*-cleavage of a catechol requires a mononuclear non-heme ferrous iron (Fig. 7). The crystal structure of HsaC confirmed that the ferrous iron was located at the active site and coordinated by two histidines and one glutamate (86). This type of coordination in extradiol

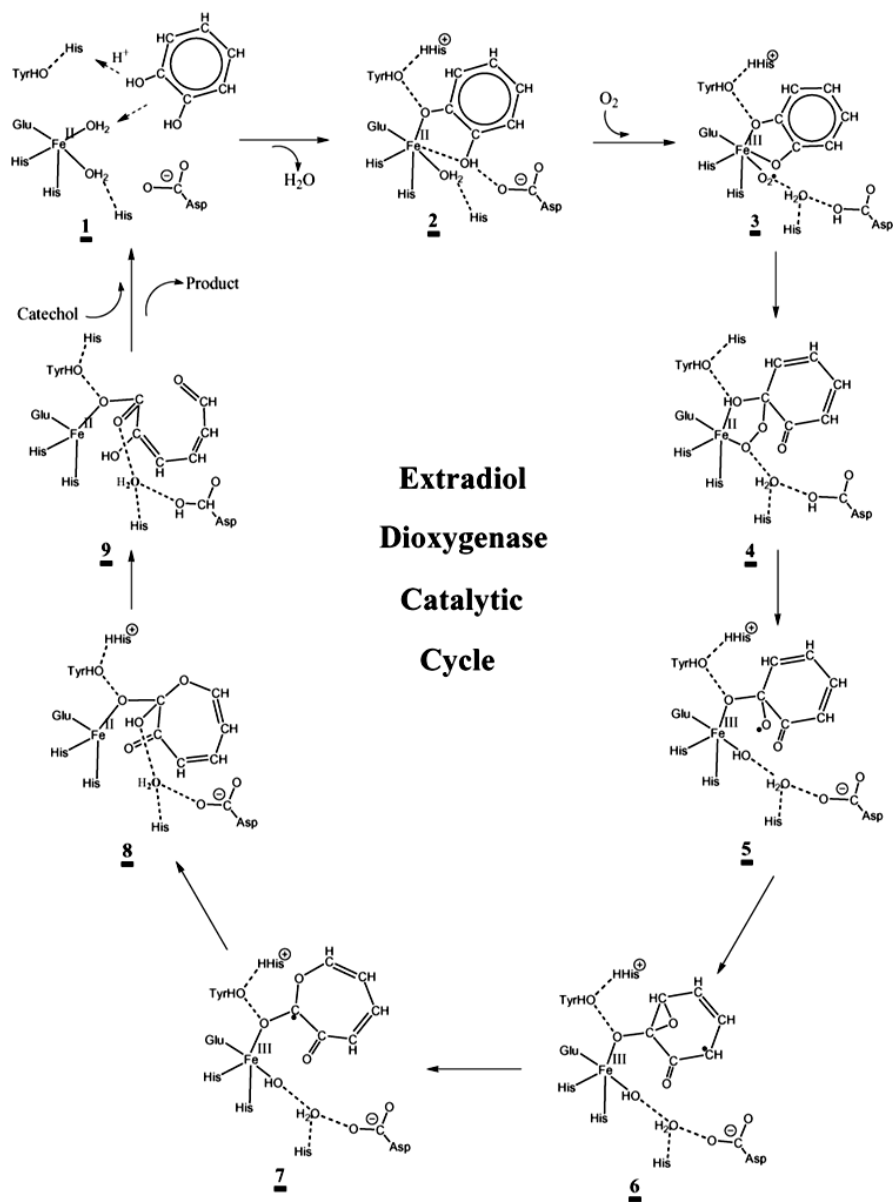
dioxygenases is known as the 2-His-1-carboxylate facial triad motif because the three residues only occupy one face of the octahedral coordination sphere (87) (Fig. 7). The other face of the sphere can be coordinated by three solvent molecules. During catalysis, the two hydroxyl groups of the catecholic substrate substitute two solvent molecules, and an oxygen molecule substitutes the third solvent molecule (Fig. 7). The reaction is featured by the formation of a peroxide bridge between the ferrous iron and the catecholic substrate, followed by homolytic O-O bond fission. Through a Criegee-type rearrangement to form a putative seven-membered lactone intermediate (85, 87), the C-C bond between C4 and C5 of 3,4-DHSA is cleaved.

The physiological role of HsaC in cholesterol degradation was confirmed by a knockout experiment (86). *Mtb*  $\Delta$ *hsaC* strain was not able to completely degrade cholesterol but accumulated catechols which could be oxidized to 1,2-benzoquinone derivatives. These compounds in turn gave the initial *Mtb* culture a pinkish color. The bacilli lost viability soon, likely because they could not further degrade the intermediates to produce enough energy and/or they accumulated toxic intermediates (86).



**Figure 6. Degradation of A/B-ring (Part 1)**

**a, Catabolic pathway and intermediates. The oxygen atoms are colored base on the catalytic steps. b, Schematic of HsaAB FAD-dependent monooxygenase.**



**Figure 7. Suggested catalytic cycle for extradiol dioxygenase, adapted from (85) with permission.  
Copyright (2004) American Chemical Society.**

#### I.4.2.5 HsaD cleaves HHD off 4,9-DSHA

The last gene in *hsa* operon encodes an  $\alpha/\beta$  hydrolase HsaD (Rv3569c) (31, 88). HsaD is responsible for cleaving 2-hydroxy-hexa-2,4-dienoic acid (HHD) from 4,9-DSHA (Fig. 6a). Interesting, Lack et al. reported that upon binding to HsaD, the maximal absorption of 4,9-DSHA underwent a redshift of 60 nm (88). The crystal structures of an inactive mutant HsaD<sub>S114A</sub> in complex with several catechol *meta*-cleavage products showed that the torsion angle about C4-C5 double bond (Fig. 8a, labeled with asterisks) dramatically deviated from planarity (88). Therefore, they proposed that the enolate substrate tautomerized into the keto form, which made C4-C5 a rotatable single bond. The keto form could also explain the redshift phenomenon. Based on the crystal structures and the optical phenomenon, Lack et al. tentatively proposed a general acid-base catalytic mechanism (Fig. 8a). It is still disputable of whether the catalytic serine uses a general base mechanism to de-protonate a water molecule (Fig. 8a) or it acts as a nucleophile to directly attack the carbonyl at C6 (89-91) (Fig. 8b).

#### I.4.2.6 HsaEFG degrades HHD

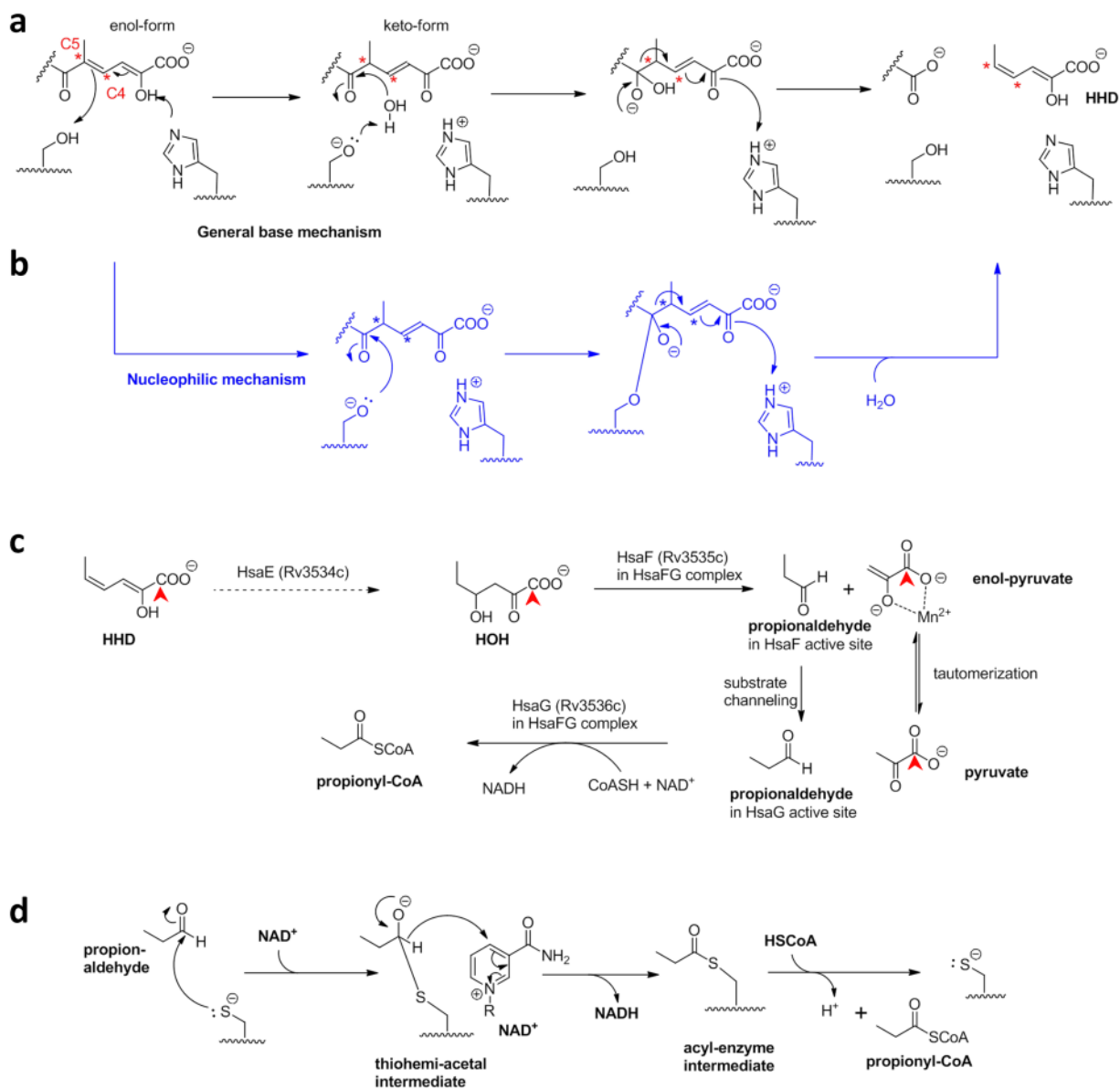
Bioinformatics analyses (31) suggested the enzymes responsible for the degradation of HHD (Fig. 6a, 8c) are encoded in the *hsaEFG* operon (rv3534c-rv3536c). The function of HsaE has not been characterized but was proposed based on other *meta*-cleavage pathway intermediates (92) (Fig. 8c). 4-hydroxy-2-oxohexanate (HOH), the putative product of HsaE, could be catabolized by HsaF-HsaG complex with specific



activity of approximately  $7.91 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  (92). HsaF catalyzes a retro-aldol cleavage in the presence of a divalent metal ion to stabilize the enol pyruvate intermediate.  $\text{Mn}^{2+}$  showed the highest activity for the enzyme. The retro-aldol reaction generates pyruvate and propionaldehyde (92).

The crystal structure of HsaF-HsaG complex showed a tunnel linking the active sites of HsaF and HsaG (92). This structure along with a homologous protein complex DmpF-DmpG structure (93) implied a substrate channeling mechanism (92) (Fig. 8c). After the retro-aldol cleavage, propionaldehyde would be channeled to the active site of HsaG for CoA-dependent dehydrogenation (92) (Fig. 8d). HsaG produces the third molecule of propionyl-CoA in cholesterol degradation.

As mentioned in section I.2, Pandey and Sasseti showed that the radioactivity of  $^{14}\text{C}_4$  labeled cholesterol was detected in  $^{14}\text{CO}_2$  (34). Based on the retro-aldol cleavage by HsaF, the C4 carbon of cholesterol is turned into the carboxyl group of pyruvate (Fig. 8c, indicated by red arrowhead). To enter the tricarboxylic acid cycle, pyruvate undergoes oxidative decarboxylation catalyzed by pyruvate dehydrogenase to form acetyl-CoA. Therefore, the carboxyl group ends up as  $\text{CO}_2$ . This eventually explains the fate of C4 carbon of cholesterol.



**Figure 8. Degradation of A/B-ring (Part 2)**

**a**, A simplified general base mechanism of HsaD as proposed in (88). Carbons C4 and C5 that undergo a double bond shift are labeled with asterisks. **b**, An alternative nucleophile mechanism of HsaD analogous to other *meta*-cleavage product hydrolases as proposed in (89). **c**, Degradation of HHD. Dashed arrow indicates the reaction is hypothesized. Red arrowheads indicate C4 of cholesterol. **d**, Proposed catalytic mechanism of HsaG.

### I.4.3 Degradation of the C/D-Ring

Removing HHD from 4,9-DSHA yields the C/D-ring-contain compound 3 $\alpha$ -H-4 $\alpha$ -(3'-propanoate)-7 $\alpha$  $\beta$ -methylhexahydro-1,5-indanedione (HIP, Fig. 2, 6a, 9). Genetic studies suggested that the genes involved in HIP degradation are clustered in *kstR2* operon, but the catabolic pathway was unknown until very recently (71). Biochemical properties of most enzymes for C/D-ring degradation are still poorly understood.

#### I.4.3.1 FadD3 initiates the degradation of C/D-ring

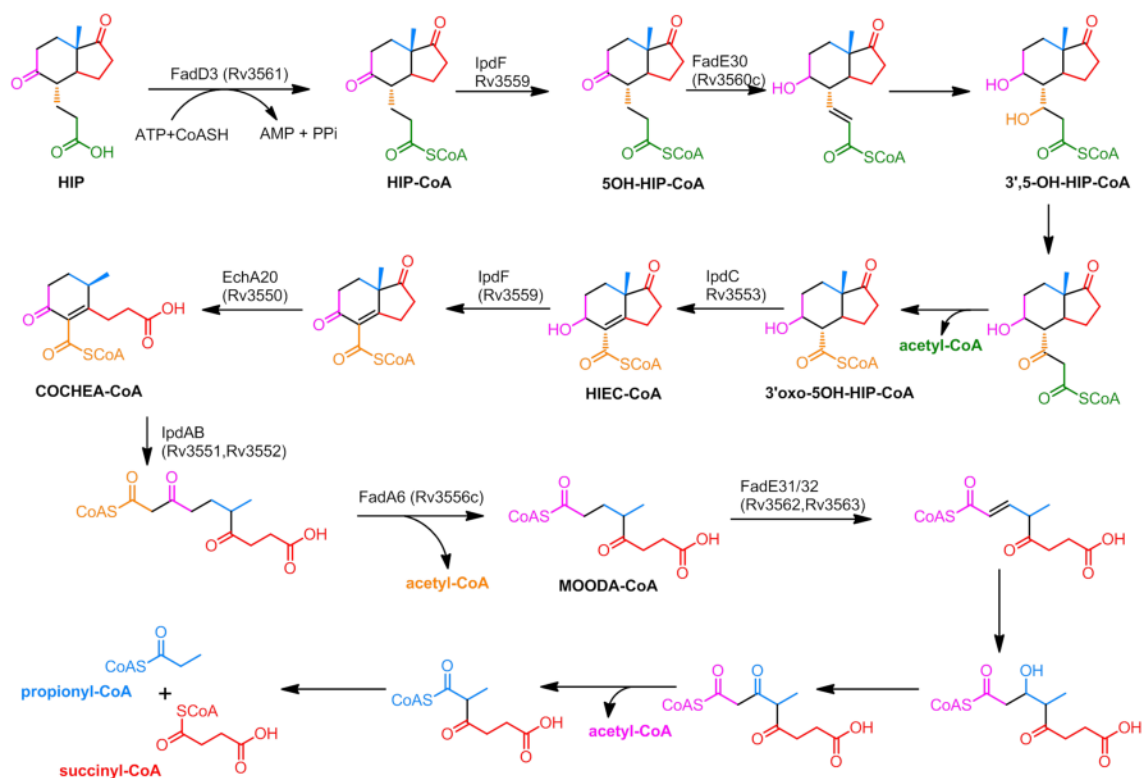
FadD3 (Rv3561) is one of the predicted CoA ligases in the cholesterol degradation gene cluster and is the only CoA ligase in the *kstR2* regulon (53, 54). The biochemical function of FadD3 was characterized by Casabon et al. in *R. jostii* RHA1 strain through a comparative metabolism of wildtype and *fadD3*-deleted strains (55). *R. jostii*  $\Delta$ *fadD3* strain grew on cholesterol showed incomplete degradation of cholesterol and the accumulation of HIP. *Mtb fadD3* was able to restore the full activity of cholesterol degradation in *R. jostii*  $\Delta$ *fadD3* strain (55). Steady-state kinetics of *Mtb* FadD3 showed that the enzyme could ligate CoA onto HIP with specific activity of  $7.3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (55).

As mentioned at the beginning of Section I.4, the cholesterol catabolism may be concurrent from the side-chain and the A/B-ring. While some A/B-ring catabolic enzymes are active against side chain partially degraded substrates, FadD3 is only active against the side chain fully removed substrates (55), indicating that the degradation of the C/D-ring is absolutely the last stage of the cholesterol catabolism.

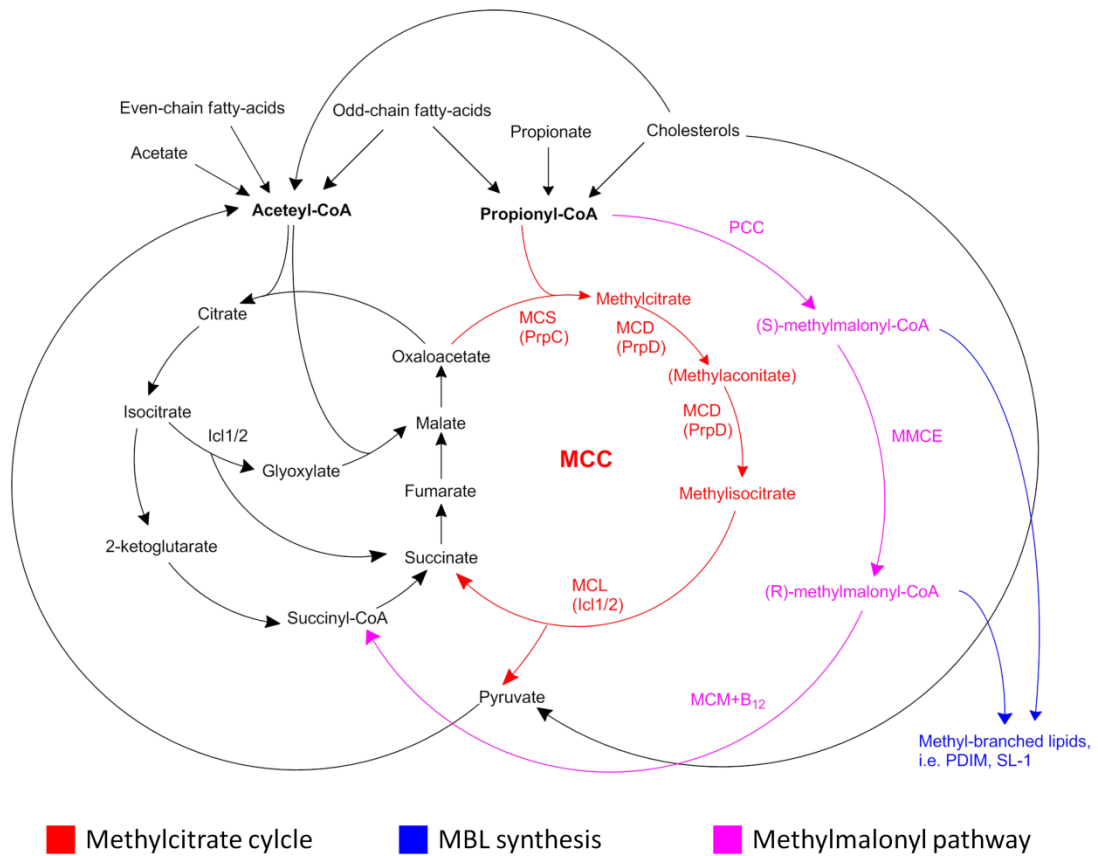
#### I.4.3.2 Degradation of the C/D-ring yields acetyl-, propionyl- and succinyl-CoA

The identities of the enzymes and intermediates during the C/D-ring degradation were elucidated in 2017 using a combination of bioinformatics, genetics and biochemical tools (71). Crowe et al. purified protein orthologs of the *Mtb kstR2* regulon from several cholesterol degrading organisms, including *Mtb*, *R. jostii* RHA1, and *Pseudomonas putida* DOC21, and reconstituted the pathway *in vitro*. They further validated the *in vivo* functions of these enzymes by tracking the intermediates accumulated in the *M. smegmatis* mutants with these genes individually deleted (71). The catabolic pathway is shown in Fig. 9. Except for the IpdAB and FadA6 (72), the biochemical properties of these enzymes have not been characterized yet. The degradation includes ring-open steps coupled by several rounds of  $\beta$ -oxidation (71). The final products contain three molecules of acetyl-CoA, one molecule each of propionyl- and succinyl-CoA.

Taking the side-chain and A/B-ring degradation together, the catabolism of cholesterol yields four molecules each of acetyl- and propionyl-CoA, one molecule of succinyl-CoA and one molecule of pyruvate. The acetyl-CoA, succinyl-CoA, and pyruvate can be readily fed into the TCA cycle and glyoxylate shunt for further catabolism or be used as synthetic blocks for anabolism. However, propionyl-CoA was considered as a potentially toxic molecule to *Mtb* and requires several pathways be assimilated. These pathways include the vitamin B<sub>12</sub>-dependent methylmalonyl pathway (MMP), the membrane-associated methyl-branched lipid (MBL) synthesis, and the methylcitrate cycle (MCC) (Fig. 10).



**Figure 9. Degradation of C/D-ring, adapted and modified from (71). The acyl-CoAs that are sequentially removed from the C/D-ring are shown in different colors.**



**Figure 10. Overview of the propionyl-CoA assimilation pathways**  
**The propionyl-CoA assimilation pathways are shown along with the TCA cycle and glyoxylate shunt in the central carbon metabolism.**

## I.5 Metabolism of the unique product propionyl-CoA in *Mtb*

### I.5.1 Methylmalonyl pathway (MMP)

#### I.5.1.1 *Mtb* has a functional MMP

MMP is a well-studied pathway that organisms use to metabolize propionyl-CoA. The pathway widely exists from prokaryotes to eukaryotes (26, 94-98). Enzymes involved in this pathway include propionyl-CoA carboxylase (PCC), methylmalonyl-CoA epimerase (MMCE, also named methylmalonyl-CoA racemase) and

methylmalonyl-CoA mutase (MCM). Genes encoding these enzymes in *Mtb* have been identified or bioinformatically assigned. *Mtb* propionyl-CoA carboxylase complex (PCC) contains  $\alpha$ ,  $\beta$  and  $\epsilon$  subunits which are encoded by three nearby genes *rv3285*, *rv3280* and *rv3281*, respectively (96, 99, 100). PCC carboxylates propionyl-CoA to generate (S)-methylmalonyl-CoA. This step is mediated by a biotin cofactor and consumes one ATP to activate a bicarbonate molecule. The conversion of (S)-methylmalonyl-CoA to (R)-methylmalonyl-CoA is catalyzed by MMCE. Bioinformatics study assigned Rv1322A as *Mtb* MMCE (96). But its biochemical function has not been experimentally validated. A critical and unique step in MMP is the reversible rearrangement between methyl-branched (R)-methylmalonyl-CoA and linear succinyl-CoA. This step is catalyzed by a vitamin B<sub>12</sub>-dependent enzyme MCM. *Rv1492* (*mutA*) and *rv1493* (*mutB*) in *Mtb* encode the  $\beta$ - and  $\alpha$ -subunits of MCM, respectively (96). The  $\alpha$ -subunit, MutB, binds to both B<sub>12</sub> cofactor and the (R)-methylmalonyl-CoA substrate. Despite sharing a similar fold with the  $\alpha$ -subunit, the  $\beta$ -subunit does not bind to B<sub>12</sub> or acyl-CoA (101). A nearby gene *rv1496* is likely to encode an MCM associated GTPase which is involved in assembling MCM-B<sub>12</sub> complex and protecting MCM from inactivation (96, 102, 103). Savvi et al. (96) and others (104, 105) showed that MMP could partially compensate the disruption of MCC for *Mtb* survival on propionate media when vitamin B<sub>12</sub> was supplemented into the media. Notably, for propionate utilization, bacteria need to activate this three-carbon molecule into the propionyl-CoA form. In *Mtb*, this reaction is presumably catalyzed by propionyl-CoA synthetase Rv3667 (105). These results together indicated that *Mtb* possesses a functional MCM pathway, although B<sub>12</sub> must be

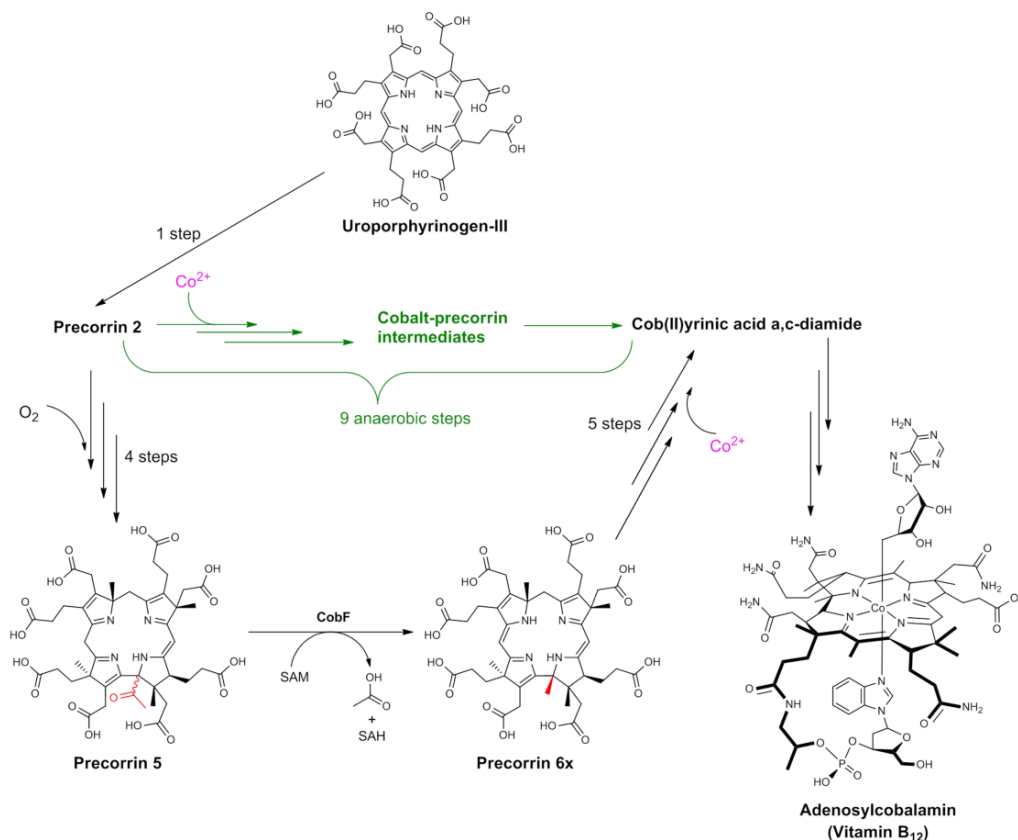
exogenously supplemented during *in vitro* growth. This raises the question of whether *Mtb in vivo* synthesizes B<sub>12</sub> cofactor or acquires it via other approaches.

#### I.5.1.2 Vitamin B<sub>12</sub> acquisition in *Mtb*

Vitamin B<sub>12</sub> is one of the most complicated natural compounds. Its *de novo* synthesis is absent in eukaryotes and is limited to only a subset of archaea and eubacteria (106). The biosynthetic pathway involves around 25 genes starting from the last common precursor uroporphyrinogen-III (107) (Fig. 11). Different organisms use either the oxygen-dependent (aerobic) route or the oxygen-independent (anaerobic) route to synthesize B<sub>12</sub> (106, 107). The major differences are that the latter incorporates cobalt before corrin synthesis while the former does it in the opposite order, and the former requires an oxygen molecule to hydroxylate precorrin 3A (106, 107) (Fig. 11).

Comparative genomics studies showed that *Mtb* seems to possess most of the gene homologs of the oxygen-dependent route except for *cobF* which is responsible for converting precorrin-5 into precorrin-6x (26, 107, 108) (Fig. 11). CobF is an S-adenosylmethionine (SAM) dependent methyltransferase (106, 109). Although it was hypothesized that this step might be compensated by other methyltransferases in *Mtb* (108), the hypothesis has not been validated. As a result, it is still arguable whether *Mtb* is able to complete the B<sub>12</sub> *de novo* synthesis.





**Figure 11. Vitamin B<sub>12</sub> biosynthesis**  
**Anaerobic pathways are colored green. The reaction catalyzed by CoF, which is missing in *Mtb*, is specified.**

Nevertheless, it is known that *Mtb* genome encodes a B<sub>12</sub> transporter. Gopinath et al. screened the transposon insertion mutant library of *Mtb* (32) on the B<sub>12</sub>-containing culture media and found that mutating *rv1819c* abolished the B<sub>12</sub> uptake. Sequence analysis suggested that *rv1819c* encodes an ABC transporter containing a transmembrane domain and a nucleotide-binding domain (NBD). *Rv1819c* was also crucial for the bacilli to use propionate when MCC was chemically inhibited (110).

### I.5.1.3 *Mtb* may circumvent B<sub>12</sub>-dependent metabolic steps

Although possessing the membrane-anchored B<sub>12</sub>-uptaking machinery, whether *Mtb*, especially the CDC1551 clinical strain, practically use B<sub>12</sub>-dependent MMP during infection and persistence is still not clear. *Mtb* laboratory strain H37Rv genome harbors two methionine synthase genes *metH* (*rv2124c*) and *metE* (*rv1133c*) (26). Both can synthesize methionine from homocysteine. The former is a B<sub>12</sub>-dependent methyltransferase, and the latter is B<sub>12</sub>-independent. However, compared to the H37Rv genome, the CDC1551 genome contains a deletion that resulted in a loss of 398 amino acids at the C-terminus of MetH. The deletion abolished the function of MetH and left the B<sub>12</sub>-independent MetE the only available methionine synthase for CDC1551 (111). Paradoxically, there is a B<sub>12</sub> riboswitch upstream of the *metE* gene, which means that in the presence of B<sub>12</sub> the *metE* mRNA will form an alternative structure and its transcription or translation should be attenuated. Warner et al. confirmed that 10 µg/mL (7.3 µM) of B<sub>12</sub> supplementation to the culture medium was enough to inhibit the growth of CDC1551 strain by targeting the *metE* B<sub>12</sub> riboswitch. Mutations at two adjacent positions of the *metE* riboswitch region conferred CDC1551 tolerance to B<sub>12</sub> (111). These results together suggest that *Mtb* CDC1551 has to restrict the intracellular level of B<sub>12</sub> in order to survive. Retaining the B<sub>12</sub> riboswitch upstream of *metE* during evolution also indicates that CDC1551 can tolerate B<sub>12</sub> starvation *in vivo*.

Besides MutAB and MetE, *Mtb* genome potentially encodes a third B<sub>12</sub>-dependent enzyme, the Class II ribonucleotide reductase (RNR) NrdZ (Rv0570). Similar to the case of MetE and MetH, there is also a B<sub>12</sub>-independent Class Ib RNR in *Mtb*

which contains the NrdE (Rv3051c) large subunit and NrdF2 (Rv3048c) small subunit. The presence of the Class Ib RNR makes Class II RNR nonessential both *in vitro* and during mice infection (112).

All the genetic redundancy raises the conundrum of whether *Mtb*, not limited to CDC1551 strain, requires B<sub>12</sub> and the B<sub>12</sub>-dependent pathways to survival in humans. Notably, the *mutAB* operon and *rv1322A* (MMCE) were neither essential in mice infection (29) nor upregulated during the infection of INF- $\gamma$  activated macrophages (30). Therefore, it is unlikely that *Mtb* actively uses the B<sub>12</sub>-dependent MMP *in vivo* to assimilate propionyl-CoA, albeit the bacterium has a fully functional MMP.

### I.5.2 Methyl-branched lipid synthesis

#### I.5.2.1 *Mtb* cell wall and the lipid-rich outer membrane

The second pathway that propionyl-CoA can be fed into is the MBL synthesis. MBL synthesis uses (S)- and potentially (R)-methylmalonyl-CoA as building blocks. During MBL synthesis, propionyl-CoA is first converted by PCC to (S)-methylmalonyl-CoA (100, 113), and thereby, this step is shared between MMP and MBL synthesis. Since *Mtb* synthesizes a large quantity of MBLs, they can act as a sink for propionyl-CoA (25, 113). Although the MMP in *Mtb* is not essential and may not even be actively used *in vivo*, the genes encoding the PCC complex are essential (29, 100). This is likely because MBLs are critical for mycobacteria virulence (see below).

The outer membrane of mycobacteria is rich in lipids (Fig. 12a). Besides the C<sub>60-90</sub> mycolic acids that could be covalently attached to the arabinogalactan layer to form

the mycolyl arabinogalactan-peptidoglycan (mAGP) complex, there are also free lipids and glycolipids that noncovalently anchor to the mAGP. Among these free lipids, phthiocerol dimycocerosates (PDIMs) and sulfolipid-1 (SL-1) contain methyl-branched fatty acyl chains (114). PDIMs are unique to the pathogenic mycobacteria including *M. tuberculosis*, *M. marium*, *M. leprea*, *M. ulcerans*, *M. kansasii*, etc. (115). Therefore, the PDIMs have been suggested to be associated with the bacterial virulence. *Mtb* mutants not being able to synthesize or transport PDIMs were highly attenuated *in vivo*. In addition, *Mtb* grown *in vitro* rapidly lost the PDIM due to mutations in the PDIM synthesis and transport genes. These results indicated that PDIM is related to *in vivo* activity (116). The SLs, on the other hand, are unique to *Mtb* (115). But they appeared to be dispensable for *Mtb* virulence *in vivo* (117).

Since the MBL synthesis shares the intermediates (S)- and/or (R)-methylmalonyl-CoA with MMP, given that MBL is essential to *Mtb in vivo*, it is important to understand why only the PCC in MMP is essential. We herein try to rationalize this question from the aspect of MBL biosynthesis.



### I.5.2.2 Stereochemistry in MBL biosynthesis and virulence of *Mtb*

The genes responsible for PDIM biosynthesis are mainly clustered within a 50-kb region of the *Mtb* genome (26, 118). The methyl-branches are introduced into the mycocerosate chains and the phthiocerol backbone of PDIM by three genes, all of which encode type-I polyketide synthases (PKSs). A giant protein complex was identified and characterized by Kolattukudy and colleagues in 1984 (119). The protein machinery was isolated from *M. bovis BCG* cell lysate. This 490-kDa protein complex was a homodimer containing two polypeptide chains each of around 238-kDa. It elongates C<sub>18</sub> or C<sub>20</sub> *n*-fatty acid primers with methylmalonyl-CoA to produce mycocerosic acids. Thus the protein was renamed mycocerosic acid synthase (MAS). MAS specifically uses methylmalonyl-CoA but not malonyl-CoA as a substrate. There are two biological epimers of methylmalonyl-CoA. Experiments showed that MAS could incorporate both the (S)- form and the (R,S)- racemic mixture into the methyl-branched chains (119). But whether the (R)- form in the racemic mixture was incorporated into the lipid was not validated. It is known that *Mtb* preferentially produces laevorotatory ((S)- form) lipids rather than dextrorotatory ((R)- form) ones (120, 121), and PCC complex is a stereospecific enzyme which only synthesizes the (S)-methylmalonyl-CoA (122). These results indicate why the PCC complex is essential in *Mtb*, but MMCE which interconverts the two methylmalonyl-CoA epimers is nonessential.

In addition to mycocerosic acid chains, methylmalonyl-CoA is also incorporated into the phthiocerol backbone of PDIM by two type-I PKSs. Specifically, the phthiocerol backbone is elongated by five PKSs, PpsA-E, using a C<sub>22</sub> *n*-fatty acid as a primer (Fig.

12b). The first three steps of the elongation are catalyzed by PpsA-C using malonyl-CoA. The fourth and fifth steps are catalyzed by PpsD and PpsE which introduce the methyl-branches through the incorporation of methylmalonyl-CoA. After elongation, the product undergoes spontaneous decarboxylation (123) that renders the terminal methylmalonyl moiety non-chiral (Fig. 12b). Therefore, C4 carbon becomes the only methyl group-attached chiral center. It has been shown that in *Mtb*, this chiral center is strictly in the (S)- configuration (120), implicating the dispensability of the nonessential MMCE.

The incorporation of methylmalonyl-CoA into SL-1 is catalyzed by another MAS-like type-I PKS, Pks2 (124). Pks2 elongates the C<sub>16</sub> or C<sub>18</sub> *n*-fatty acids with methylmalonyl-CoA, and the resultant product is known as phthioceranic acid (Fig. 12c). Pks2 can incorporate four to nine methyl-branches into phthioceranic acid, depending on the abundance of methylmalonyl-CoA (113). Genetically disrupting *pks2* abolishes phthioceranic acid formation and consequently the SL-1 synthesis in *Mtb*. However, this did not affect the infection and persistence of *Mtb* in mice or guinea pigs, indicating that SL-1 *per se* is not an essential virulence factor (117, 125). The stereospecificity of Pks2 has not been characterized. Since SL-1 appears nonessential for the survival and virulence of *Mtb*, the stereospecificity may not be a concern.

It has been reported by Jain et al. that the dynamics of the MBL composition was affected by the level of methylmalonyl-CoA, and the MBL composition, in turn, influenced the virulence of *Mtb in vivo* (113). However, they also noticed that through two alternative ways of increasing the methylmalonyl-CoA level, the outcomes were

quite different. Specifically, when propionate was provided as the carbon source, both the molecular weight and abundance of MBLs increased. The increase in the molecular weight was due to an extension of the mycocerosate chain length by MAS. The lipid mass spectrometry showed that the PDIM pattern, in this case, was similar to that of the wildtype *Mtb* isolated from lung infection. Alternatively, if *mutAB* were overexpressed to convert more succinyl-CoA into (R)-methylmalonyl-CoA, the molecular weight of PDIM also increased. However, the abundance of methyl-branched lipids decreased by 30% and 80% for PDIM and SL-1, respectively. Moreover, the bacteria growth was highly attenuated during mice infection (113). Although the author attributed the decreased viability as a disturbance of the metabolism, it is also intriguing to consider the differences in stereochemistry. Propionate supplementation will lead to a direct increase of (S)-methylmalonyl-CoA synthesized by PCC. However, *mutAB* overexpression leads to direct production of (R)-methylmalonyl-CoA from succinyl-CoA. While the efficient incorporation of the (S)- form has been confirmed, the utilization of the (R)- form is not clear (119). Since *Mtb* is a laevorotatory ((S)- form) lipid-synthesizing bacterium (120, 121), its preference of the (S)-methylmalonyl-CoA is likely to have physiological significance.

In summary, propionyl-CoA is not only a nutrient and energy source of *Mtb*, but it is also indispensable for the virulence of the pathogen. The essentiality of PCC complex may be explained by its role in the MBL synthesis, but not in the MMP. It will be of great interest and significance to investigate the relationship between the virulence and stereochemistry of MBLs.

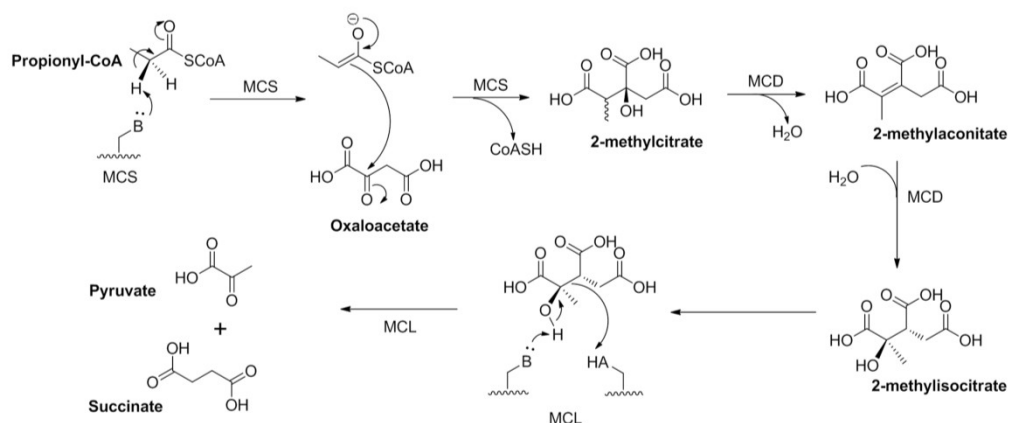


### I.5.3 The methylcitrate cycle

#### I.5.3.1 Enzymes of MCC in general and in *Mtb*

The methylcitrate cycle (MCC) is a third pathway for *Mtb* to assimilate propionyl-CoA. MCC widely exists in prokaryotes but is absent in eukaryotes. The enzymes involved in the pathway are usually encoded within an operon, namely the propionate metabolic operon (*prp* operon). Typically, the operon contains a group of three genes *prpB*, *prpC* and *prpD*. Some organisms, i.e., *Salmonella enteric*, contain a fourth gene encoding a propionyl-CoA synthetase, PrpE, which prepares the substrate for MCC (126).

The cycle starts with an aldol condensation reaction catalyzed by methylcitrate synthase (MCS, PrpC) (Fig. 13). Analogous to citrate synthase in the TCA cycle, MCS condenses propionyl-CoA with oxaloacetate and forms 2-methylcitrate (2-MC). In the second step, 2-MC is isomerized into 2-methylisocitrate (2-MIC). The isomerization could be catalyzed by a [4Fe4S] cluster-containing aconitase-like enzyme AcnD coupled with a hydratase PrpF. These two enzymes in some organisms, including *Mtb*, are substituted by a single enzyme methylcitrate dehydratase (MCD, PrpD) (126). PrpD is iron-sulfur-free and oxygen-tolerant. In the third step, 2-MIC is cleaved by methylisocitrate lyase (MCL, PrpB) via a retro-aldol reaction to form pyruvate and succinate. The two products can be used in anabolism or fed into the TCA cycle.



**Figure 13. Consensus enzyme catalytic mechanisms of the methylcitrate cycle**  
**The residues are generalized as HA for a catalytic acid and B for a catalytic base.**

In *Mtb*, the *prp* operon only contains *prpC* (*rv1131*) and *prpD* (*rv1130*) without the methylisocitrate lyase gene *prpB*. This is because the isocitrate lyase 1 (Icl1, Rv0467) which commonly functions in the glyoxylate shunt also possesses MCL activity (127). Therefore, Icl1 is a bifunctional enzyme in *Mtb*. Notably, *Mtb* genome encodes a second copy of isocitrate lyase (Icl2). Deletion and complementation experiments suggested that the activity of Icl2 against 2-MIC was poor and Icl1 contributed to more than 90% of the MCL activity (25).

#### I.5.3.2 Toxicity of propionyl-CoA and related metabolic intermediates

Propionyl-CoA has been heavily documented to create cytotoxicity if it cannot be properly metabolized. Munoz-Elias et al. observed that propionyl-CoA accumulation could cause *Mtb* growth inhibition (25). They showed that *Mtb* deleting *prpDC* operon failed to grow on 0.1% propionate or valerate as the sole carbon source. Valerate

degradation could produce equal mole of propionyl-CoA and acetyl-CoA, and assimilation of the latter should not be impaired in the  $\Delta prpDC$  bacilli. Therefore, the authors concluded that the accumulation of propionyl-CoA was the cause of growth inhibition (25), and the role of MCC was to detoxify propionyl-CoA (128). However, VanderVen et al. suggested that propionyl-CoA *per se* was not toxic. They showed that *Mtb* deleting both *prpC* and *icl1* was able to grow on cholesterol or propionate-containing rich media, where propionyl-CoA was produced and accumulated. However, if only *icl1* was deleted, where propionyl-CoA could be converted into 2-MC or 2-MIC without further degradation, *Mtb* growth on the rich media was strongly inhibited. Therefore, they inferred that the metabolic intermediates, likely the 2-MC or 2-MIC, were toxic to *Mtb* (129). Eoh and Rhee applied metabolomics technique to potentially unravel a bigger picture of the propionate toxicity (130). They observed that the toxicity of propionate or propionyl-CoA was multifaceted. In *icl1* deleted *Mtb*, propionyl-CoA could deplete the TCA cycle intermediate oxaloacetate and route the metabolism into a “dead-end” MCC pathway. This further resulted in the depletion of gluconeogenesis and TCA cycle substrates as well as amino acid synthesis precursors, accumulation of methylmalonyl-CoA, and shifting the  $NAD^+/NADH$  ratio and membrane potential (130). However, the propionate or cholesterol inhibition was also observed in OADC-based rich media where nutrition and central carbon metabolic substrates are plenty, and substrate depletion should not be a concern (129).

Propionyl-CoA-related toxicity has also been observed in other organisms. In *E. coli*, propionyl-CoA was shown to bind to the TCA cycle enzyme citrate synthase with

an affinity comparable to that of the true substrate acetyl-CoA. Nevertheless, *E. coli* citrate synthase had no detectable activity against propionyl-CoA, which made propionyl-CoA a potential inhibitor of *E. coli* citrate synthase (131). In addition, propionyl-CoA could inhibit several CoA-ester involved enzymes in fungi, among which pyruvate dehydrogenase was most strongly inhibited (132). Propionyl-CoA-related toxicity was probably best studied in *S. enterica*. Unlike *E. coli* citrate synthase, the *S. enterica* citrate synthase appeared to be able to synthesize 2-MC. However, the 2-MC synthesized by *S. enterica* citrate synthase was more toxic than synthesized by methylcitrate synthase PrpC (133, 134). Knocking down citrate synthase alleviated the cytotoxicity (133). Since citrate does not have a chiral center but 2-MC has two chiral centers, it is possible that citrate synthase may synthesize a mixture of 2-MC diastereomers, while PrpC may synthesize a stereospecific 2-MC. Further studies identified the enzyme target of the toxic 2-MC in *S. enterica*  $\Delta prpC$  strain to be the gluconeogenesis enzyme fructose-1,6-bisphosphatase (FBPase) (134). Overexpression of FBPase relieved the toxicity. Moreover, a single mutation S123F in FBPase made the  $\Delta prpC$  *S. enterica* resistant to propionate (134).

The toxicity of 2-MC against *Mtb* FBPase was performed by Eoh and Rhee (130), who showed that 2-MC was a non-competitive inhibitor of FBPase with an  $IC_{50}$  value below 1 mM. But since the commercially available 2-MC is typically a racemic mixture of diastereomers, the true inhibitory constant was probably well below 1 mM. It will be very intriguing and necessary to find out the configuration of 2-MC produced by *Mtb* PrpC and the two citrate synthases (Rv0889c/CitA and Rv0896/GltA2). At present, the

reason of propionyl-CoA-related toxicity is unknown, and the toxicity of the intermediates appears to be the most satisfactory explanation.

#### I.5.3.3 Role of MCC *in vivo*

Paradoxically, although the *prpDC* operon was required for the *in vitro* growth of *Mtb* on propionate (25, 104, 135), odd-chain fatty acids (25) or cholesterol (104, 129) containing media and in murine bone marrow-derived macrophages (25, 104), the operon was dispensable for *Mtb* infection and persistence in mice (25). On the contrary, *icl1* deleted *Mtb* was not viable *in vivo* or on the media containing acetate, propionate, fatty acids or cholesterol (25, 96, 105, 130). Since Icl1 is bifunctional which also participate in the glyoxylate shunt, it is conceivable that Icl1 performs a pivotal role in central carbon metabolism more than detoxifying MCC intermediates. Nevertheless, the role of Icl1 on the MCC is not negligible. Notably, *prpDC* and *icl1* genes were highly upregulated during resting or INF- $\gamma$  activated macrophage infections. (30) Therefore, it is likely that the overexpression of *prpDC* during infection is enforced regardless of its essentiality. A large amount of 2-MC and 2-MIC will consequently be generated, which requires an intact and fully functioning Icl1 to detoxify the MCC intermediates.

MCC is a double-edged sword for *Mtb* growth *in vivo*. On the one hand, MCC is crucial for assimilating the high energy propionyl-CoA which is massively produced during cholesterol and odd-chain fatty acid degradation. On the other hand, MCC may lead to the production and accumulation of toxic intermediates. As a result, MCC needs to be precisely regulated. The regulation at the transcriptional level is achieved by two

local regulators, PrpR and RamB, which are encoded neighboring to the *prpDC* operon and *icl1*, respectively. In the next chapter, we will investigate the structure and biochemical properties of these regulators and unravel the regulatory mechanism of the MCC pathway.

CHAPTER II  
SHORT-CHAIN FATTY ACYL COENZYME A DERIVATIVES REGULATE  
*MYCOBACTERIUM TUBERCULOSIS* METHYLCITRATE CYCLE AND  
GLYOXYLATE SHUNT

## II.1 Introduction

*Mycobacterium tuberculosis* (*Mtb*) remains a major global health problem. In 2016, *Mtb* led to an estimated 1.3 million deaths among HIV-negative people, and 6.3 million new tuberculosis (TB) cases were reported (136). It is well accepted that *Mtb* utilizes fatty acids and cholesterols as its primary source of nutrition when infecting humans (22, 34, 105). A common metabolite from the catabolism of both odd-chain fatty acids and cholesterols is propionyl coenzyme A (propionyl-CoA) (57, 71, 73, 96, 104). It has been shown that the accumulation and metabolism of propionyl CoA leads to the toxicity of the bacilli (128, 130).

There are several major pathways that utilize propionyl-CoA including the methylmalonyl pathway (MMP), methyl-branched chain fatty acids synthesis and the methylcitrate cycle (MCC) (Fig. 10). In the MMP, propionyl-CoA is sequentially converted to (S)- and (R)-methylmalonyl-CoA, which is then rearranged to the TCA cycle intermediate succinyl-CoA by a vitamin B<sub>12</sub>-dependent enzyme methylmalonyl-CoA mutase (96). Since *Mtb* genome does not encode *cobF*, an essential enzyme in the de novo B<sub>12</sub> synthesis (108, 137), it leads to the argument that the bacilli is not able to synthesize B<sub>12</sub>, and thus has to acquire B<sub>12</sub> from the media (96) or scavenge from the

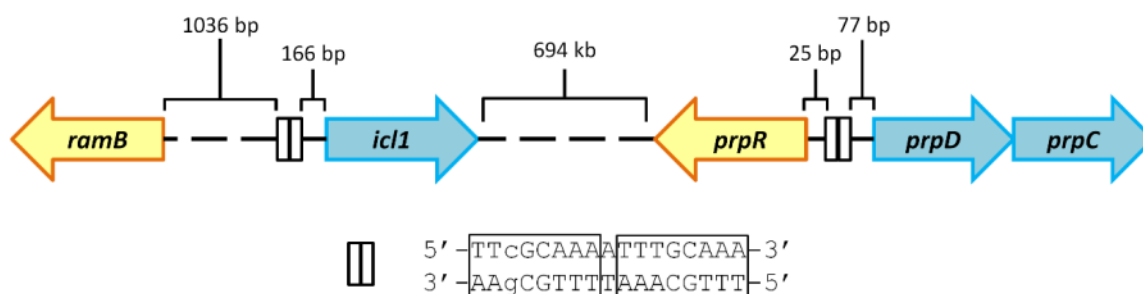
host (110). An alternative route to bypass the B<sub>12</sub>-dependent step is to incorporate methylmalonyl-CoA into the methyl-branched lipids pool (113). The other major pathway to catabolize propionyl-CoA is the MCC (25). Although MCC is dispensable in a mouse infection (25), it is essential for *Mtb* growth in the macrophage (25, 104), and thus is still considered as a candidate pathway for inhibitor development (129). The key enzymes of the MCC include methylcitrate synthase (MCS, also named PrpC), methylcitrate dehydratase (MCD, also named PrpD), and methylisocitrate lyase (MCL, also named PrpB in some organisms, i.e. *M. smegmatis*) (25, 128). In *Mtb*, the former two genes are encoded in the propionate (*prp*) metabolic operon. However, the *Mtb* genome does not encode a functionally unique MCL, instead its function is accomplished by isocitrate lyase 1 (Icl1), a well known enzyme in the anaplerotic glyoxylate shunt (25, 127, 130).

In *Mtb*, the transcription of the *prp* operon was proposed to be mediated by the master transcription sigma factor E (SigE) (138) and was stringently controlled by a local transcriptional regulator, PrpR (propionate regulator, Rv1129c, MtPrpR) (104, 135, 138). The *prpR* gene is adjacent to, but in the opposite direction of, the *prp* operon (Fig. 14). MtPrpR activates the transcription of both itself and the *prp* operon in response to propionate or cholesterol (104, 135). Deleting *prpR* from the *Mtb* genome made the bacterium unable to respond to propionate or cholesterol *in vitro* and fails to grow on media containing these molecules as sole carbon sources (104, 135). Interestingly, *prpR* deficient mutants have been found to be relatively prevalent in isoniazid resistant, as well as other drug resistant *Mtb* clinical isolates (139, 140), possibly because these

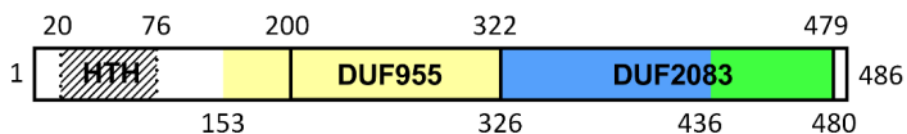


mutations slow down bacterial growth as a fitness cost of drug tolerance (140). These paradoxical phenomena intrigue us to understand the molecular characteristics and regulatory mechanism of this protein.

The regulatory mechanisms of MtPrpR and its homologs at the molecular level have not been resolved in any organism due to the paucity of structural information of these proteins. While we still do not understand how MtPrpR is stimulated, 2-methylcitrate was suggested as a coactivator of PrpR in different organisms (141-143). cAMP has also been suggested to be indirectly involved in the response (129, 144). Conserved domain analysis (145) was only able to annotate the N-terminus of MtPrpR as the helix-turn-helix DNA binding domain (Fig. 15). The rest of the protein was previously annotated as two domains of unknown function, DUF955 and DUF2083 (138). DUF955 was recently updated to be an IrrE N-terminal-like domain, which was reported as a Zn-dependent metalloprotease (146, 147).



**Figure 14. Genome organization of *ramB-icl1* regulon and the *prpR-prpDC* regulon. Dashed lines indicate the presence of genes in between. The distances of the intergenic regions are indicated. The tandem repeats are zoomed in with the sequence spelled out. Lower case indicates the imperfect palindrome.**



**Figure 15. Domain organization of MtPrpR.**

**Black blocks and the numbers above indicate the predicted domain borders; color-filled blocks and the numbers below indicate the domains observed in the structure. The D<sub>II</sub> domain is colored in light yellow; GAF domain is colored in blue; C-terminal region is colored in green. The HTH domain (hatched) was truncated from the protein construct.**

Notably, in *Mtb* there is a *prpR* paralog, *ramB* (regulator of acetate metabolism B, *rv0465c*), located upstream of *icl1* (Fig. 14). Unlike MtPrpR, which is largely accepted as a transcriptional activator (104, 135, 138), MtRamB is more likely to be a transcriptional repressor in the presence of a fermentable carbon source such as dextrose (138, 148).

A 17 bp length of DNA between *prpR* and *prpD* has been proposed as the primary recognition sequence of MtPrpR (135) (Fig. 14). It consists of a perfect (TTTGCAA) and an imperfect (TTTGCgAA) palindrome separated by one base pair. The exact 17 bp DNA sequence also exists upstream of *icl1* but nowhere else in the *Mtb* genome. In addition, very similar sequences have been mapped to MtRamB binding sites (148-150). This genome context may provide a precise regulatory mechanism, and cross-regulation by MtPrpR and MtRamB might also exist. To unravel the potential structure-function relationship of MtPrpR and its homologs, we solved the crystal structures of different forms of MtPrpR. The structure revealed that MtPrpR is an iron-sulfur protein that binds to Coenzyme A or its derivative molecules tightly, which may be a clue to its

regulatory mechanism. MtPrpR and MtRamB showed a concerted and carbon source dependent regulation upon *icl1* transcription.

## II.2 Materials and Methods

### II.2.1 Bacterial strains, media, and growth conditions

*M. tuberculosis* strains were maintained in complete 7H9 media (Middlebrook 7H9 salts supplemented with 0.2% glycerol, 0.05% Tween-80 and 10% Middlebrook OADC). The *prpR* deletion mutant of *M. tuberculosis* H37Rv (H37Rv $\Delta$ *prpR*) was obtained from the Sassetti lab as described previously (104). *PrpR* deletion mutants carrying a rescuing wildtype or mutant allele were then constructed as described previously (140). Briefly, a kanamycin resistant plasmid carrying the *prpR* wildtype or mutant gene and 95 bp of the sequence upstream of the proposed translational start site was integrated at the L5 phage integration site of the *prpR* deletion strain.

Knockdown of *ramB* was performed using the CRISPR interference (CRISPRi) system optimized for *M. tuberculosis* (151) targeting the sequence GCGGGCAGGTGCCCGCTGGA within the *ramB* ORF. The anhydrous tetracycline (ATc) inducible *ramB* knockdown vector was transformed into both wildtype H37Rv and the H37Rv $\Delta$ *prpR* mutant strains.

### II.2.2 Cloning, protein expression, and purification

The gene *Rv1129c* (*prpR*) encoding the full-length PrpR protein (MtPrpR\_M1) was cloned as previously described (135). Briefly, the gene was amplified from the *Mtb*

H37Rv genome using PrpR\_M1\_Fw and PrpR\_Rv primers. (All primers were synthesized by Integrated DNA Technologies, and are listed in Table 2.) The PCR product was treated with the BamHI and XhoI restriction enzymes (New England Biolabs) and cloned into a pET28a vector (Novagen).

A re-annotation of *Mtb* ORF (152) suggested that the ninth amino acid valine might be the translational start site. *PrpR\_V9* construct removing first eight amino acids from the N-terminus was subcloned into a pET28a vector between the BamHI and XhoI sites using PrpR\_V9\_Fw and PrpR\_Rv primers. Both the MtPrpR\_M1 and MtPrpR\_V9 were expressed in *Escherichia coli* BL21(DE3). The cells were grown in LB media supplemented with 50 µg/ml kanamycin to an OD<sub>600</sub> of 0.8, induced with 0.08 mM IPTG, and shaken at 18 °C for another 12-16 hours before being pelleted. The cells were resuspended in Lysis Buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.5, 200 mM NaCl, 30 mM imidazole), homogenized using a Microfluidizer M-100P (Microfluidics). Cell debris was clarified and the supernatant was applied to a Ni-NTA resin (Qiagen). The resin was washed with 20 column volumes of Wash Buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.5, 200 mM NaCl, 60 mM imidazole). An isocratic elution was performed using Elution Buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.5, 200 mM NaCl, 300 mM imidazole).

**Table 2. Primers used in this study**

<b>Cloning</b>	
PrpR_M1_Fw	5'-GGTCTAGAGGGATCCATGACGCGGAGTAATGTCTTA-3'
PrpR_V9_Fw	5'-CGCGGATCCGTGGCTAGGACGTATTCGAG-3'
PrpR_Rv	5'-GGGGTACCCTCGAGTCAAACCGATTGCTCGGTCTGA-3'
PrpR_D81_Fw	5'-TACTTCCAATCCAATGCCGACGCACGGTTGGTGG-3'
PrpR_V486_Rv	5'-TTATCCACTTCCAATGTTAAACCGATTGCTCGGT-3'
PrpR_F155_Fw	5'-TACTTCCAATCCAATGCCTTCGAGGAGGTCCGCGAC-3'
PrpR_E440_Rv	5'-TTATCCACTTCCAATGTTACTCCGTGCTCGGGTCTGTC-3'
PrpR_F240A_Fw	5'-CCCGGACAGCGCGCCGCCAGATCGCCACCCAA-3'
PrpR_F240A_Rv	5'-TTGGGTGGCGATCTGGGCGGCGCGCTGTCCGGG-3'
PrpR_H319A_Fw	5'-TTCGAAACCGTCTGCGCCCGGCTCTCCACACTG-3'
PrpR_H319A_Rv	5'-CAGTGTGGAGAGCCGGGCGCAGACGGTTTCGAA-3'
PrpR_F155H_Fw	5'-GCGCGGCCGATGCCGACGAGGAGGTCCGCGAC-3'
PrpR_F155H_Rv	5'-GTCGCGGACCTCCTCGTGCGGCATCGGCCGCGC-3'
PrpR_F155A_Fw	5'-GCGCGGCCGATGCCGCGCCGAGGAGGTCCGCGAC-3'
PrpR_F155A_Rv	5'-GTCGCGGACCTCCTCGGCCGGCATCGGCCGCGC-3'
PrpR_F155W_Fw	5'-GCGCGGCCGATGCCGTGGGAGGAGGTCCGCGAC-3'
PrpR_F155W_Rv	5'-GTCGCGGACCTCCTCCCACGGCATCGGCCGCGC-3'
PrpR_F155Y_Fw	5'-GCGCGGCCGATGCCGTATGAGGAGGTCCGCGAC-3'
PrpR_F155Y_Rv	5'-GTCGCGGACCTCCTCATAACGGCATCGGCCGCGC-3'
PrpR_C363A_Fw	5'-CGGGTCGGCGGCAGCGCCCCGCTGTGGGTGGTC-3'
PrpR_C363A_Rv	5'-GACCACCCACAGCGGGGCGCTGCCGCCGACCCG-3'
PrpR_C450A_Fw	5'-GCGGGCTGCAAGATCGCCAACCGAACGTCGTGC-3'
PrpR_C450A_Rv	5'-GCACGACGTTTCGGTTGGCGATCTTGCAGCCCGC-3'
p1129c_Fw	5'-GACGTCAACCGGATCGGCAGC-3'
p1129c_Rv	5'-GGCACCGGAAAACGTCTCTCA-3'
<b>RT-qPCR for <i>M. tuberculosis</i></b>	
prpR_Fw	5'-ATGTCAACCAGTTGGAGAATGA-3'
prpR_Rv	5'-CGGAATCCGAGGAGAAATACTG-3'
prpD_Fw	5'-TGACTTTCACGACACGTTTCTGGC-3'
prpD_Rv	5'-TGTGGATCTCATAGGCGGTTACCA-3'
icl1_Fw	5'-CAGCACATCCGCACTTTGAC-3'
icl1_Rv	5'-ATCACCACCGTGGGAACATC-3'

**Table 2. Continued**

<b>RT-qPCR for <i>M. tuberculosis</i></b>	
ramB_Fw	5'-CCTGCCCGCTGTGGAA-3'
ramB_Rv	5'-GGCGATTTGCACCAAGATCT-3'
sigA_Fw	5'-CAAGTTCTCCACCTACGCTAC-3'
sigA_Rv	5'-GTTGATCACCTCGACCATGT-3'
<b>RT-qPCR for <i>M. smegmatis</i></b>	
rv1129_Fw	5'-GTTACGCTATGACATCGACCTG-3'
rv1129_Rv	5'-GAAGATGAACGGTATCCCTCG-3'
MsprpR_Fw	5'-CACCCAGTCAGAGCAGTTATC-3'
MsprpR_Rv	5'-CGAATAGATCAGCCGGTGAG-3'
MsprpD_Fw	5'-AACAAGTTCGTCACCCTGG-3'
MsprpD_Rv	5'-ACTTGTCCAACCCGTCCTTG-3'
MsprpB_Fw	5'-AAGTCTTCGACGCTGGAAC-3'
MsprpB_Rv	5'-GAAGTTGTAGATGTCCGAGTCG-3'
MsprpC_Fw	5'-AGGAGAAGGTCATGGGATTTG-3'
MsprpC_Rv	5'-TTGTAGATGTCCGAGCCAACG-3'
MssigA_Fw	5'-GACTACACCAAGGGCTACAAG-3'
MssigA_Rv	5'-TTGATCACCTCGACCATGTG-3'
<b>Binding Oligos</b>	
Pos_Pal30mer	5'-GCACATTTTGCAAATTTTGCGAACATGACG-3'
Neg_Pal30mer	5'-CGTCATGTTTCGAAAATTTGCAAATGTGC-3'
Pos_HalfPal30mer	5'-GCACATTTTGCAA T CCTGTGTACATGACG-3'
Neg_HalfPal30mer	5'-CGTCATGTACACAGGATTTGCAAATGTGC-3'
Pos_/Cy3/Pal30mer	5'-/Cy3/GCACATTTTGCAAATTTTGCGAACATGACG-3'

The protein truncations were determined by domain and secondary structure analysis. The domain annotation was performed in InterPro (145) (<https://www.ebi.ac.uk/interpro/>). Protein secondary structure was predicted by PSIPRED (153) (<http://bioinf.cs.ucl.ac.uk/psipred/>)

*PrpR*<sub>81-486</sub> truncating the N-terminal helix-turn-helix domain was subcloned into a ligation-independent cloning (LIC) vector pMCSG7 (154) with primers PrpR\_D81\_Fw and PrpR\_V486\_Rv. This construct contains a His<sub>6</sub>-tag and a TEV cleavage site to the N-terminus of the target protein. The plasmid was transformed into *E. coli* BL21(DE3) competent cells. The cells were grown at 37 °C in LB medium supplemented with 100 µg/ml carbenicillin to an OD<sub>600</sub> of 0.8. Protein expression was induced by 0.3 mM IPTG. Cells were grown at 18 °C for another 12-16 hours before being pelleted. Protein purification was analogous to that of the full-length proteins except that 20 mM Tris-HCl pH7.8 was used instead of 100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.5 in all the buffers. The salt and imidazole concentrations remained unchanged. PrpR<sub>81-486</sub> was about 95% pure according to SDS-PAGE analysis.

The protein was diluted in IEX Buffer A (20 mM Tris-HCl pH 7.8) until the NaCl concentration was below 100 mM and was loaded to an anion exchange HiTrap Q FF column (GE Healthcare). The column was washed by mixing the IEX Buffer A with ten column volumes of 20% IEX Buffer B (20 mM Tris-HCl pH 7.8, 1 M NaCl). An isocratic elution was performed with 30% IEX Buffer B.

Protein was concentrated immediately to approximately 30 mg/ml and injected onto a HiLoad 26/600 Superdex-200 column (GE Healthcare) equilibrated with Gel

Filtration Buffer (20 mM Tris-HCl pH 7.8, 100 mM NaCl, 3 mM DTT). The majority of the protein appeared as a tetramer in solution. Peak fractions were pooled and concentrated to 24 mg/ml for crystallization trials.

MtPrpR<sub>81-486</sub> point mutations were introduced using the primers listed in Table 2 and the QuikChange site-directed mutagenesis kit (Agilenet). The F155H variant was expressed and purified in particular following the same procedure for the wildtype protein.

Selenomethionine-derived MtPrpR<sub>81-486</sub> was expressed by growing the bacteria in M9 minimal salt media supplemented with 0.4% glucose, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 100 µg/ml carbenicillin. Cells were grown to an OD<sub>600</sub> of 0.6 and chilled on ice. 100 mg l<sup>-1</sup> each of L-lysine, L-threonine, L-phenylalanine, 50 mg l<sup>-1</sup> each of L-leucine, L-isoleucine, L-valine, and 60 mg l<sup>-1</sup> L-selenomethionine, were added before induction along with 100 mg/l ferrous ammonium sulfate. Cells were shaken at 18 °C for 15 min, and protein expression was induced with 0.3 mM IPTG for 12-16 hours. The selenomethionine-derived protein purification procedure was identical to that of the native protein and its behavior was also very similar to that of the native protein.

The transition metal ion evaluations was performed by growing the *E. coli* cells in M9 minimal salt media supplemented with 0.4% glucose, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 100 µg/ml carbenicillin to an OD<sub>600</sub> of 0.8. Each flask was supplied with 25 µM one of the chloride salts of Fe<sup>3+</sup>, Co<sup>2+</sup>, Ni<sup>2+</sup>, Cu<sup>2+</sup> and Zn<sup>2+</sup> 15 min before induction. Protein expression was induced with 0.3 mM IPTG for 16 hrs at 18 °C. The total protein



expression in the whole cell lysate and the soluble protein in the supernatant were analyzed on a SDS PAGE.

*PrpR*<sub>155-440</sub> was subcloned into a pMCSG7 vector using primers PrpR\_F155\_Fw and PrpR\_E440\_Rv. Procedures for protein expression and affinity purification were similar to those of MtPrpR<sub>81-486</sub> except that all purification buffers contain 500 mM NaCl. The protein was dialyzed overnight against 20 mM Tris-HCl pH7.8, 500 mM NaCl while the His<sub>6</sub>-tag was cleaved with TEV protease. The protein was reapplied to the Ni-NTA resin to remove the His<sub>6</sub>-tag and the residual impurities. The protein was concentrated to about 10 mg/ml and injected to a HiLoad 26/600 Superdex-75 column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl pH 7.8, and 400 mM NaCl. The protein appeared as a monomer in solution. The peak fractions were pooled and concentrated to approximately 10 mg/ml.

Analytical size exclusion chromatography (SEC) for each protein construct was performed using the Superdex 200 10/300 GL (GE Healthcare). SEC buffer for each protein construct was identical to their preparative SEC buffer.

### II.2.3 *Crystallization*

Freshly purified His-tagged MtPrpR<sub>81-486</sub> or MtPrpR<sub>81-486</sub>\_F155H variant proteins at 24 mg/ml was screened against over 900 crystallization conditions in sitting-drop vapor diffusion using a Mosquito Crystal liquid handler (TTP Labtech Inc). Protein was equilibrated with the crystallization conditions in 17 °C. Brown color crystals appeared

within two days. The best growth condition contained either 100 mM HEPES pH 7.5 or 100 mM Bicine pH 9.5, 1 M sodium acetate, 0.08 mM Zwittergent 3-14.

The MtPrpR<sub>155-440</sub> protein was screened against over 900 conditions at about 10 mg/ml in the presence and absence of 1 mM CoA. Crystals of identical morphology for both formulations appeared in about one week in 0.2 M CaCl<sub>2</sub>•2H<sub>2</sub>O, 0.1 M HEPES pH 7.5, 28% v/v PEG 400.

#### II.2.4 *Data collection and structure determination*

Crystal X-ray diffraction data were collected at Beamline 19-ID, 19-BM, or 23-ID at the Advance Photon Source, Argonne National Laboratory in Chicago. Anomalous scattering for selenomethionine derived MtPrpR<sub>81-486</sub> was collected around the peak position at 0.97949 Å. (We planned for iron phasing and selenium phasing, but were informed that the beam at low energy was not stable at the time.) The crystal diffracted to approximately 2.7 Å resolution. Crystal symmetry and diffraction intensity were analyzed with Denzo in HKL2000, scaled and reduced with Scalepack in HKL2000(155). The selenomethionine derived crystal belonged to the P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> space group with a = 96.19 Å, b = 142.37 Å, c = 145.97 Å,  $\alpha = \beta = \gamma = 90^\circ$ , and contained four molecules per asymmetric unit (ASU) in D<sub>2</sub> symmetry. Experimental phasing was achieved using single-wavelength anomalous dispersion (SAD). Heavy atom sites were searched in SHELXD (156) (SHELXC/D/E pipeline in CCP4 (157)). Seventeen out of 28 possible heavy atom sites were found, all of which were correct when compared to the final structure. The electron density map was calculated with SHELXE(158),

followed by density modification with Parrot (159) in CCP4 by applying non-crystallographic symmetry (NCS) calculated from the heavy atom substructure. Automated initial model building plus refinement was conducted using Buccaneer (160) in CCP4. Manual model building was performed in Coot (161). The [4Fe4S] cluster and Coenzyme A were built into the electron density in all four chains in the ASU.

The native PrpR<sub>81-486</sub> crystal was collected to 2.5 Å and had slightly different unit cell dimensions. The crystal was indexed to a higher symmetry of the P4<sub>1</sub>2<sub>1</sub>2 space group with a = b = 144.88 Å, c = 97.77 Å. In this case, a tetramer was formed with two molecules relating to the other two in twofold crystallographic symmetry. Since the native crystal is non-isomorphous to the heavy atom derived one, a molecular replacement search was performed in Phaser-MR (162) in PHENIX (163) using one polypeptide chain from the Se-PrpR<sub>81-486</sub> tetramer. Refinement was carried out using phenix.refine (164), by randomizing the B-factor of the MR solution and applying simulated annealing to remove the trace of refinement of the initial model. Real space refinement was performed in phenix.refine and Coot. Fifteen TLS groups were defined by phenix.refine and applied in the refinement. X-ray/stereochemistry and X-ray/ADP weights were optimized in the final cycles of refinement to improve the refinement statistics.

The MtPrpR<sub>81-486</sub>\_F155H variant crystal diffracted to 2.7 Å, and the unit cell was isomorphous to the native PrpR<sub>81-486</sub>. The structure of the wildtype protein was directly refined against the mutant data by transferring the test set from the wildtype data. The process of refinement was analogous to that of the wildtype protein.

The MtPrpR<sub>155-440</sub> crystal diffracted to about 1.8 Å and belonged to the I222 space group with  $a = 75.03$ ,  $b = 81.71$ ,  $c = 95.37$ ,  $\alpha = \beta = \gamma = 90^\circ$ . A molecular replacement search was performed in Phaser-MR using one molecule of MtPrpR<sub>81-486</sub> trimmed to appropriate positions as the search model. The manual model building and refinement were performed in Coot and phenix.refine analogous to that of the native MtPrpR<sub>81-486</sub>.

The interface analysis was performed with PISA (165). The 3D structure similarity search was performed in Dali (165) (<http://ekhidna2.biocenter.helsinki.fi/dali/>). The structure figures were generated using UCSF Chimera (166) (<http://www.rbvi.ucsf.edu/chimera>).

### II.2.5 *Iron-sulfur cluster identification and redox characterisation*

The iron-sulfur cluster was initially identified with the UV-visible spectroscopy. MtPrpR<sub>81-486</sub> protein was diluted to 30 µM and transferred to a quartz cuvette of 1 cm path length. The UV-visible spectra were recorded from 200 nm to 800 nm using a Cary 50 UV-Vis spectrophotometer (Varian/Agilent Technologies, Santa Clara, CA, USA). A single peak was observed at 410 nm, indicating the presence of the [4Fe-4S] cluster. Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) was conducted using an ELAN DRC II ICP Mass Spectrometer (Perkin Elmer) to further confirm the identity and the stoichiometry of the metal ion. Iron standard was diluted in 1% HNO<sub>3</sub> to 25, 50, 100, 150, 200, and 250 ppb, and plotted as the standard curve. The protein sample (10 mg/ml) was denatured and 200-fold and 400-fold diluted in 1%

HNO<sub>3</sub> to around 1.04 and 0.52 μM and centrifuged to remove the protein and was subjected into the ICP-MS system. The iron concentration was calculated based on the standard curve.

The H<sub>2</sub>O<sub>2</sub> oxidation test was performed by mixing 20 μM of purified MtPrpR<sub>81-486</sub> with different concentrations of H<sub>2</sub>O<sub>2</sub> (Sigma-Aldrich) in the SEC buffer. The property of iron-sulfur cluster was monitored optically using the UV-visible spectrometry at indicated time points.

The NO response was performed as described before (167). Briefly, PROLI NONOate (Cayman Chemical), the NO donor compound, was dissolved in 25 mM NaOH. And it was introduced into 50 μM of MtPrpR<sub>81-486</sub> in 20 mM Tris-HCl pH 7.8, 300 mM NaCl at indicated concentrations. Each molecule of PROLI NONOate can release two molecules of NO ( $t_{1/2}$ =13 s at 25 °C) (168). The UV-visible spectrum was measured after 10 min incubation.

### II.2.6 Coenzyme A derivative isolation and identification

The CoA derivative isolation procedure was performed at 4 °C to maintain the integrity of the molecules. 100 mg of 25 mg/ml of the wildtype MtPrpR<sub>81-486</sub> or the F155H variant was denatured in 10 volumes of 8 M urea acidified with 0.1 M formic acid to stabilize the potential thioester bond. The urea was freshly made and purified with mixed bed resin (Sigma Aldridge) to minimize the content of isocyanate and other ions. Denatured protein solution was passed through a 30 kDa cut-off concentrator to remove the protein. Flow-through was collected and loaded onto Q fast flow sepharose

anion exchange resin (GE Healthcare Life Sciences). The resin was washed with five column volumes of 0.1 M formic acid, and eluted with two column volumes of 2 M ammonium formate in 0.1 M formic acid. The eluate was lyophilized and dissolved in 0.2 ml pure water for identification. The LC-MS analysis was adapted from previous research on CoA derivatives. Specifically, a 10  $\mu$ L sample solution was injected into a *Kinetex* 2.6  $\mu$ m EVO C18 100A LC column 100 x 4.6 mm and separated by a mobile phase gradient elution of 50 mM ammonium acetate (A) and acetonitrile + 0.1% formic acid (B) at a flow rate of 0.5 mL/min. An Agilent 1200 Series LC system was used to start (B) at 10% and increase linearly to 100% from 0 to 8 minutes, hold (B) at 100% from 8–12 minutes, decrease linearly from 100% to 10% from 12–14 minutes, and held at 10% for an additional three minutes (total run time 17 minutes). A Bruker Daltonics micrOTOF-Q II LC-MS with an electrospray ionization system was operated in positive mode, at a source temperature of 220 °C, dry nitrogen gas flow of 11 l/min and under 3.5 bars of pressure. A 0.4 mg/ml sodium formate external calibration solution was directly injected at 13 minutes for post data analysis and calibration. Quantification was based on integrating peak area corresponding to the elution of the target compound in the extracted product ion chromatograms. Data was processed using Bruker Hystar Software 4.1 and extracted ion chromatograms were created for each target compound.

### II.2.7 Protein sequence alignment and analysis

MtPrpR and a group of representative homologous protein sequences were aligned in Clustal Omega (169) (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The output

alignment file was further analyzed and illustrated using ESPript 3.0 (170) (<http://esprict.ibcp.fr/ESPript/ESPript/index.php>). The proteins with names are those whose functions are either reported before, or recognizable based on the genome context. The proteins with question marks are those with unknown functions and adjacent to uncharacterized genes or operons.

### II.2.8 *Transcription level quantification*

Quantification of gene expression was performed using reverse transcription, quantitative PCR (RT-qPCR) as describe previously (140). *PrpR* variant strains were grown to log phase in 7H9 complete media, spun down, and resuspended to an OD<sub>600</sub> of 0.1 in 10 ml of 7H12 media (7H9 salts, 0.2% glycerol, 0.1% casamino acids, and 0.05% tyloxapol) with 0.02% of the indicated short chain fatty acid. After two days of exposure, RNA was isolated using the Direct-Zol RNA miniprep kit (Zymo Research) and cDNA was generated using Superscript IV reverse transcriptase (Invitrogen) with 200-300 ng of total RNA as a template. The transcription of *prpD*, *prpR icl1*, and *sigA* was measured with the primers included in Table 2. Transcriptional level of *prpD*, *prpR*, and *icl1* was quantified using the delta-delta CT method using *sigA* as a housekeeping gene to normalize for input. Expression of the *prpR* wildtype complement in acetate was set to one in each case.

In the *ramB* knockdown experiments, wildtype H37Rv and the H37Rv $\Delta$ *prpR* mutant strains transformed with *ramB* knockdown vector were cultured in acetate or propionate media as describe above simultaneously with 100ng/ml ATc where indicated

to induce knockdown. RNA was collected after 48 hours of media exposure and induction, and was quantified as described above.

### II.2.9 Recognition DNA preparation and DNA-binding assay

The 30 bp duplex DNA pieces were prepared by annealing Pos\_30mer with equal molar of Neg\_30mer in a thermocycler. The oligo mixtures were heated to 95 °C for 10 min and slowly cooled down to 25 °C in 35 min. The quality of the duplex DNA was evaluated in pre-cast 8% TBE gel (ThermoFisher Scientific).

The 260 pb promoter region of the *prp* operon were amplified as described in the literature(135) using primers p1129c\_Fw and p1129c\_Rv.

The interaction between MtPrpR and the suggested recognition DNA was tested using eletrophoretic mobility shift assay (EMSA). MtPrpR\_V9 construct was used in the assay. The purified protein was prepared in serial dilutions as indicated in figures. Protein-DNA mix were loaded to pre-cast 8% TBE gel and resolved at 4 °C. The binding buffer, loading dye and other reagents were used following the supplier's manual (ThermoFisher Scientific). Small molecule ligands were added when indicated.

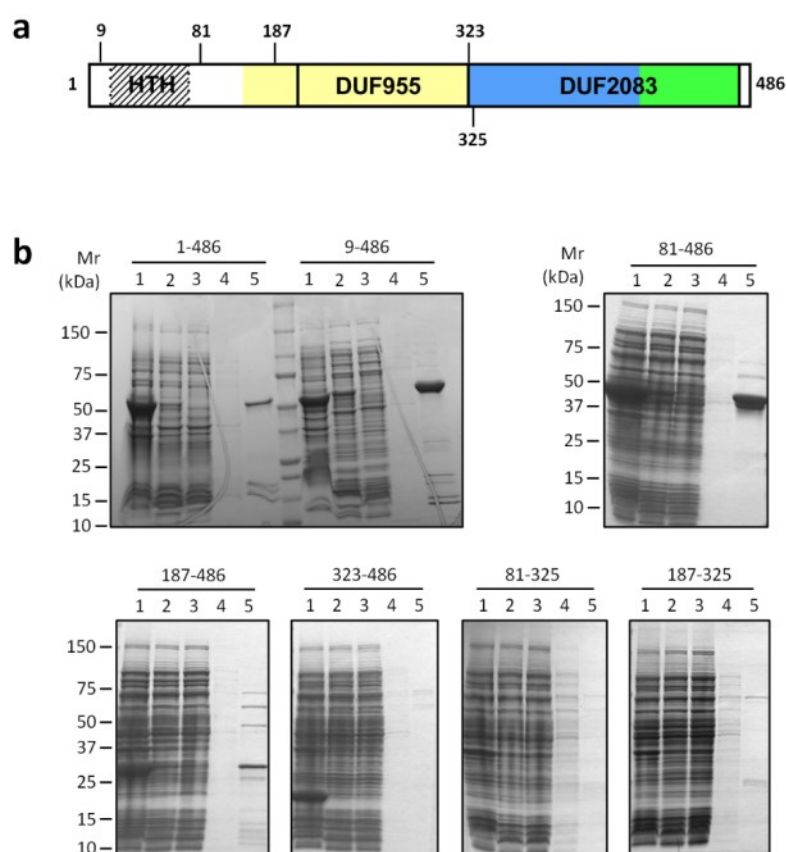
## II.3 Results

### II.3.1 Identification of an iron-sulfur cluster in MtPrpR

To understand the transcription activation of MCC at the molecular level, we employed X-ray crystallography to study the structure and function of its transcriptional regulator MtPrpR. We generated a number of different recombinant protein constructs of



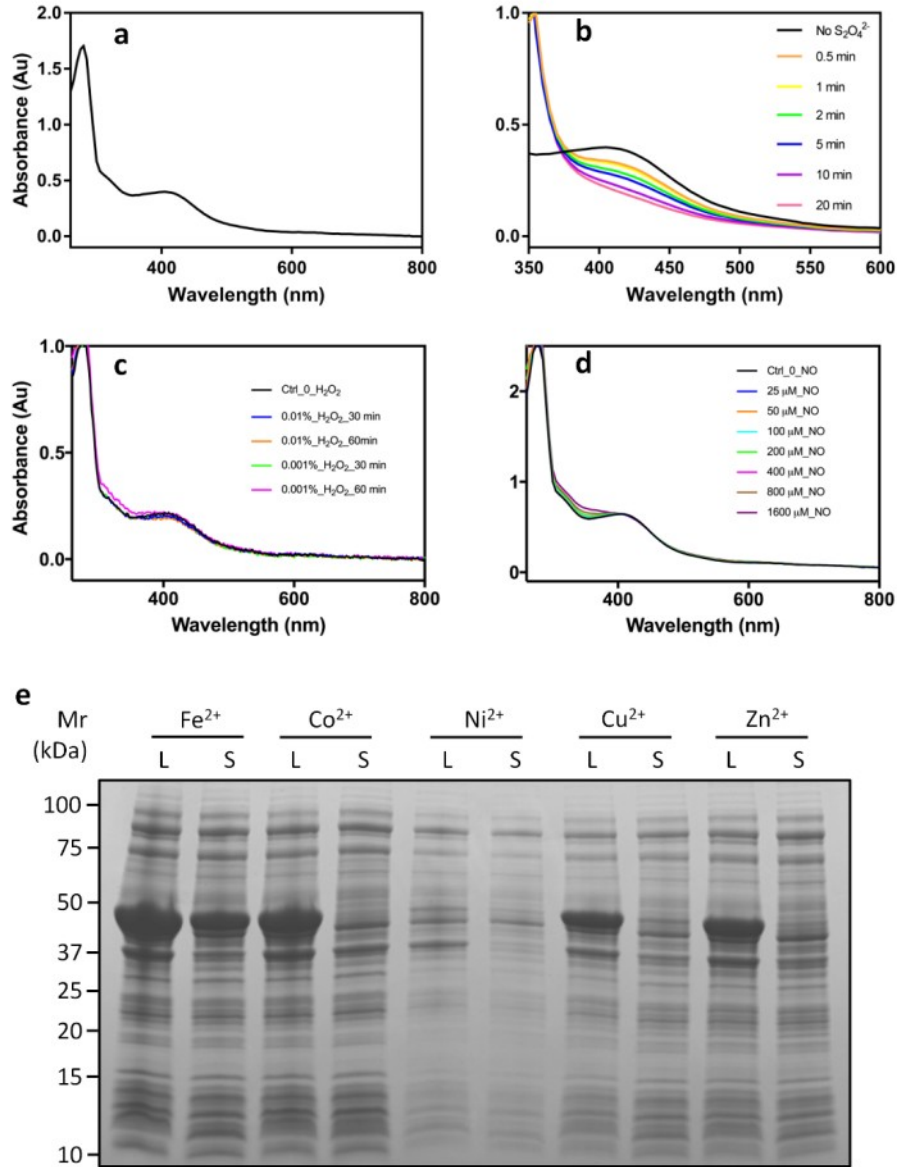
MtPrpR that were subsequently overexpressed in *E. coli* (Fig. 16a), and most of the resultant proteins were either insoluble or easily aggregated (Fig. 16b). However, when we truncated the first 80 residues that contained the N-terminal helix-turn-helix DNA-binding domain (residues 20-76) and its flanking loops, the resultant MtPrpR<sub>81-486</sub> protein was both soluble and monodispersed even at > 20 mg/ml.



**Figure 16. Expression and purification of MtPrpR constructs**  
**a, Truncations of MtPrpR. Numbers indicate the of N- (above) and C-termini (below) of the constructs. Protein domains are illustrated as in Fig. 14. b, Affinity purification of each construct in a. (1. Lysate, 2. Supernatant, 3. Flow-through, 4. Wash, 5. Elution).**

The MtPrpR<sub>81-486</sub> had a yellowish brown color, and showed an absorbance peak at 410 nm in UV-visible spectrometry (Fig. 17a). The extinction coefficient at 410 nm ( $\epsilon_{410 \text{ nm}}$ ) was approximately  $13,700 \text{ cm}^{-1} \text{ M}^{-1}$ , indicating the potential presence of the iron-sulfur cluster. Inductively coupled plasma mass spectrometry (ICP-MS) confirmed the identity of the metal as Fe, and each polypeptide chain contains approximately 3.8 Fe (data not shown). Iron-sulfur clusters are typically oxygen labile and require stringent anaerobic conditions for the protein purification or cluster reconstitution. However, aerobically and anaerobically purified MtPrpR<sub>81-486</sub> showed no apparent difference in solubility or UV-visible spectra. Moreover, the iron-sulfur cluster could tolerate 0.01 % (approximately 3 mM) of H<sub>2</sub>O<sub>2</sub> (Fig. 17c) or 16-fold molar excess of nitric oxide (Fig. 17d) for hours without showing a decrease of the absorbance peak at 410 nm. The cluster could only be reduced by > 30-fold molar excess of dithionite ion (Fig. 17b). This high level of stability of the iron-sulfur cluster suggests that MtPrpR was unlikely to be an oxidation stress sensor.

We then assessed the importance of iron to MtPrpR stability by expressing MtPrpR<sub>81-486</sub> in M9-dextrose media supplemented with common transition metal (Fe, Co, Ni, Cu and Zn) ions and we could only observe soluble MtPrpR when iron was added to the media (Fig. 17e).



**Figure 17. Characterization of the iron-sulfur cluster of MtPrpR<sub>81-486</sub>**  
**a**, UV-visible spectrophotometry of MtPrpR<sub>81-486</sub>. A peak at ~410 nm was observed.  
**b**, Sodium dithionite reduction of the iron-sulfur cluster in MtPrpR<sub>81-486</sub>. The absorbance at 410 nm was bleached by 30 fold molar excess of sodium dithionite over the time indicated. No bleaching effect was observed for twofold to tenfold molar excess of sodium dithionite. **c**, H<sub>2</sub>O<sub>2</sub> at ≤ 0.01% did not decrease the the 410 nm absorption peak of MtPrpR. **d**, Nitric oxide up to 32 molar excess did not decrease the 410 nm absorption peak of MtPrpR. **e**, Several common transition metals were tested to substitute iron in MtPrpR<sub>81-486</sub> expression. Only iron could yield soluble protein (L, lysate; S, supernatant).

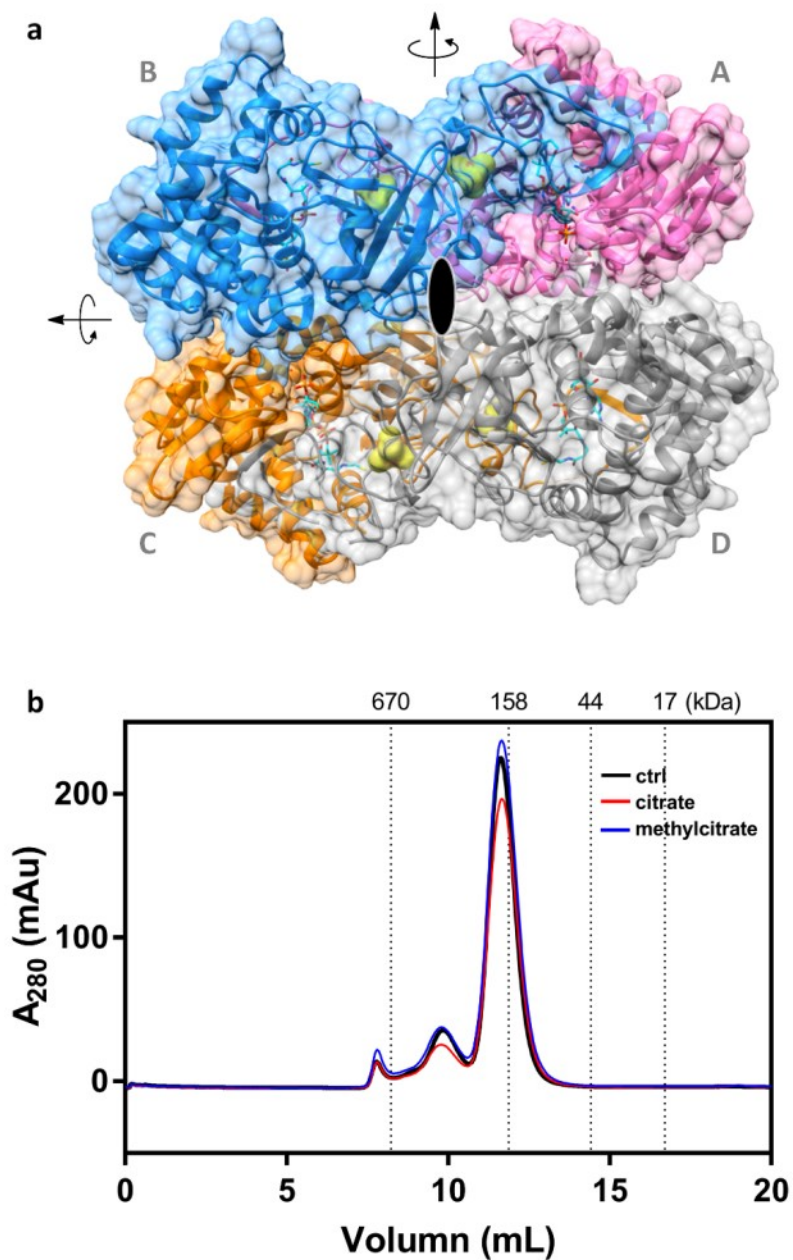
### II.3.2 *Crystal structure of MtPrpR<sub>81-486</sub>*

To visualize the iron-sulfur cluster binding mode in MtPrpR, we conducted crystallization trials on several of the above mentioned protein constructs, and only the stable MtPrpR<sub>81-486</sub> produced protein crystals. An X-ray diffraction dataset to 2.70 Å resolution was collected and phased using the selenium-single-wavelength anomalous diffraction (Se-SAD) methods (171, 172) with the selenomethionine-derived (Se-Met) protein crystals (Table 3). Interestingly, the Se-Met protein crystallized in the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> space group with two axes of very similar lengths (a = 142.37 Å, b = 145.96 Å). However, the native protein crystallized in a similarly packed unit cell but in the higher symmetry space group P4<sub>1</sub>2<sub>1</sub>2 with a = b = 144.87 Å. The asymmetric unit (ASU) of the P4<sub>1</sub>2<sub>1</sub>2 crystal contained two protein molecules and was equivalent to half of the ASU of the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal. Nevertheless, the overall crystal unit cell volumes and the structures of both forms were nearly identical. The biological assembly of MtPrpR<sub>81-486</sub> in the crystals was a homotetramer suggested by PISA(173) (Fig. 18a), which was consistent with our size exclusion chromatography (SEC) results (Fig. 18b). The four polypeptide chains organized in D<sub>2</sub> symmetry (Fig. 18a). The shared surface area of the interface on each chain were around 2,400 Å<sup>2</sup> between Chain A and B, around 1,250 Å<sup>2</sup> between Chain A and C, and around 340 Å<sup>2</sup> between Chain A and D (Fig. 18b). Since the four chains in the Se-Met structure were virtually identical (Cα\_RMSD < 0.3 Å), we focus on Chain A for subsequent depictions.

**Table 3. Crystallographic data collection and refinement statistics**

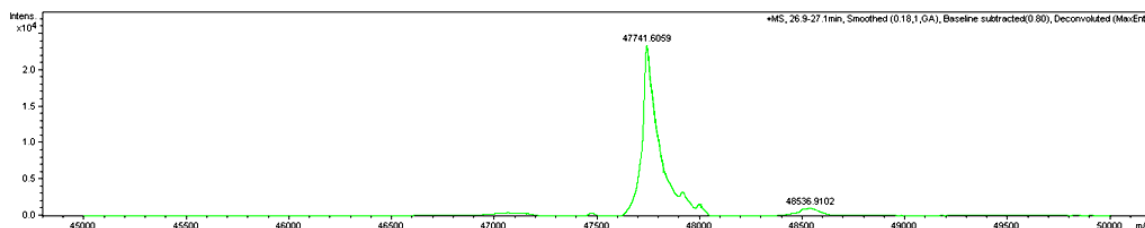
	SeMet PrpR <sub>81-486</sub>	Native PrpR <sub>81-486</sub>	PrpR <sub>81-486</sub> _F155H	PrpR <sub>155-440</sub>
PDB code	6CZ6	6CYY	6CYJ	6D2S
Data Collection				
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	I222
Unit cell demensions				
a, b, c (Å)	96.19, 142.37, 145.96	144.87, 144.87, 97.27	144.49, 144.49, 96.19	75.00, 81.72, 95.36
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Wavelength (Å)	0.97949	1.07507	1.03328	0.97935
Resolution range (Å)	50.00-2.70 (2.75-2.70)	50.00-2.50 (2.54-2.50)	50.00-2.70 (2.75-2.70)	50.00-1.82 (1.85-1.82)
Unique reflections *	55,505 (2,747)	35,534 (1,773)	27,986 (1,414)	26,651 (1,274)
Multiplicity *	7.4 (7.4)	21.2 (22.6)	17.0 (17.1)	9.5 (7.3)
Completeness (%) *	100.0 (100.0)	98.4 (100.0)	97.8 (100.0)	99.8 (100)
R <sub>merge</sub> *	0.108 (1.132)	0.059 (0.733)	0.078 (1.052)	0.043 (0.241)
R <sub>pim</sub> *	0.043 (0.446)	0.013 (0.157)	0.019 (0.258)	0.015 (0.095)
I/σ (I) *	21.80 (1.38)	66.7 (3.77)	45.60 (2.44)	61.53 (10.11)
Wilson B value (Å <sup>2</sup> )	65.88	70.26	78.74	20.30
Refinement				
Resolution (Å)	45.56-2.70 (2.80- 2.70)	46.10-2.51 (2.60- 2.51)	45.63-2.70 (2.80- 2.70)	32.09-1.82 (1.88- 1.82)
R <sub>work</sub> /R <sub>free</sub> *	0.182/0.212 (0.302/0.345)	0.183/0.214 (0.259/0.307)	0.185/0.218 (0.281/0.364)	0.179/0.214 (0.205/0.269)
No. of atoms	10,515	5,288	5,286	2,484
Protein	10,267	5,172	5,158	2,287
Ligand	224	112	126	5
Solvent	24	4	2	192
Mean B factors (Å <sup>2</sup> )	67.29	77.75	85.96	27.72
Protein	67.28	77.85	86.05	26.87
Ligand	69.43	73.98	82.36	24.43
Solvent	52.00	62.62	69.59	37.83
Rmsd bond length (Å)	0.005	0.004	0.003	0.007
Rmsd bond angle (°)	0.763	0.823	0.721	0.834
Clashscore	5.90	3.55	6.15	5.67
Number of TLS groups	28	15	15	2
Ramachandran plot (%) (favored/allowed/outliers)	97.69/2.31/0.00	97.87/2.13/0.00	98.47/1.53/0.00	97.79/2.21/0.00
Rotamer outliers (%)	0.94	0.00	0.56	0.00

\* Values in parentheses indicate highest-resolution bin.



**Figure 18. Structure of MtPrpR<sub>81-486</sub> and analytical SEC of MtPrpR<sub>81-486</sub> in solution**  
**a, Crystal structure of MtPrpR<sub>81-486</sub> tetramer. The four chains are labeled and shown in different colors. Arrows and ellipse indicate the positions and directions of the two-fold rotation axes of the tetramer. b, Size exclusion chromatography analysis of MtPrpR<sub>81-486</sub> with or without TCA and MCC intermediates citrate and methylcitrate.**

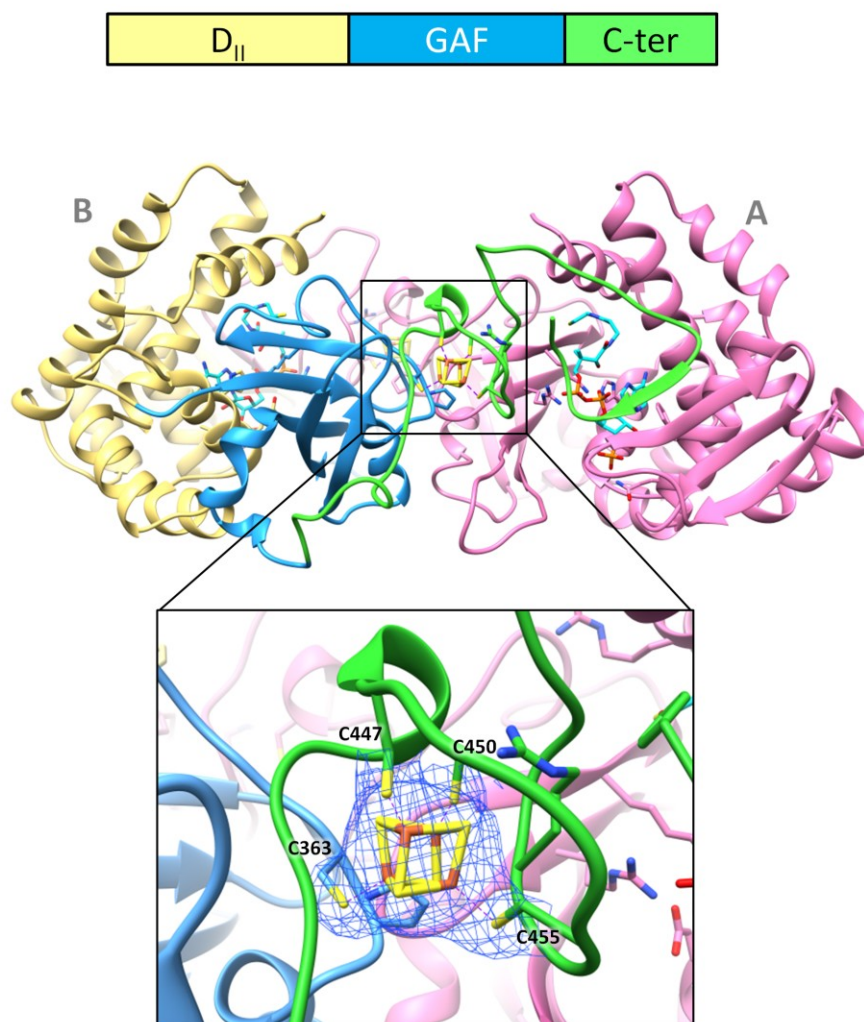
Based on the visible part of the truncated protein structure, MtPrpR contains at least three domains (Fig. 20a), including the N-terminal DNA-binding domain (residues 20-76) that was removed in our protein construct. A disordered region (residues 81-152) connected the DBD to the second domain (designated D<sub>II</sub>, residues 153-324). The electron density for the disordered region was not visible in our crystal structure, but based on the length of this region, it is likely to form a domain. Protein mass spectrometry confirmed that this region was present in the crystallized protein (Fig. 19). The D<sub>II</sub> domain is composed of a nine-helix bundle and a three-stranded  $\beta$ -sheet (Fig. 20). A loop (residues 325-334) links the D<sub>II</sub> to a GAF-like (cGMP phosphodiesterase/adenylyl cyclase/FhlA) (174) domain (residues 335-419).



**Figure 19. Mass spectrometry of the His-tagged MtPrpR<sub>81-486</sub> crystals**  
Protein intact MS showed a de-convoluted peak with a  $[M+H]^+$   $m/z$  of 47,743, indicating that no proteolysis had occurred.

MtPrpR also contains a long C-terminal region (residues 420-486) consisting primarily of loops plus several fairly short secondary structure elements. The C-terminal region of Chain A interacts with the D<sub>II</sub>-GAF di-domain of Chain B (Fig. 18a, Fig. 20). The interaction is stabilized by an extended  $\beta$ -sheet formation between a short  $\beta$ -strand

at the end of the C-terminal region and the three  $\beta$ -strands in the D<sub>II</sub> domain of the neighboring chain (Fig. 20).

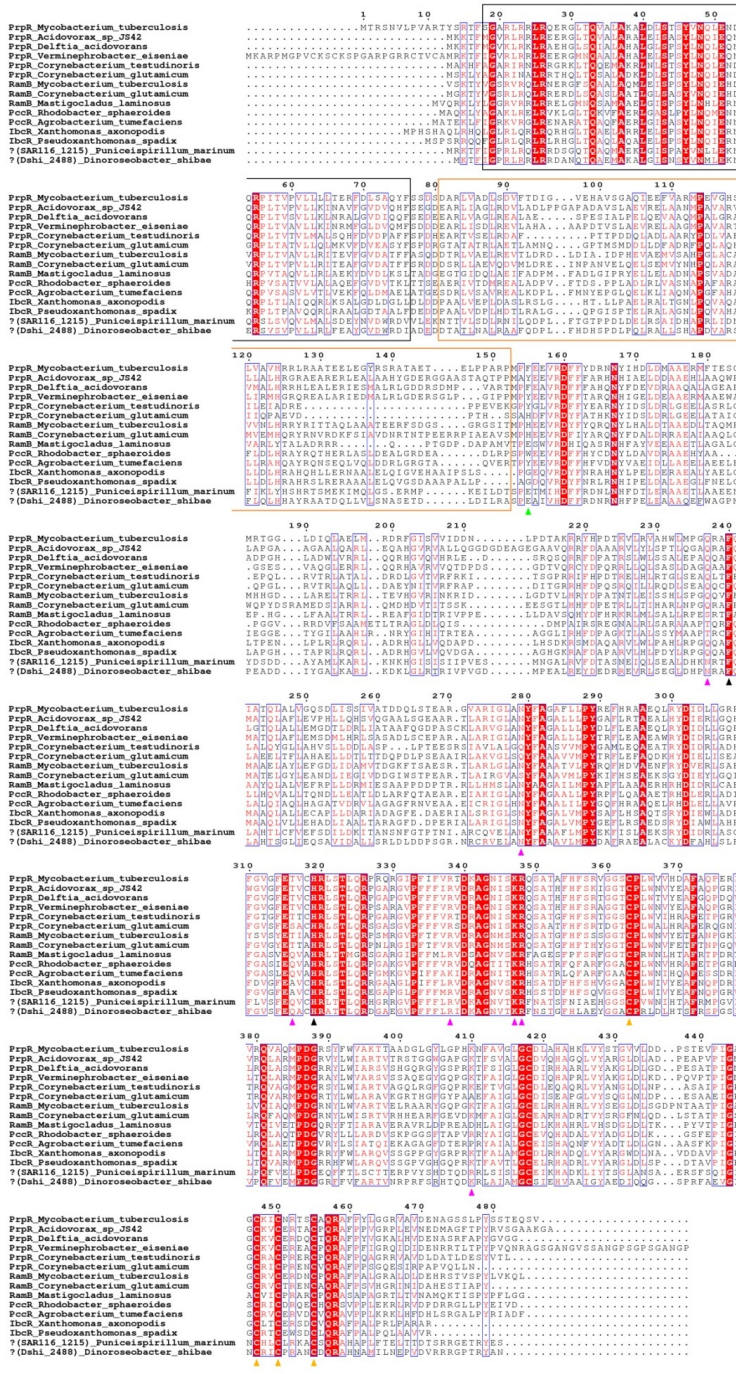


**Figure 20. Iron-sulfur cluster binding mode of MtPrpR<sub>81-486</sub>**  
Chain A (color in pink) and Chain B (color as in the domain organization) are extracted from the tetramer. The iron-sulfur cluster binding site is zoomed in. 2Fo-Fc electron density of iron-sulfur cluster is shown at a contour level of 1  $\sigma$ .

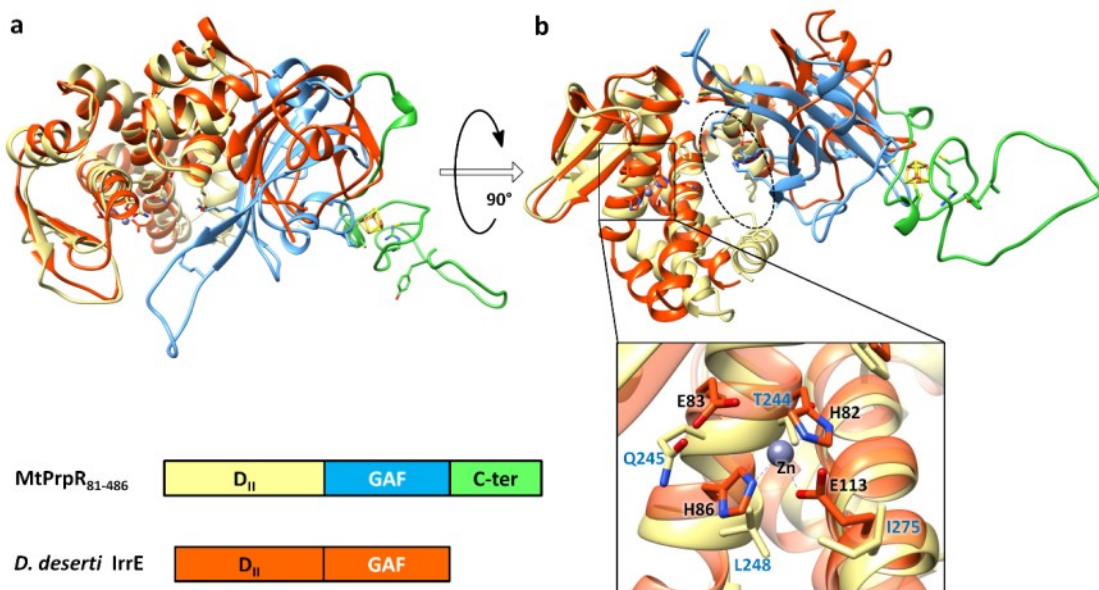


The cubane-type [4Fe4S] cluster resides in the C-terminal region (Fig. 20) with three of the four irons ligated by Cys447, Cys450 and Cys455 in a C-X<sub>2</sub>-C-X<sub>4</sub>-C motif which appears to be conserved among all MtPrpR homologs (Fig. 21). Cys363 from the GAF-like domain of the same chain coordinates the fourth iron (Fig. 20). The loop harboring the C-X<sub>2</sub>-C-X<sub>4</sub>-C motif wraps around the iron-sulfur cluster and buries the majority of it inside the protein, which we postulate as the major contribution to the stability of the cluster when the protein was exposed to air.

A 3D structure similarity search of the MtPrpR<sub>81-486</sub> structure against the Protein Data Bank (PDB) (165) suggested that apart from the hits on the well-studied GAF domain, the only structure that showed significant similarity to MtPrpR<sub>81-486</sub> was a *Deinococcus deserti* DNA damage repair regulator IrrE (146) (PDB code: 3DTI and related entries, Z-score 15.5, rmsd 5.0 Å, Fig. 22a, b). There were only around 240 residues within the D<sub>II</sub>-GAF region of MtPrpR<sub>81-486</sub> that could be structurally aligned to IrrE, and the sequence identity of the region was only 14%. IrrE was proposed as a metalloprotease with an HEXXH motif coordinating a Zn ion (146, 147) (Fig. 22b). However, those residues are not conserved in MtPrpR.



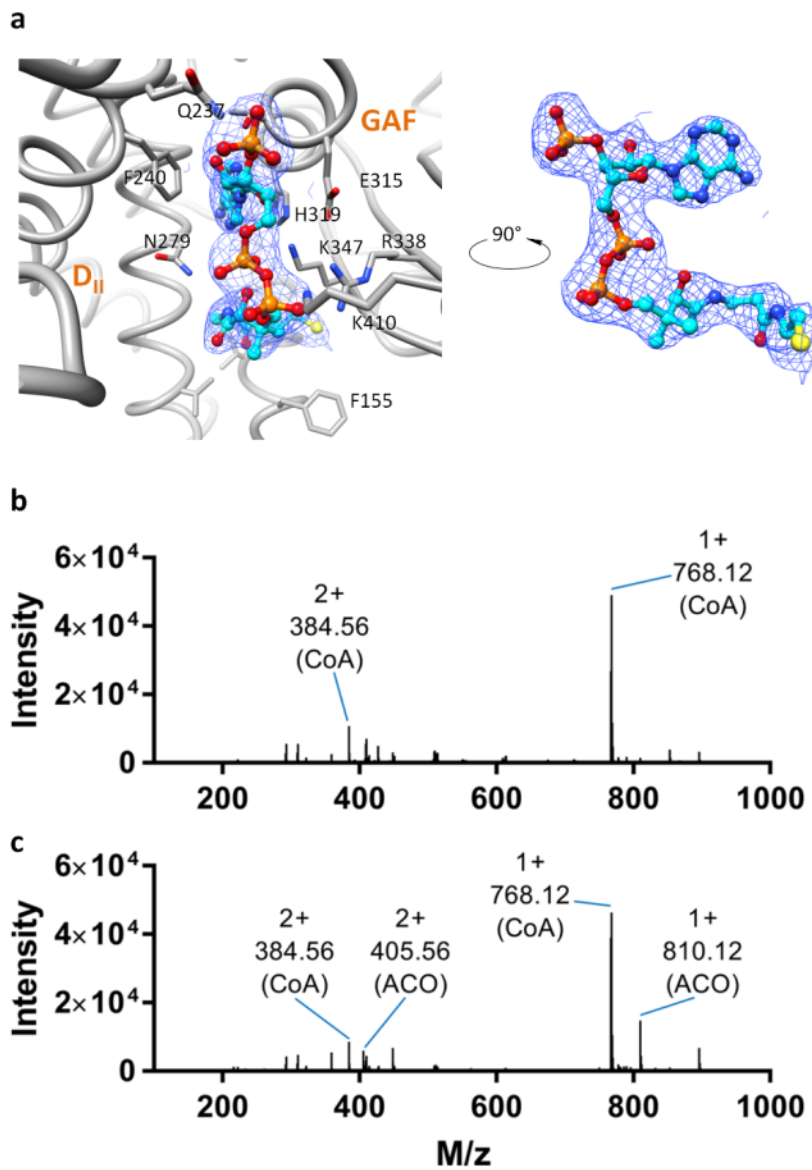
**Figure 21. Sequence alignment of MtPrpR and homologs**  
**The CoA adenine group interacting residues are pointed with black triangles. The CoA hydrogen bonding residues are pointed with purple triangles. The CoA tail contacting residue is pointed with green triangle. The cysteinyl residues ligating the iron-sulfur cluster are pointed with yellow triangles.**



**Figure 22. Structure comparison between MtPrpR<sub>81-486</sub> and *D. deserti* IrrE**  
**a, Superposition of MtPrpR<sub>81-486</sub> (colored by the domains as in Fig. 20) and IrrE from *D. deserti* (orange, PDB accession number: 3DTI). The RMSD of 160 aligned Ca of D<sub>II</sub> domain is 2.33 Å, and 7.79 Å over all 229 Ca pairs between the two proteins. b, Rotating (a) by 90 degree to show the ligand binding cavity in MtPrpR<sub>81-486</sub> (dashed ellipse), and the zinc binding site of the putative metalloprotease domain in IrrE (square). A close-in view of the zinc-binding site shows that the residues are not conserved in MtPrpR. The IrrE residues are labeled in black; the MtPrpR residues are labeled in blue.**

### II.3.3 Identification of Coenzyme A in MtPrpR

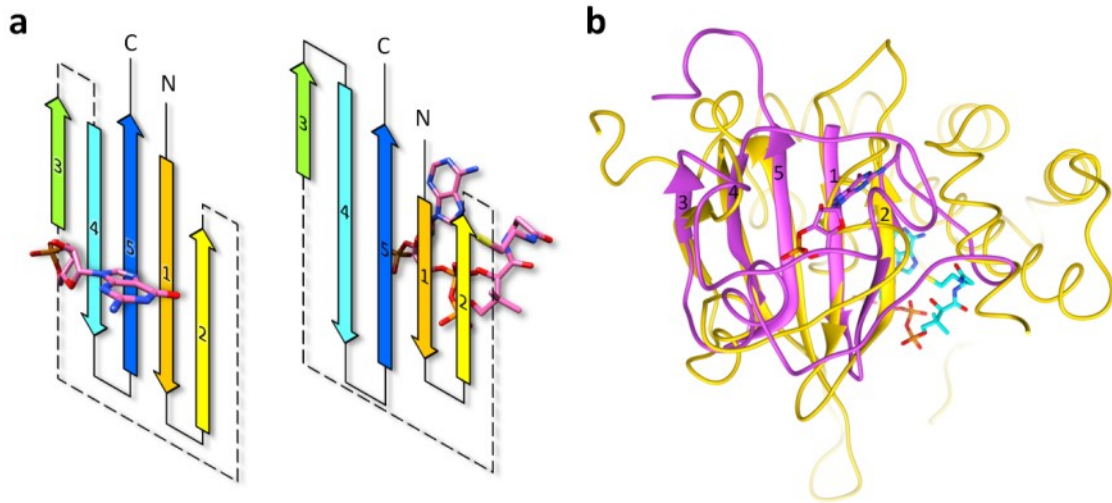
After building and refining the protein atoms and the iron-sulfur clusters into the electron density map, we observed a large electron density blob in each polypeptide chain that could not be accounted for by protein. These electron density peaks were nearly identical in shape, and they were located in the cleft between the D<sub>II</sub> and GAF domains in both the Fo-Fc and 2Fo-Fc difference electron density maps (Fig. 23a), indicating a ligand molecule had been cocrystallized with the protein. To identify the ligand, MtPrpR<sub>81-486</sub> was denatured, and the ligand was enriched by ion exchange resin and subjected to liquid chromatography-high resolution mass spectrometry (LC-MS). The [M+H]<sup>+</sup> m/z value showed that the majority of the ligand was Coenzyme A (768.12) from protein expressed in LB media (Fig. 23b). When the protein was expressed in M9-dextrose-iron media, acetyl-CoA ([M+H]<sup>+</sup> m/z = 810.12) was also detected besides CoA (Fig. 23c). The ratio of acetyl-CoA to CoA was about 1:8 according to the HPLC UV chromatogram (data not shown). However, we did not observe well defined electron density of the acetyl group, although there was room for it in the binding cavity. Therefore, the predominant CoA was built into all protein chains, and the CoA molecules fit very nicely into the electron density map with occupancies of 1.0 and reasonable refined B-factors.



**Figure 23. Identification of CoA in MtPrpR<sub>81-486</sub>**

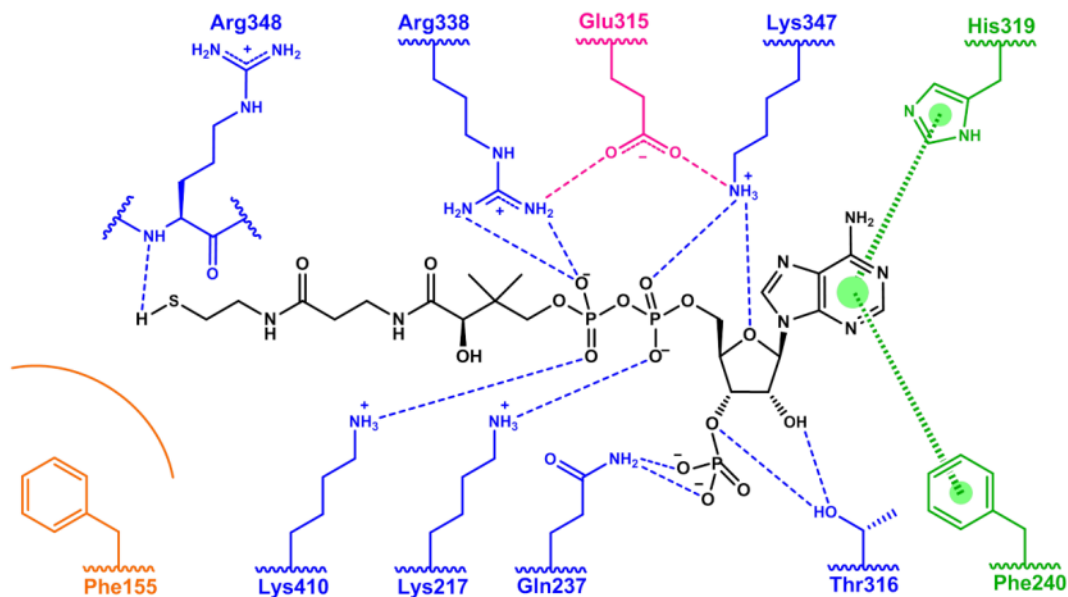
**a**, Coenzyme A binding cavity in MtPrpR<sub>81-486</sub>; protein chain is colored in grey. The binding cavity is cooperatively formed by the D<sub>II</sub> (left, mainly helix bundles) and GAF (right, mainly  $\beta$  sheets) domains. Key interacting residues are labeled. CoA is shown in stick and ball; 2Fo-Fc map of CoA is contoured at 1  $\sigma$ . Electron density of CoA is also displayed by a 90 degree rotation. **b** and **c**, Mass spectrum of CoA and its derivatives extracted from purified MtPrpR<sub>81-486</sub> expressed in LB (**b**) or in M9-dextrose (**c**) media.

Although the GAF-like domain in MtPrpR participated in the CoA binding, the binding mode was quite different from those found in typical GAF domain structures. The canonical GAF domain and the closely related PAS (Per-Arnt-Sim) domain have been reported to bind to a variety of signaling molecules (174-178) on the concave face of the central  $\beta$ -sheet of the GAF domain (Fig. 24). However, the CoA binding cavity in MtPrpR was cooperatively formed between the convex face of the central  $\beta$ -sheet of the GAF-like domain and the helix bundle of the D<sub>II</sub> domain (Fig. 23a, 24a, b). The CoA molecule embedded snugly in the cavity. The adenine group was sandwiched between Phe240 and His319, and was held in place by  $\pi$ - $\pi$  interactions. The rest of CoA bound via 13 hydrogen bonds and electrostatic interactions contributed by the side chains or backbone atoms of seven amino acids in the D<sub>II</sub> and GAF domains (Fig. 25). The diphosphate group in particular was surrounded by four basic residues, Lys217, 347 and 410 and Arg338. Phe155 located at the bottom of CoA binding cavity and contacted the cysteamine thiol through van der Waals interactions. The C-terminal loop region of the neighboring Chain B forms a hinged triangular lid that covers the CoA binding cavity of Chain A (Fig. 26c).



**Figure 24. Structure and ligand binding mode comparisons between canonical GAF and MtPrpR GAF-like domains**

**a, Topology diagrams of canonical GAF domain (left) and the GAF-like domain in MtPrpR (179) indicate the ligand binding modes are different. b, Comparison of the canonical cGMP bound GAF domain structure (magenta, PDB code: 1MC0) with the GAF-like domain in MtPrpR structure (protein colored in yellow and ligand colored in cyan).**

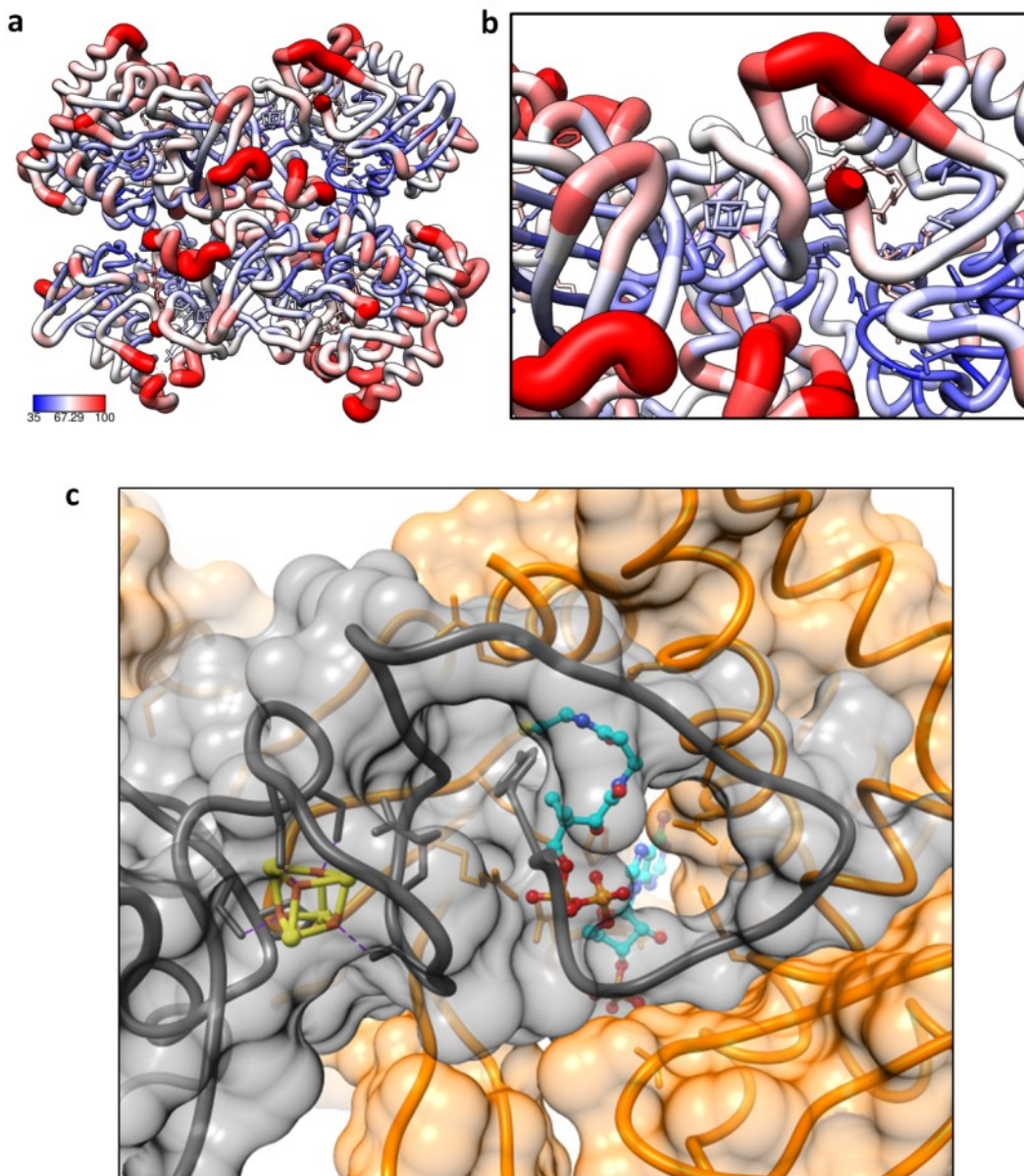


**Figure 25. Interactions between MtPrpR<sub>81-486</sub> and CoA**

The residues contributing to  $\pi$ - $\pi$  interactions are colored in green. The residues and hydrogen bonds between protein and CoA are colored in blue, and Glu315 which forms intramolecular hydrogen bonds to stabilize Arg348 and Lys347 is colored in magenta. The CoA tail contacting residue Phe155 is colored in orange.

The hinge region harbors the C-X<sub>2</sub>-C-X<sub>4</sub>-C motif where the iron-sulfur cluster is ligated. The bound of the iron-sulfur cluster appears to rigidify the hinge region, as indicated by the lower local b-factors (Fig. 26a, b). This makes it difficult for the lid to open, and therefore CoA appears to be locked inside the cavity (Fig. 26c). Common ligand stripping procedures such as dialysis failed to dissociate CoA from the protein. From the structural aspect, the integrity of the iron-sulfur cluster appears to be pivotal for protein oligomerization and the retention of CoA in the cavity.





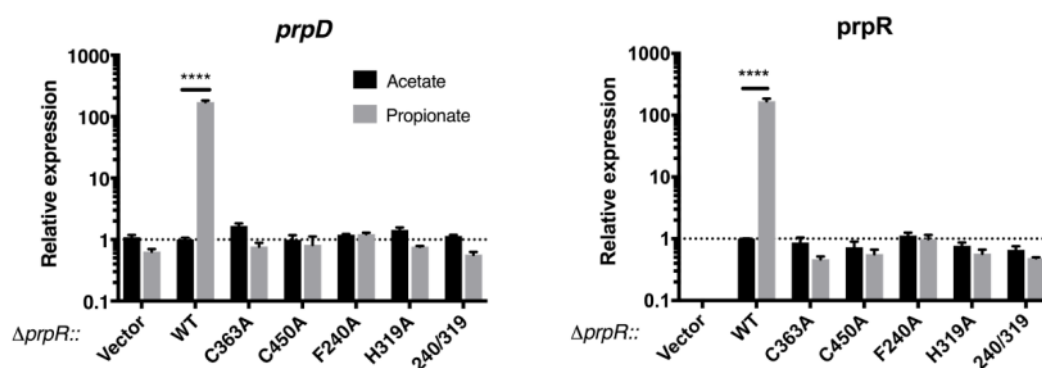
**Figure 26. Relationship between iron-sulfur cluster and CoA binding**  
**a, MtPrpR<sub>81-486</sub> rendered by b-factor (Å<sup>2</sup>). b, A close-in view of a around the iron-sulfur cluster. c, The hinged triangle lid (dark grey) that covers the CoA binding pocket of the neighboring chain (orange). The iron-sulfur cluster is located at the hinge region and stabilizes the hinge.**

### II.3.4 The iron-sulfur cluster is critical for CoA binding and transcription activation

To determine if the iron-sulfur cluster in the structure was important for transcriptional regulation, we picked two iron-coordinating cysteine residues, Cys363 in the GAF-like domain and Cys450 in the C-terminal C-X<sub>2</sub>-C-X<sub>4</sub>-C motif, and mutated them individually into alanine (C363A and C450A) in a *prpR*-containing shuttle vector (140). We chromosomally integrated the wildtype *prpR*, each mutant, or an empty vector into the *prpR* deleted *M. tuberculosis* laboratory strain H37Rv (*H37RvΔprpR*) (104). The upregulation of *prp* operon, represented by *prpD*, or *prpR* in response to propionate were completely abolished in the *ΔprpR::C363A*, *ΔprpR::C450A*, and *ΔprpR::vector* strains (Fig. 27).

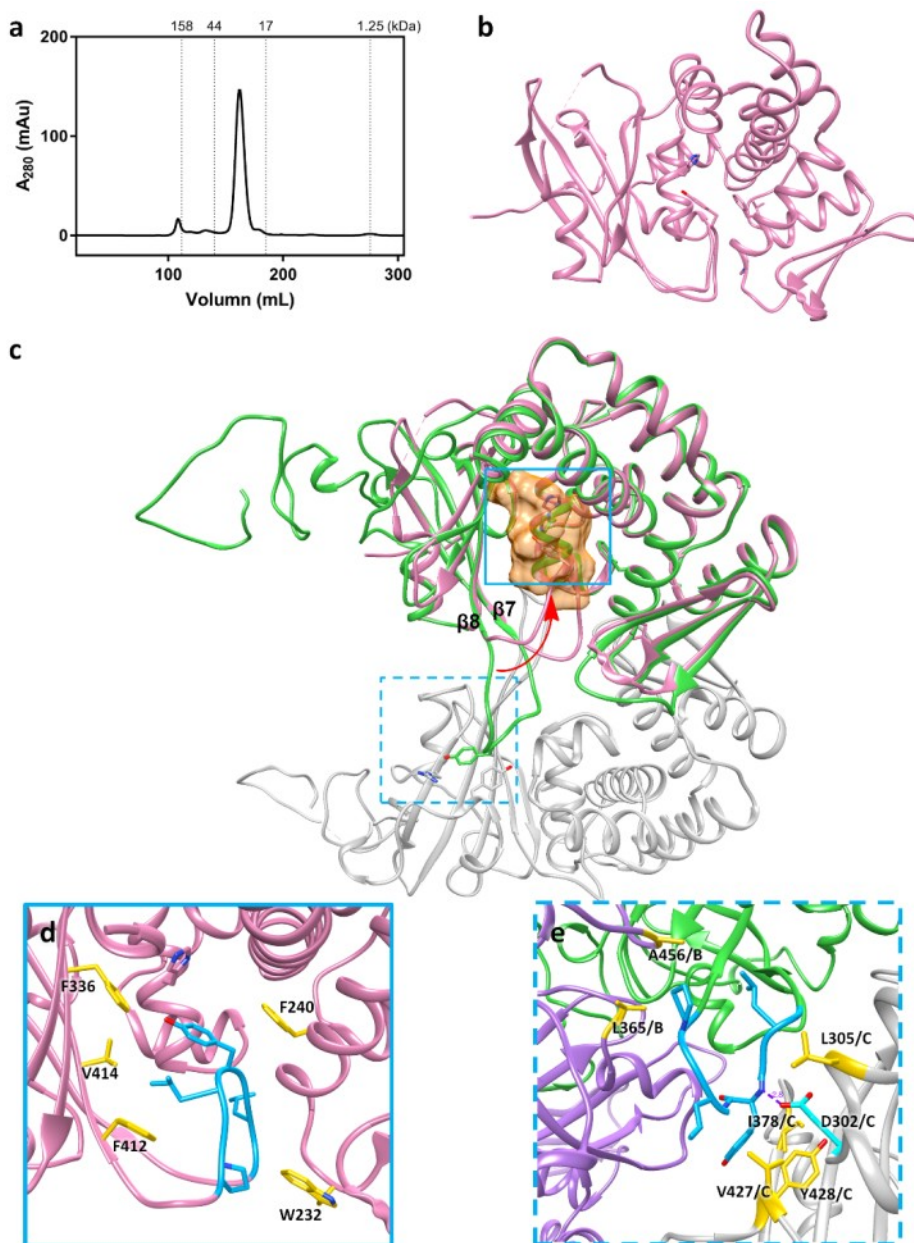
We next sought to visualize the iron-sulfur free form of the protein. Since overexpressing MtPrpR<sub>81-486</sub> in the iron-free M9-dextrose media failed to produce soluble protein, we truncated a large portion of the C-terminal region (residues 441-486) containing the C-X<sub>2</sub>-C-X<sub>4</sub>-C iron-sulfur cluster-binding motif. In addition, residues 81-154, which were disordered and not visible in the MtPrpR<sub>81-486</sub> structure, were also removed from the expression construct. The resultant *prpR*<sub>155-440</sub> construct produced soluble protein with no color, and the MtPrpR<sub>155-440</sub> protein was a monomer in solution as shown by the size exclusion chromatography (Fig. 28a). The crystal structure of MtPrpR<sub>155-440</sub> showed that the protein retained the D<sub>II</sub>-GAF core structure of MtPrpR<sub>81-486</sub> and, as expected, had no bound iron-sulfur cluster (Fig. 28b). Interestingly, the CoA was also absent in the structure, even though 1 mM CoA was mixed with the protein during crystallization. The structure of the CoA binding cavity revealed a dramatic

conformational change of the protein (Fig. 28c). The loop between  $\beta 7$  and  $\beta 8$  of the GAF-like domain flipped back nearly 180 degrees, thereby blocking the CoA cavity entrance. A hydrophobic fragment of the loop with the amino acid sequence of  $G_{402}L_{403}G_{404}Y_{405}L_{406}G_{407}P_{408}$  inserts into the CoA binding cavity and interacts with Trp232, Phe240, 336, 412, and Val414, all within van der Waals distances to the loop (Fig. 28d). Gly404, Tyr405 and Leu406 at the tip of the loop occupy the space where the 3'-phosphoadenosine diphosphate moiety of CoA resided in the CoA bound form of the protein (Fig. 28c). While in the tetrameric MtPrpR<sub>81-486</sub> structure, the aromatic Tyr405 and the aliphatic Leu403, 406 and Pro408 in the loop in Chain A were surrounded by the hydrophobic residues Leu365 and Ala456 in Chain B, and Leu305, Ile378, Val427 and Tyr428 in Chain C, all within 5 Å distances (Fig. 28e). In addition, Asp302 in Chain C hydrogen bonded (2.8 Å) to the backbone nitrogen of Tyr405 (Fig. 28e). These interactions would essentially hold the hydrophobic loop in an open conformation



**Figure 27. Transcription levels of *prpD* and *prpR* in MtPrpR variants under acetate or propionate conditions**

Values presented as mean  $\pm$  SD ( $N=3$  for wildtype and each variant, Data were analyzed with two-way ANOVA followed by Sidak's test, \*\*\*\*  $p<0.0001$ ).



**Figure 28. Structural characterization of MtPrpR<sub>155-440</sub>**  
**a**, Size exclusion chromatography analysis of MtPrpR<sub>155-400</sub>. **b**, Crystal structure of MtPrpR<sub>155-400</sub>. **c**, Conformational change of MtPrpR in the presence (green and grey for the two neighboring chains) or absence (180) of the iron-sulfur cluster binding region. The surface of CoA is shown in orange. Red arrow indicates the open and closed forms of the loop. Solid-line square is zoomed in (**d**) to show the hydrophobic interactions in the loop-closed conformation. Dashed-line square is zoomed in (**e**, Chain A in green, Chain B in violet, Chain C in grey) to show the interactions in the loop-open conformation.

allowing for CoA binding. Taken together, this conformational change of the protein, likely induced by the lack of the iron sulfur cluster, prevented CoA from binding.

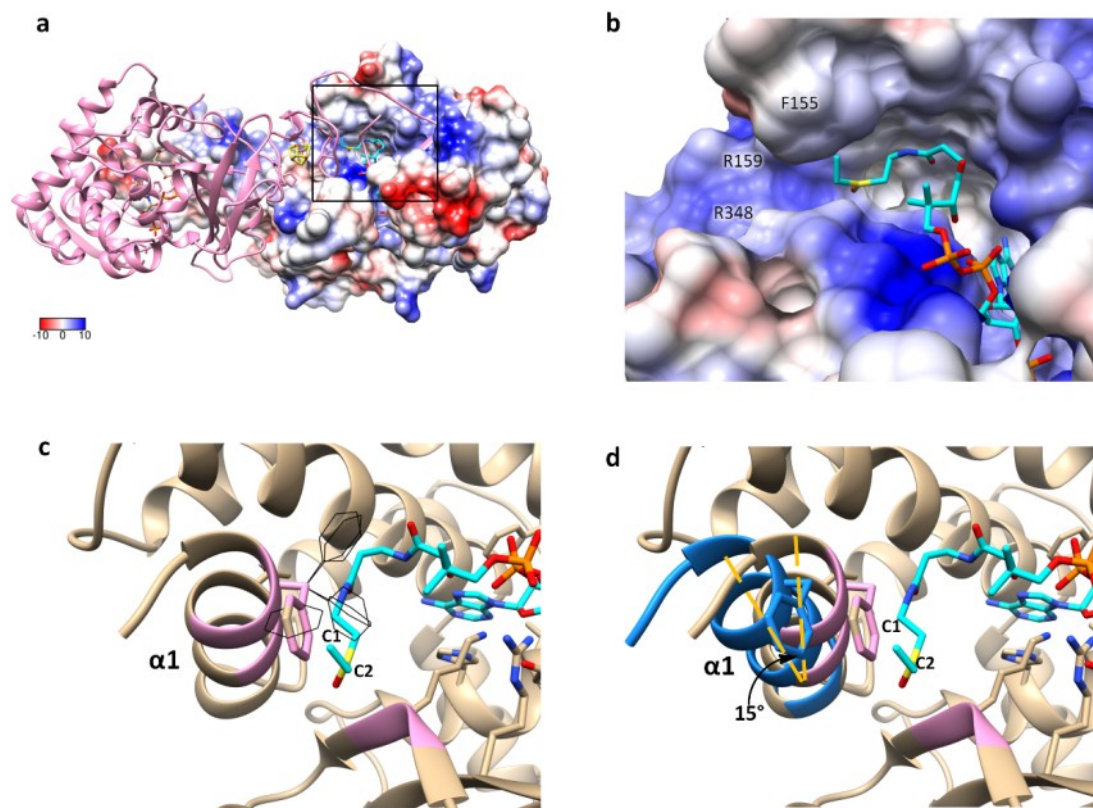
### II.3.5 CoA derivatives control the MCC pathway activation

The  $\pi$ - $\pi$  interactions between CoA adenine group and the residues Phe240 and His319 may play critical roles during CoA binding. These two residues appear to be invariable in most MtPrpR homologs (Fig. 21). To test the importance of these interactions, we complemented the H37Rv $\Delta$ *prpR* strain with each of the *prpR\_F240A*, *prpR\_H319A* single mutations or *prpR\_F240A/H319A* double mutation. None of the resultant strains could upregulate *prpD* or *prpR* in response to propionate (Fig. 27).

There is good agreement between our studies (Fig. 27, 30, 32) and the literature (104, 135) on *prpR* that it activates the transcription of the *prp* operon only under propionate or cholesterol carbon utilization, where propionyl-CoA is the common metabolite. Therefore, we speculated that the CoA bound MtPrpR was not the active form, but that propionyl-CoA was the switch molecule used to activate the *prp* operon transcription. We tried to crystallize propionyl-CoA bound MtPrpR<sub>81-486</sub> by mixing the protein with different concentrations of propionyl-CoA, or by purifying MtPrpR<sub>81-486</sub> expressed in M9-propionate media. Nonetheless, the desired propionyl group was not discernible in the electron density map.

We therefore virtually modeled propionyl-CoA into the MtPrpR<sub>81-486</sub> structure by treating the protein as a rigid body (Fig. 29a, b). This led to steric clashes between the propionyl group and Phe155 (Fig. 29c). These clashes cannot be overcome by simply

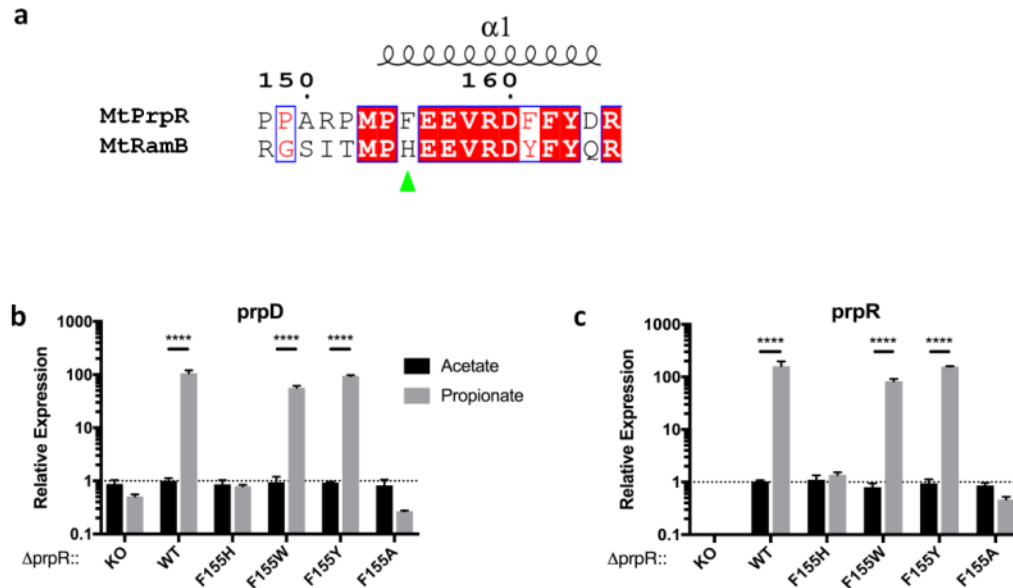
changing the rotamer conformation of Phe155 (Fig. 29d). Instead, a roughly fifteen-degree movement of the  $\alpha$ -helix ( $\alpha$ 1, the first visible helix) where Phe155 locates is required for the protein to accommodate propionyl-CoA. This also implies that Phe155 may directly participate in the ligand selection.



**Figure 29. MtPrpR<sub>81-486</sub> with propionyl-CoA computationally modeled into the CoA binding cavity**

**a**, Propionyl-CoA was modeled into the MtPrpR<sub>81-486</sub>\_WT structure. Electrostatic surface is shown (red, negative charge; blue, positive charge; unit kcal\* $\text{mol}^{-1}$ \* $e^{-1}$ ). **b**, Close-in view of a. **c**, Propionyl-CoA clashes with protein chain. Atoms within 5 Å to the C1 and C2 of the propionyl group are colored in pink. Distance between Phe155 and C1 is  $\sim 2.3$  Å. Black wires represent the possible rotamers of Phe155. **d**, A roughly 15° movement of  $\alpha$ 1 (blue, new position) is required to overcome the clashes.

Through a structure guided sequence comparison, we noticed that Phe155 in MtPrpR is replaced by a histidine in its paralog MtRamB (Phe155→His143, Fig. 30a), while other CoA contacting residues remains identical. Recall that MtRamB has been shown as a transcriptional repressor of *icl1* in dextrose containing media (148). We hypothesized that the Phe→His substitution would allow MtRamB to bind to a different CoA derivative, and that F155H mutation of MtPrpR would affect its ability to respond to propionyl-CoA. We therefore integrated a *prpR\_F155H* mutant allele into the H37RvΔ*prpR* genome. The Δ*prpR::F155H* retained no activity in response to propionate in two experimental runs (Fig.30b). Mutating Phe155 into alanine, which has an even smaller side chain, completely abrogated the upregulation of *prpD* and *prpR*. While mutating Phe155 into the bulky aromatic residues tryptophan or tyrosine restored the upregulation of *prpD* and *prpR* to the wildtype level (Fig 30b). In fact, these bulky aromatic residues naturally exist in MtPrpR homologs activating either the methylcitrate cycle (i.e. Ajs\_1637 protein of *Acidovorax sp. JS42* strain harbors a Tyr (181)) or the methylmalonyl pathway (i.e. the PccR protein of *Rhodobacter sphaeroides* harbors a Trp (150)) (Fig. 21). Notably, both pathways are used by bacteria to assimilate propionyl-CoA. Therefore, it leads to the postulate that propionyl-CoA clashes with those aromatic residues and induces a similar conformational change of the proteins.



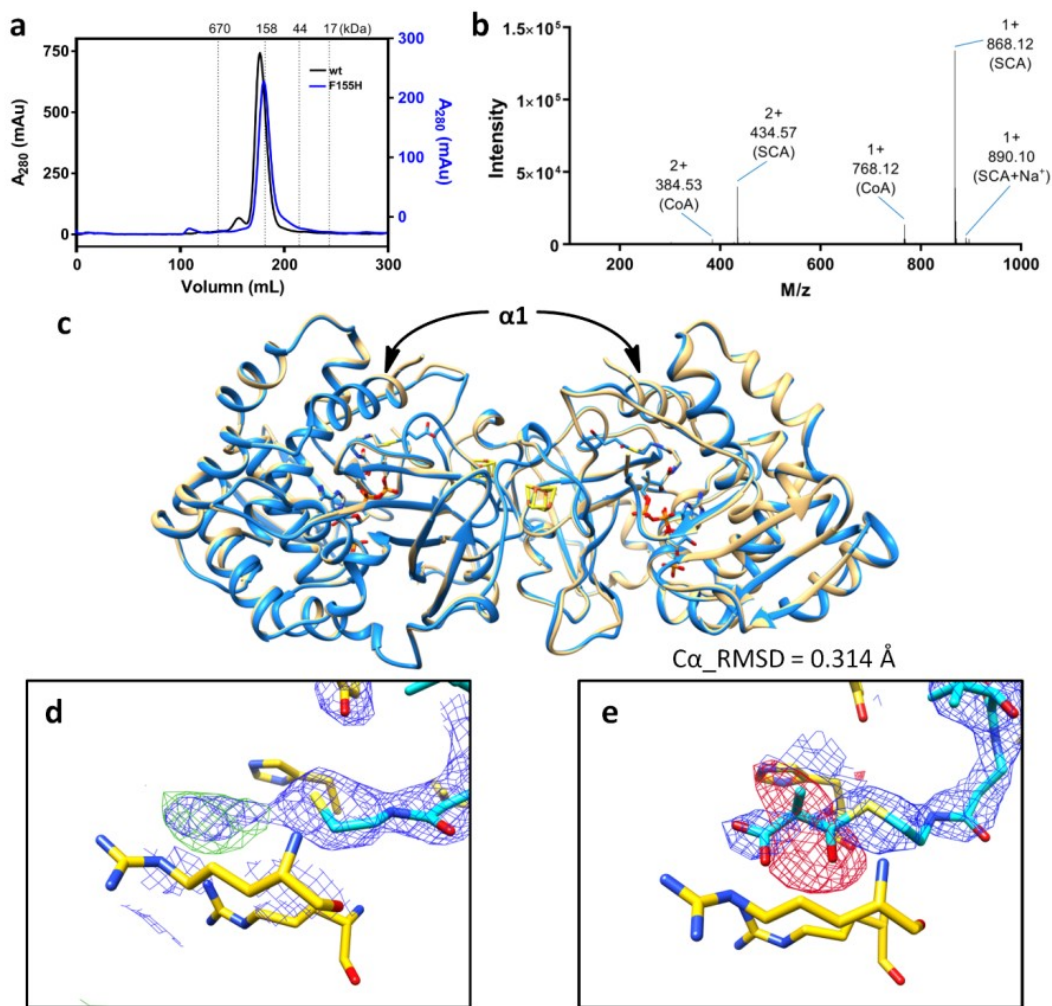
**Figure 30. Study of the role of Phe155 in CoA derivative binding**  
**a**, Sequence alignment of the helix  $\alpha 1$  between MtPrpR and MtRamB, highlighting Phe155 in MtPrpR and its counterpart His143 in MtRamB. **e**, and **f**, Transcription level of *prpD* ( $N=3$ ) and *prpR* ( $N=3$ ) in MtPrpR\_F155 variants under acetate or propionate conditions. Values presented as mean  $\pm$  SD (Data were analyzed with two-way ANOVA followed by Sidak's test, \*\*\*\*  $p<0.0001$ ).

### II.3.6 MtRamB may recognise succinyl-CoA

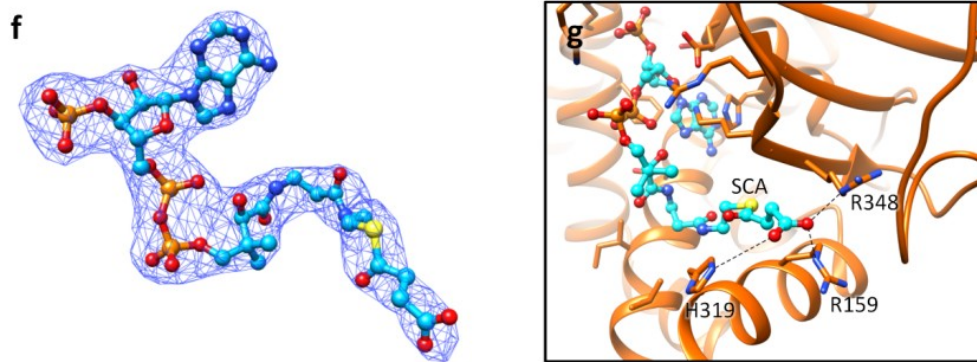
We were particularly interested in the identity of the ligand recognized by MtRamB. However, the insolubility of the full-length MtRamB and the aggregation of MtRamB<sub>69-474</sub> with the helix-turn-helix DNA-binding domain truncated similar to MtPrpR<sub>81-486</sub> obstructed a direct study of MtRamB. We instead used MtPrpR<sub>81-486</sub>\_F155H as a mimic based on the fact that the other CoA binding residues in both proteins are identical. MtPrpR<sub>81-486</sub>\_F155H remained a tetramer in solution as shown by the size exclusion chromatography (Fig. 31a) and crystallized in the same condition as



the wildtype protein. The overall structure of the polypeptide chain superimposed perfectly to the CoA-bound wildtype protein, including the helix  $\alpha 1$  where residue 155 is located ( $C_{\alpha}$ \_RMSD = 0.314 Å, Fig. 31c). The 2Fo-Fc and Fo-Fc electron density maps showed that the ligand was longer than CoA (Fig. 31d), indicated a fatty acyl CoA derivative was bound to this mutant protein. The ligand was purified and identified using LC-MS. A monoisotopic molecular ion peak with a  $[M+H]^+$  m/z value of 868.12 was detected (Fig. 31b) besides CoA. The amount of CoA was less than a quarter of the new compound according to the UV chromatogram (data not shown). This mass corresponds to both epimers of methylmalonyl-CoA as well as succinyl-CoA, which share an identical chemical formula. Attempting to build (R)- or (S)- methylmalonyl-CoA into the electron density resulted in a strong negative electron density peak ( $> 4 \sigma$ ) around the methyl branch in Fo-Fc map (Extended Data Fig. 31e). The methyl branch also caused steric clashes with adjacent residues His155 and Arg159, while succinyl-CoA nicely fit the electron density with no steric overlap (Fig. 31f). The terminal carboxyl group of the succinyl moiety interacts with His155 ( $\sim 4$  Å), Arg348, and Arg159 ( $\sim 2.8$ - $3.2$  Å) (Fig. 31g).



**Figure 31. Characterization of MtPrpR<sub>81-486</sub>\_F155H as a mimic of MtRamB**  
**a**, Size exclusion chromatography of MtPrpR<sub>81-486</sub> wildtype and F155H variant. **b**, Mass spectrum of the ligands. M/z and ligand identities are labeled (SCA for succinyl CoA; SCA+Na<sup>+</sup> for the sodium adduct form of SCA). **c**, Structure comparison between MtPrpR<sub>81-486</sub>\_WT and the F155H variant. The first visible helix  $\alpha 1$  in the two neighboring chains, where residue 155 is located, are indicated with black arrows. **d**, 2Fo-Fc map (blue, contoured at 1.0  $\sigma$ ) and positive Fo-Fc map (green, contoured at 3.0  $\sigma$ ) of MtPrpR\_F155H with CoA built into the electron density. **e**, 2Fo-Fc map (blue, contoured at 1.0  $\sigma$ ) and negative Fo-Fc map (green, contoured at 3.0  $\sigma$ ) of MtPrpR\_F155H with (S)-methylmalonyl-CoA built into the electron density. **f**, 2Fo-Fc electron density of succinyl CoA bound by MtPrpR<sub>81-486</sub>\_F155H mutant, electron density contoured at 1  $\sigma$ . **g**, Succinyl CoA binding environment. Dashed lines indicate the hydrogen bond and electrostatic interactions between protein and succinyl moiety.



**Figure 31. Continued.**

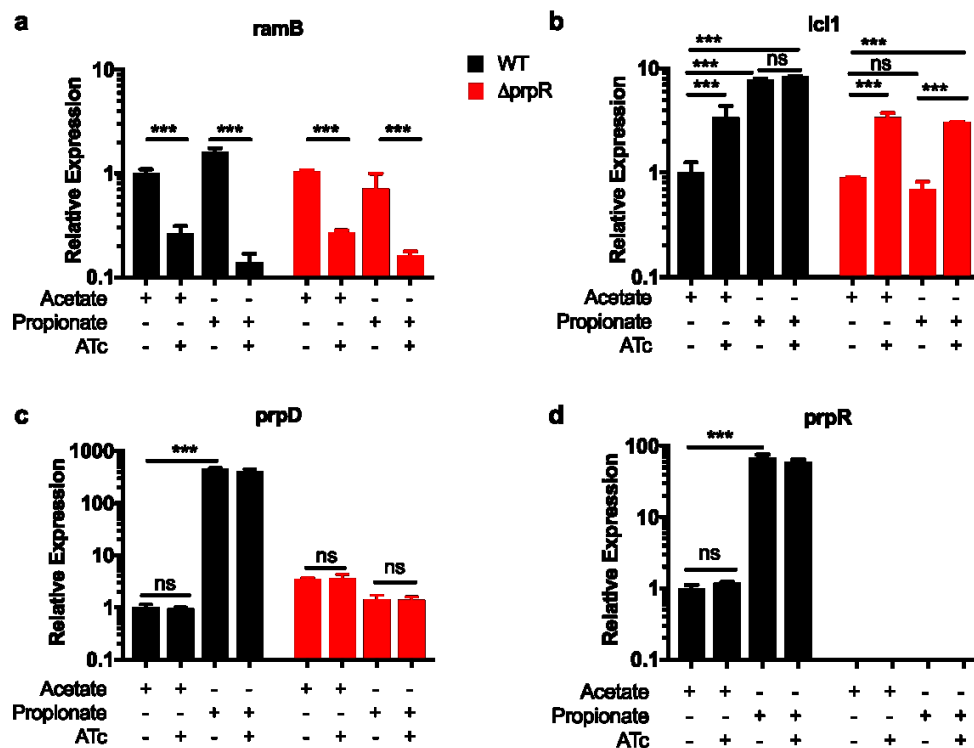
### II.3.7 *PrpR* and *RamB* cross-regulate *icl1* but not *MCC*

Given that *Mtb* has two homologous regulators, crosstalk between *prpR* and *ramB* is very likely. To understand if this was the case, we incorporated a tunable *ramB* knockdown into both wildtype and the *prpR* deleted backgrounds using a CRISPR interference (CRISPRi) system optimized for *Mtb* (151). The knockdown of *ramB* was confirmed as measured by a depletion of 80 to 90 % of its transcription after anhydrous tetracycline (ATc) induction (Fig. 32a).

Our studies showed that *icl1* transcription was controlled by both MtRamB and MtPrpR. MtRamB acted as a consistent repressor under all tested conditions and genetic backgrounds. By depleting *ramB*, *icl1* was upregulated by at least three-fold compared to the wildtype background in acetate media and without ATc induction regardless of the carbon source (Fig. 32b). MtPrpR, on the other hand, could upregulate *icl1* by about seven-fold, but the upregulation only occurred in response to propionate exposure,

irrespective of the level of *ramB* (Fig. 32b). This is because *icl1* is also the MCL (25, 127) in *Mtb*, it needs to be upregulated by MtPrpR during propionyl-CoA exposure.

Unlike *icl1*, *prpD* and *prpR*, which specifically participate in MCC, were only robustly upregulated by MtPrpR under propionate condition (Fig. 32c, d). Depleting *ramB* had no noticeable effect on the transcription level of *prpD* or *prpR* under either carbon sources.



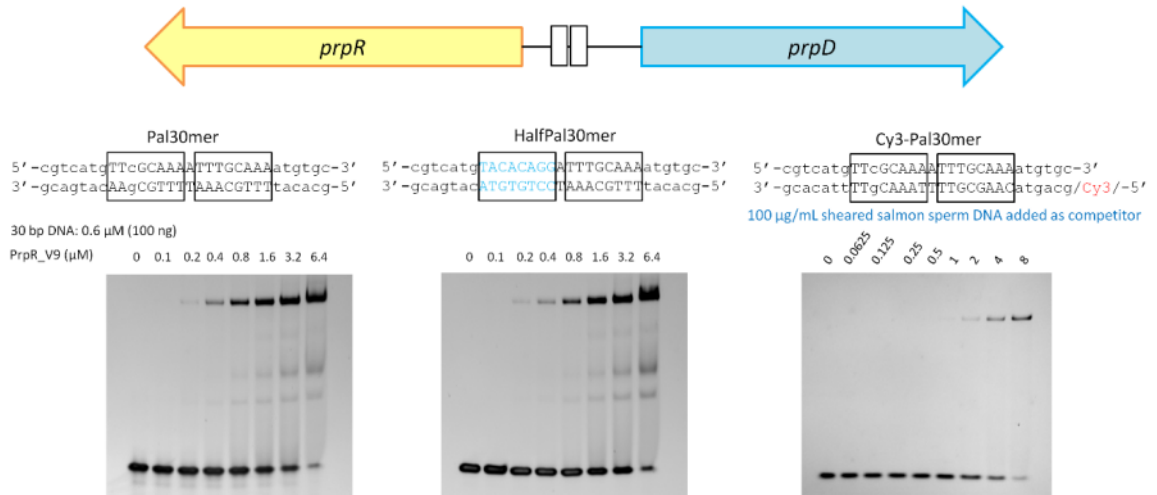
**Figure 32. Cross-regulation of MCC and glyoxylate shunt by *prpR* and *ramB***  
**a**, Transcription level of *ramB* in *Mtb* H37Rv or  $\Delta prpR$  strains with or without anhydrous tetracycline (ATc) induction and under acetate or propionate carbon sources. **b**, **c**, and **d**, Transcription levels of *icl1*, *prpD*, and *prpR* in the same situations as in **a**. Black bars are WT H37Rv transformed with ATc inducible vector. Red bars are H37Rv $\Delta prpR$  transformed with the same vector.  $N=2$  for each strain, treatment and carbon source combination. Data were analyzed with two-way ANOVA followed by Tukey's test. \*\*\*  $p=0.0002$ .

### II.3.8 *MtPrpR* may require propionyl-CoA to specifically bind to the promoter region of *prp* operon

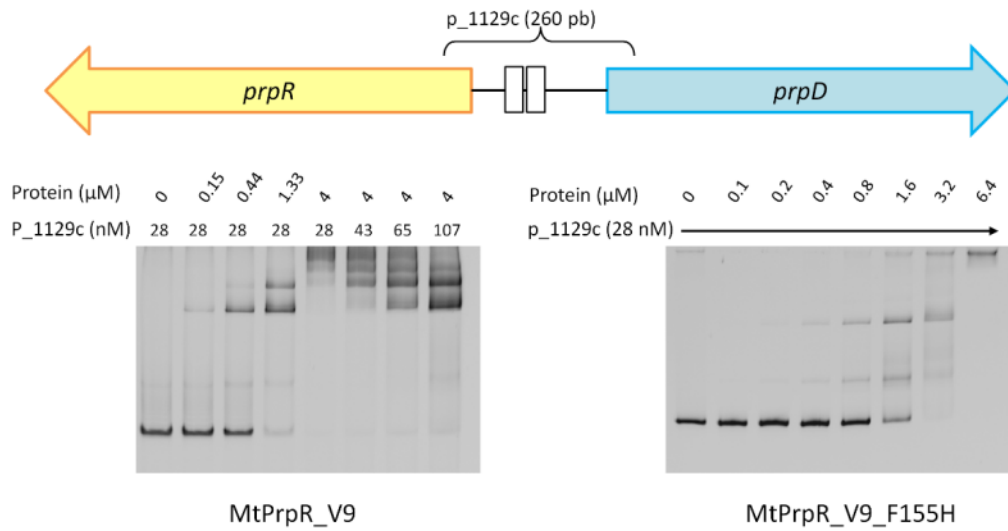
It has been shown that the recognition sites of both MtPrpR and MtRamB harbor tandem palindromic repeats TTCGCAAATTTGCAA (Fig. 14) (135, 148). Similar DNA sequences were also proposed as the operator sites of MtPrpR homologs in other organisms(150). To analyze the DNA binding by MtPrpR, the N-terminal His<sub>6</sub>-tagged MtPrpR\_V9 recombinant protein was used throughout EMSAs, since Val9 was predicted to be the first amino acid of the ORF after genome re-annotation, and MtPrpR\_V9 was more soluble than MtPrpR\_M1 (Fig. 16b). But similar to MtPrpR\_M1, MtPrpR\_V9 was still not very stable. Therefore, we had to purify fresh proteins each time for EMSA, and this led to some level of variability in the experiments due to batch-to-batch variability of the protein quality. We evaluated the binding between the purified MtPrpR\_V9 and a 30 bp oligonucleotides of the *prp* operon promoter region centered by the tandem repeats (Fig. 33). EMSA showed that the affinity between MtPrpR\_V9 and the tandem repeats was weak with a dissociation constant roughly estimated to > 800 nM based on the ratio of shifted and unshifted bands. In addition, the dual palindromic and half palindromic DNA pieces did not show noticeable difference in binding. When the sheared salmon sperm DNA (100 µg/ml) was added into the buffer to reduce nonspecific binding, the affinity became even weaker. This seems to indicate that the DNA binding in the tested condition was likely not specific, and thus was not physiologically relevant,. When the DNA length was extended to 260 bp to cover the promoter region (p\_1129c), the binding seemed to be slightly improved (Fig. 34). However, the F155H mutant form

could also bind to p\_1129c with similar affinity. Given that F155H does not bind to propionyl-CoA but preferentially binds to succinyl-CoA, we conclude that the poor binding that we observed was not relevant to the robust transcriptional activation induced by propionyl-CoA. In addition, MtPrpR could also bind to irrelevant DNA pieces under the experimental conditions (Fig. 35).

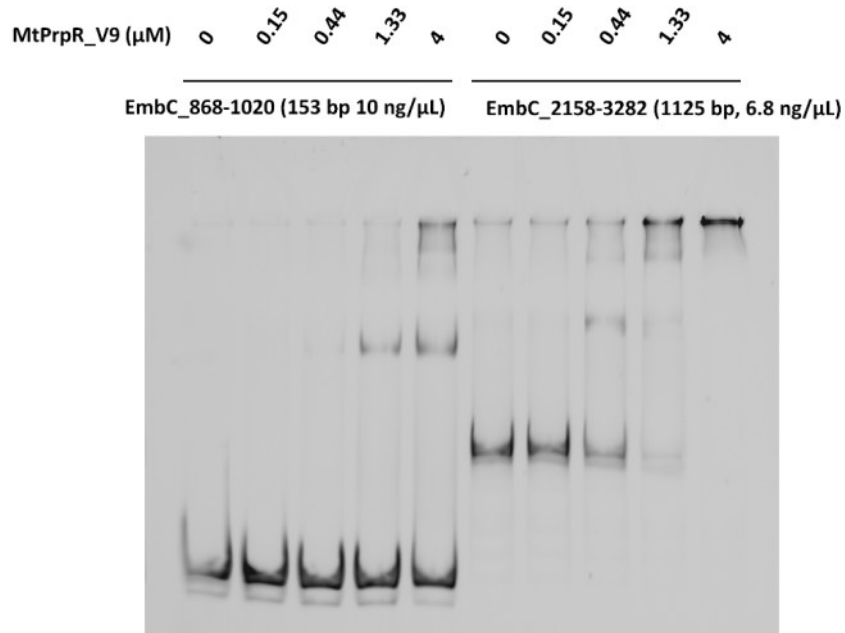
Next we investigated whether the small molecules that had been proposed as PrpR effector could affect the binding or not. The small molecules include 2-methylcitrate and cAMP. We picked an intermediate protein concentration where both free and shifted DNA fragments were present. This would allow us to tell easily if the DNA binding was enhanced or disrupted in the presence of the small molecules. We did not notice obvious difference in the DNA binding with the addition of either small molecule, although the MCC intermediate 2-methylcitrate at 500  $\mu$ M showed a more intense shifted band (Fig. 36). It is worth mentioning that the commercially available 2-methylcitrate is a racemic mixture of diastereomers, which means the bioactive forms may be one half of the total concentration being used. 2-methylcitrate did not significantly affect the shape or oligomeric state of MtPrpR<sub>81-486</sub> based on the SEC analysis (Fig. 18b). The effect of 2-methylcitrate may need further investigations.



**Figure 33. Binding of MtPrpR\_V9 to different forms of 30 bp DNA**  
**Binding between MtPrpR\_V9 and the original dual palindrome DNA (left) was similar to binding between MtPrpR\_V9 and the half palindrome DNA (middle). The DNA was stained with SYBR Green. When sheared salmon sperm DNA (sSSD) was added as competitor (right), the affinity appeared even lower. The intensities of the Cy3-Pal30mer bands were weak because we did not have a gel documentation instrument with light source or filter compatible with Cy3 dye, and staining with SYBR green gave too strong background due to sSSD.**

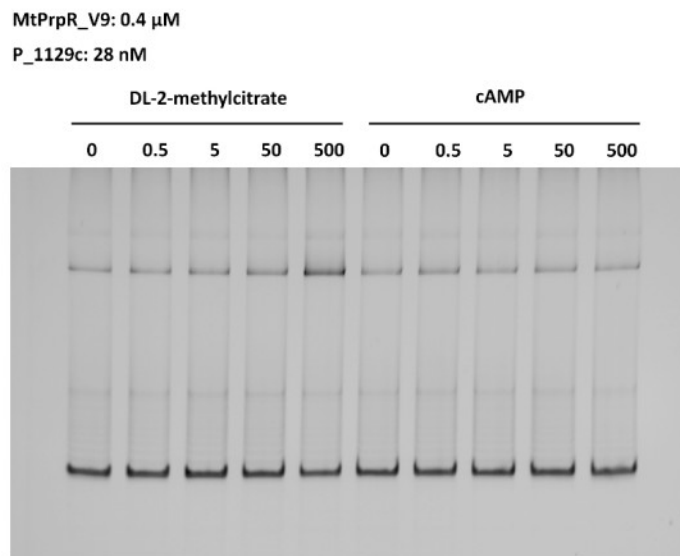


**Figure 34. Binding of wildtype and F155H variant MtPrpR\_V9 to the 260 bp promoter region of the *prp* operon**



**Figure 35. Binding of MtPrpR\_V9 to physiologically irrelevant DNA fragments of the *Mtb embC* gene**





**Figure 36. The effects of 2-methylcitrate and cAMP on the DNA affinity of MtPrpR\_V9**  
Neither molecule showed obvious trend of increased or decreased DNA affinity.

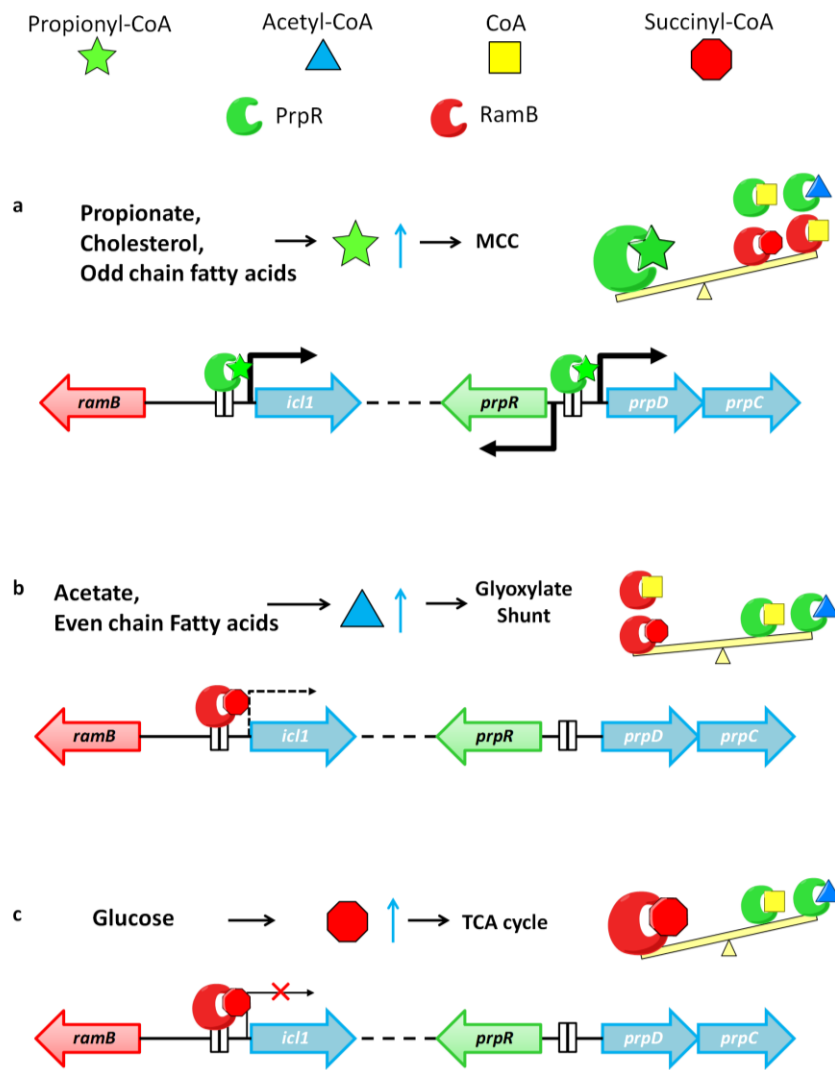
## II.4 Discussion

### II.4.1 Regulatory model of MCC and glyoxylate shunt by MtPrpR and MtRamB

*Mtb*'s dependence on lipid catabolism (22), especially cholesterol (34), is well documented, as is its reliance on MCC to assimilate propionyl-CoA and detoxify the metabolic intermediates (25, 129, 130). The local transcriptional regulator MtPrpR determines the on and off states of MCC (104, 135, 138). However, the triggering molecule that activates MCC has not been found yet, and consequently, the biochemical mechanism of the regulation has remained a mystery. In our studies, we showed that the major regulator of MCC, MtPrpR directly binds to CoA and three CoA derivatives, and

that the binding is under the control of an unexpectedly identified iron-sulfur cluster which was located at the C-terminal region of the protein.

All genetic studies so far pointed to propionyl-CoA as the triggering molecule of MCC since the pathway was only activated in propionyl-CoA producing conditions. However, our structure and LC-MS data showed that MtPrpR also readily bound to CoA and acetyl-CoA. This was not surprising because the CoA binding cavity of MtPrpR does not possess any dedicated feature, such as a potential hydrophobic pocket, to accommodate the extra propionyl group. Therefore, this strongly suggests that the selection of propionyl-CoA verse other CoA derivatives is a thermodynamic driven process mainly determined by the proportion of propionyl-CoA in the intracellular CoA pool. This explains the earlier finding that the relative balance, but not the absolute abundance, of the three- to two-carbon intermediates impacts the fitness of the bacilli (105). In addition, it becomes apparent that the transcription of all three MCC enzymes is controlled by the levels of intracellular short-chain fatty acyl-CoAs (SCFAs), especially propionyl- and acetyl-CoA (Fig. 32). We herein propose a regulatory model of the transcriptional regulation of MCC based on different carbon sources (Fig. 37).



**Figure 37. Schematic of MtPrpR/MtRamB regulation mediated by short chain fatty acyl CoA molecules**

**a**, Under propionyl-CoA-producing conditions, the active form of MtPrpR (MtPrpRa, propionyl-CoA bound) dominates the transcription of the *prp* operon and *icl1* to detoxify propionyl-CoA, regardless of the presence of MtRamB (succinyl-CoA or CoA bound) and inactive forms of MtPrpR (MtPrpRi, CoA or acetyl-CoA bound). **b**, Under acetate and even chain fatty acids conditions, acetyl-CoA in large quantity would be produced. MtPrpRi cannot upregulate *prp* operon or *icl1*, while MtRamB could still repress the transcription of *icl1* to some extent. But since the glyoxylate shunt is required, the *icl1* transcription needs to be maintained at a certain level. **c**, During dextrose utilization, succinyl-CoA is produced in TCA cycle. MtRamB with succinyl-CoA bound is expected to repress the transcription of *icl1*, which is not required in this case.

One remaining question is what ratio of intracellular propionyl- to acetyl-CoA can turn on MCC. Although we were not able to investigate it biochemically, a previous *icl1* deleted *Mtb* whole cell experiment showed that over 80-fold molar excess of acetate in the growth media of *Mtb* was required to remove the growth inhibition by propionate (105). As a comparison, each molecule of cholesterol, a prominent carbon source of *Mtb in vivo* (34), was predicted to produce four molecules each of acetyl- and propionyl-CoA (57, 71). In this case, the propionyl-CoA level would be much higher than the threshold to turn on MCC. This was recently demonstrated during macrophage infection (41).

An intriguing fact about *Mtb* is that the bacilli also possess a *prpR* paralog, *ramB*, whose protein product MtRamB is a transcriptional repressor of *icl1* during dextrose utilization (148). Our experiments showed that MtRamB likely bound to succinyl-CoA, a TCA cycle intermediate. It is known that *Mtb* synthesizes succinyl-CoA from  $\alpha$ -ketoglutarate (182) using both the KG dehydrogenase (KDH) (183, 184) and KG:ferredoxin oxidoreductase (KOR)<sup>47</sup>. We propose that an increased level of succinyl-CoA may act as a permit for a temporary shutdown of glyoxylate shunt mediated by MtRamB. Furthermore, it was recently reported the detection of Icl1 lysine succinylation when *Mtb* was cultured in a rich medium with albumin-dextrose-catalase as the carbon source (185). The succinylation could potentially decrease the enzymatic activity of Icl1 (185, 186). Taken together, it appears that an increased level of succinyl-CoA reduces the concentration of active *icl1* at both the mRNA and the protein levels. Notably, succinyl-CoA is also produced during the breakdown of the C/D-ring of cholesterol. However, the MCC is enforced to be upregulated during cholesterol utilization

regardless of the presence of succinyl-CoA, indicating that the degradation and detoxification of propionyl-CoA is a higher priority for *Mtb*.

#### II.4.2 *The role of iron-sulfur cluster*

A sequence based classification suggested that MtPrpR homologs form a distinct clade of the short-chain fatty acyl-CoA regulators (ScfRs) (150). MtPrpR contained an iron-sulfur cluster that was ligated by a featured C-X<sub>2</sub>-C-X<sub>4</sub>-C motif in the C-terminal region of the protein. The motif appears to be conserved among all MtPrpR homologs, which leads to the postulate that all members of this clade bind to the iron-sulfur cluster. We herein suggest the term iron-sulfur-dependent ScfR (IssR) for this clade of regulators in order to distinguish them from the other non-iron-sulfur ScfR clades (144, 181). We showed that without the iron-sulfur cluster, MtPrpR could not bind to CoA derivatives or activate MCC. This was supported by a recent *Mtb* transcriptome study (187). The transcription of *prp* operon showed strong positive correlation with the iron concentration (187). Since the cluster was extremely stable under air exposure, it is unlikely to respond to oxidation stress in the canonical manner. Stable iron-sulfur clusters have been previously reported in several proteins (188), among which a most well-studied case was the *E. coli* endonuclease III. The iron-sulfur cluster in endonuclease III is also completely ligated by cysteinyl sulfur. The cluster had been regarded as to serve structural role (189, 190) for more than a decade before the advent of DNA-mediated charge transport that demonstrated the cluster was to sense DNA lesion (191). Although MtPrpR does not seem to bear enzyme function, the role of the

iron-sulfur cluster is potentially not limited to structural. Its biochemical function remains to be learned.

#### II.4.3 *The protein folding of IssR in other proteins*

Our structure of the MtPrpR represents a prototype of the IssR transcriptional regulators. Moreover, it shed light on the functions of other structurally related proteins. The 3D-folding of the combined D<sub>II</sub>-GAF di-domain was previously observed in the *Deinococcus* IrrE (146), a radiation responsive transcriptional regulator (192). So far no small molecule ligand has been reported to bind to IrrE. Notably, similar to the CoA binding cavity in the MtPrpR structure, IrrE structure shows a wide cleft between its proposed Zn-dependent metalloprotease domain (corresponding to D<sub>II</sub> in MtPrpR) and the GAF-like domain (Fig. 22a, b). Mutations in these domains sensitized the bacteria to irradiation. Implanting the structural information of MtPrpR into IrrE, it is likely that the cleft is the nest of a light sensing chromophore. Taken together, despite the sequence dissimilarity, the architecture of MtPrpR could be repurposed and may be widely used to harness different signaling pathways.

## CHAPTER III

### NATURAL PRODUCT MZ721 IS A POTENT PROTEASOME INHIBITOR

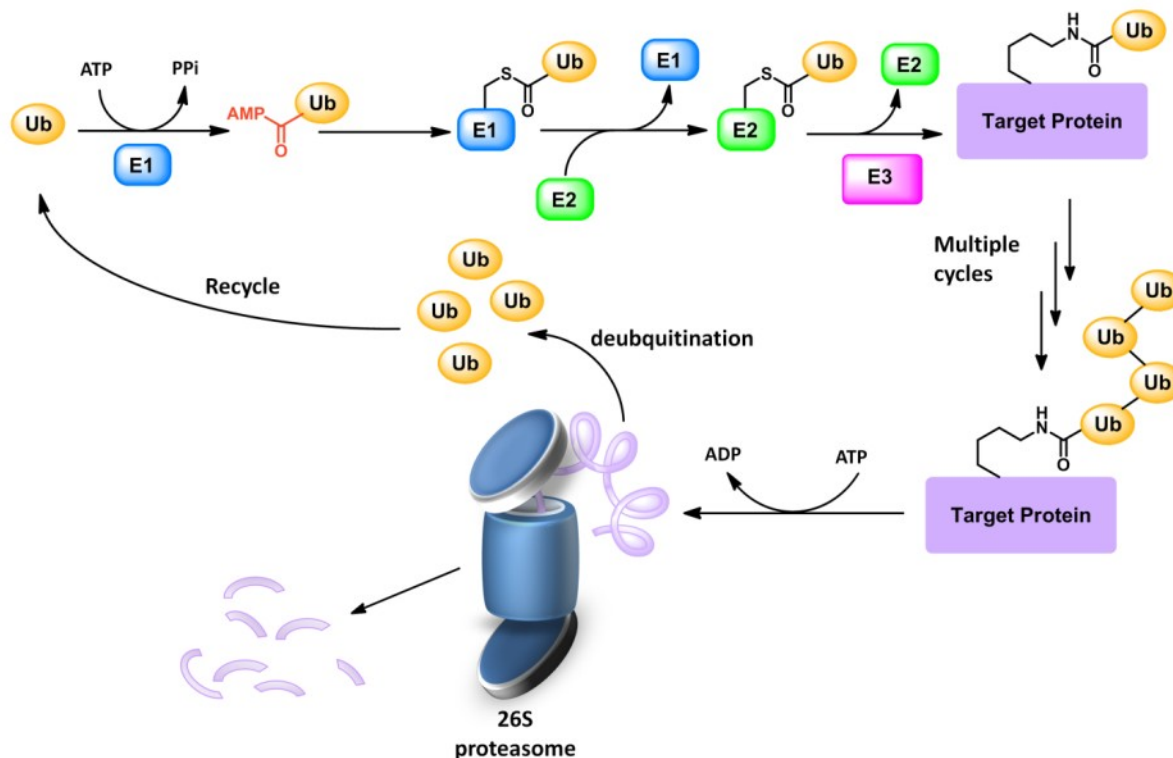
#### **III.1 Introduction**

##### *III.1.1 The Ubiquitin-Proteasome Pathway*

For a very long time, intracellular proteins were largely believed to be long-lived and barely turned over, although the lysosome-dependent proteolysis could degrade a small portion of total proteins, typically exogenous proteins that are endocytosed into the cell (193, 194). This sacrosanct principle was barely challenged until the 1970s, when the ubiquitin-proteasome pathway was identified. In 1977, Etlinger and Goldberg observed that abnormal proteins containing valine and lysine analogs were degraded much faster than native proteins in rabbit reticulocytes and cell free systems. They also confirmed that the protein degradation was ATP-dependent (195). One year later, Ciechanover et al. fractionated the protein degradation system into a heat-stable fraction I and heat un-stable fraction II. Neither fraction had significant protein degradation activity. However, once they were combined, in the presence of ATP, the degradation system was restored (196). The heat-stable fraction I was therefore renamed ATP-dependent proteolysis factor-1 (APF-1). APF-1 was later on confirmed to have the same identity as ubiquitin (197), which had been found earlier to covalently attached to the histone H2A protein (198), although the biological significance of this attachment was unknown. Ubiquitin is a 76-amino-acid polypeptide whose C-terminal carboxyl group

could be covalently conjugated to protein lysine residues via an isopeptide bond (198, 199). Studies focusing on the fraction II further identified an ATP-dependent cascade including E1-E2- E3 ligases that conjugate ubiquitin molecules to the target proteins (199-203). In fraction II, there was another ATP-dependent component which exhibited protease activity, and it was later known as the 26S proteasome (204-210). The process of the ubiquitinated protein degradation is summarized in Fig. 38. Specifically, the C-terminal carboxyl group of ubiquitin is first adenylated consuming an ATP molecule and forms an acyl-AMP. The phospho-carboxylic anhydride bond is then substituted by a thioester bond between the C-terminal carboxyl group of ubiquitin and a cysteinyl thiolate of the E1 ligase (199, 200, 202). The ubiquitin is then passed onto a cysteinyl residue of the E2 ligase to form the second thioester bond. This step does not consume ATP (203). The E3 ligase conjugates the ubiquitin from the E2 to the  $\epsilon$ -amino groups of lysine residues of the target proteins, and this step is also ATP-independent (203, 205). Subsequent ubiquitinations occur through the Lys48-Gly76 isopeptide bond between two ubiquitin molecules (211). Proteins are successively ubiquitinated to reach a minimum of tetraubiquitin for efficient proteasomal targeting (212). The poly-ubiquitinated proteins will then be recognized by the proteasome 19S regulatory particle (19S RP, also termed 19S proteasome) in an ATP-dependent manner (213). The protein is then deubiquitinated, unfolded and shuttled into the lumen of the proteasome 20S core particle (20S CP, also termed 20S proteasome), where they are degraded into peptides ranging from 3 to 25 amino acids (214, 215).



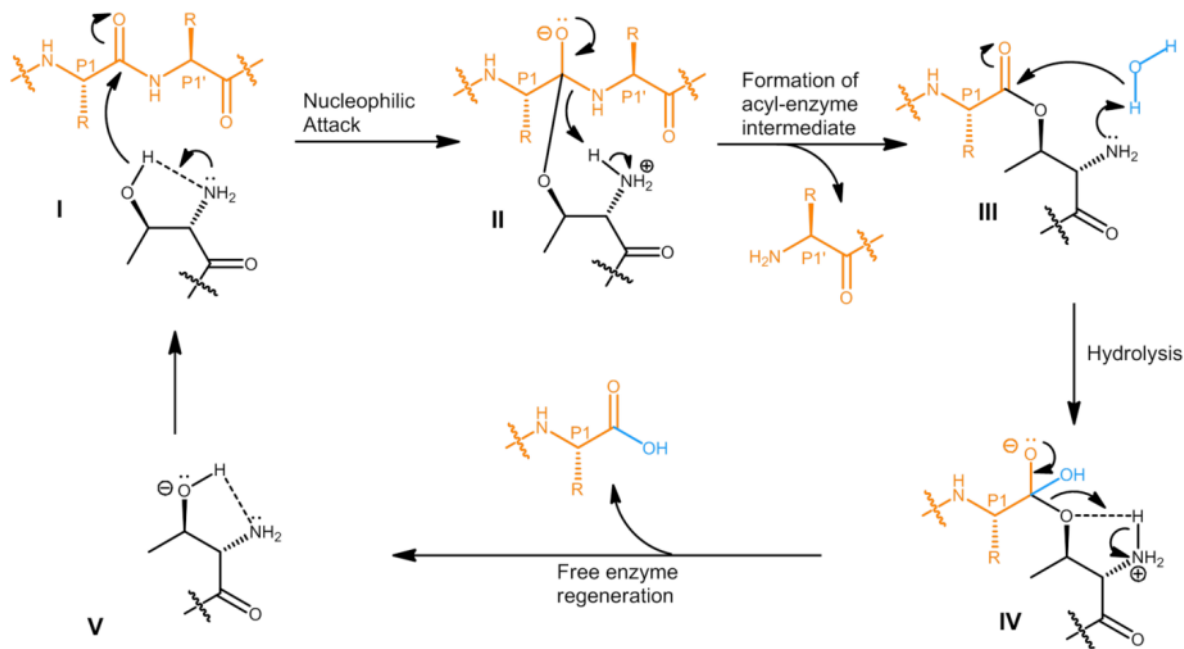


**Figure 38. Schematic of the ubiquitin-proteasome system**

### III.1.2 Structure and Function of Proteasome

The 26S proteasome holoenzyme is a protease complex of approximately 2,000 kDa (206-210). The holoenzyme consists of a cylindrical 20S CP with 19S RPs capping at each end (216). The 19S RP can be subdivided into a lid and a base. The lid is responsible for poly-ubiquitinated substrate recognition and ubiquitin removal. The base is attached to the CP (216), and contains multiple ATPases for ATP-dependent opening of the channel in the  $\alpha$ -ring of the 20S proteasome. The target proteins are translocated into the CP for breakdown (217). The 20S CP is composed of two sets of 14 individual proteins stacked into four heptameric rings under the assembly of  $\alpha_7\beta_7\beta_7\alpha_7$  (209, 218,

219). The eukaryotic 20S CP harbors three types of proteolytic activity on each  $\beta$ -ring, designated as postglutamyl-peptide hydrolyzing (PGPH, also termed caspase-like, C-L) activity for  $\beta$ 1 subunits, trypsin-like (T-L) activity for  $\beta$ 2 subunits, and chymotrypsin-like (CT-L) activity for  $\beta$ 5 subunits. They are named for their preferential cleavage sites immediately after acidic, basic, and hydrophobic residues respectively, resembling their serine protease counterparts (204, 219-222). Unlike serine proteases, all three active sites of the proteasome employ the hydroxyl group of N-terminal threonines (Thr1) as their catalytic nucleophile, likely because the secondary hydroxyl group of Thr1 allows for more efficient proteolysis than the primary alcohol group of serine (223). Serine proteases use the catalytic triad including serine, histidine and aspartate to catalyze the hydrolysis, whereas proteasomes use the catalytic diad consisting of the side chain hydroxyl group of the Thr1 and the N-terminal amine (Fig. 39).



**Figure 39. Mechanism of proteolysis catalyzed by proteasome Thr 1**  
 Proteasome Thr1 is colored is black. Protein or peptide substrate is colored in orange. State I: the N-terminal amine abstracts the proton of the Thr1 side chain hydroxyl group, and generates the hydroxylate nucleophile; State II: formation of the enzyme-substrate tetrahedral intermediate; State III: cleavage of the downstream peptide substrate, leaving the acyl-enzyme intermediate; State IV: hydrolysis of the acyl-enzyme intermediate and formation of the hydroxylated tetrahedral intermediate; State V: release of the upstream peptide substrate and regeneration of the free enzyme for another round of proteolysis.

### III.1.3 UPS Functions in Tumorigenesis

Proteasome has been considered as an important anti-cancer drug target.

Inhibition of proteasome kills multiple cancer cell lines and rescues mice from various tumors (224). The mechanisms of why tumor cells are more susceptible to proteasome inhibitors are not fully understood, but it has been well accepted that tumor cells have a faster compound uptake, and inhibiting the proteasome activity could arrest the cell

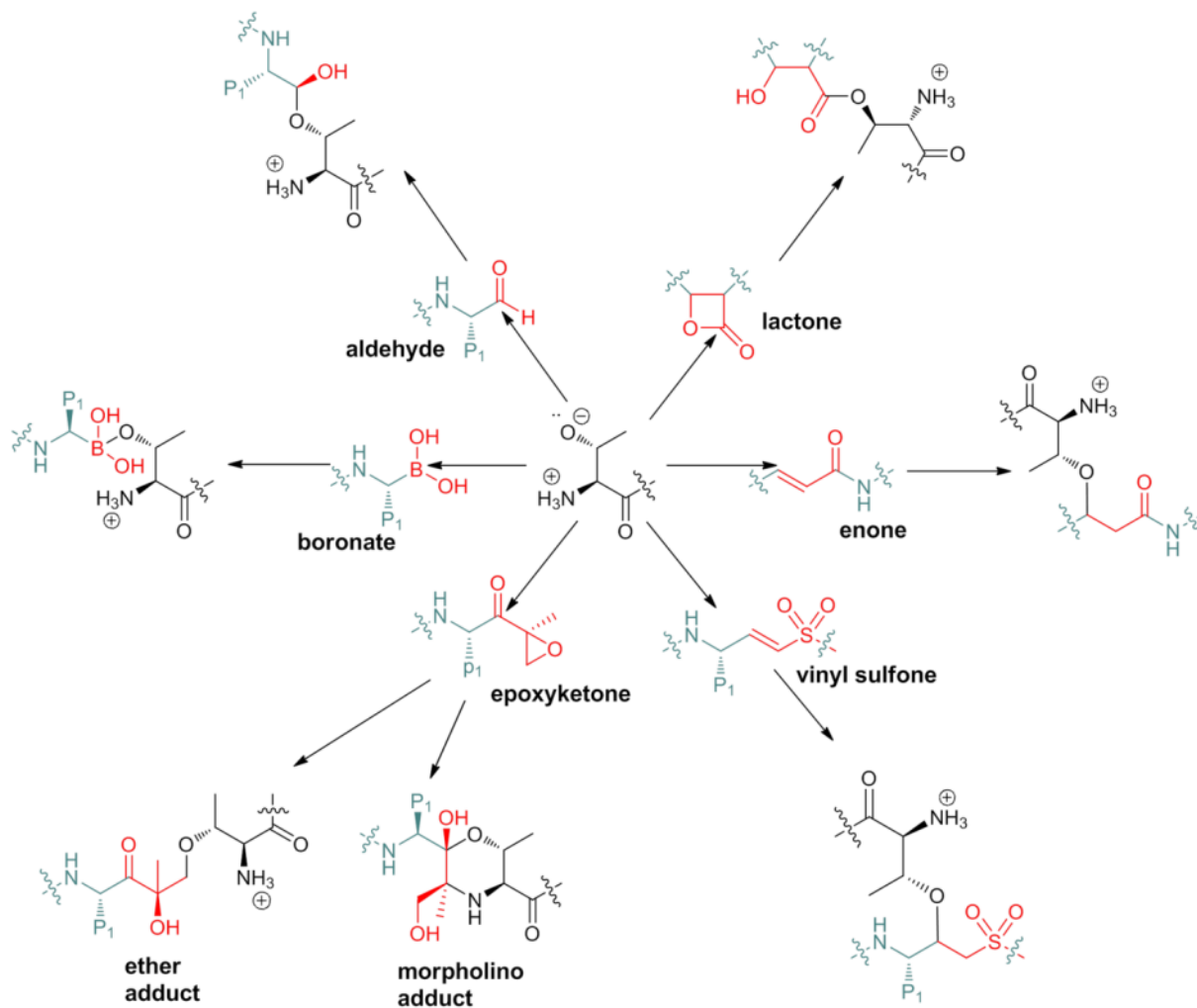
cycle and induce apoptosis (224). At cell cycle checkpoints, positive regulators such as cyclin proteins, and negative regulators, p27 and p21<sup>WAF1/CIP1</sup>, need to be degraded to drive the cell proliferation forward. Digestion of these proteins is executed by the proteasome (224-226). On the other hand, for apoptosis, the proteasome also plays the crucial role by activating the transcription factor NF- $\kappa$ B (Nuclear Factor- $\kappa$ B). NF- $\kappa$ B, a heterodimer consisting of p65 and RelA, is deactivated in normal situations by forming a complex with its inhibitory protein I $\kappa$ B $\alpha$  (Inhibitory  $\kappa$ -B chain  $\alpha$ ) via its p65 subunit (227-230). In tumor cells, extracellular stimuli such as TNF- $\alpha$  (Tumor Necrosis Factor  $\alpha$ ) trigger I $\kappa$ B $\alpha$  phosphorylation catalyzed by I $\kappa$ B kinases (IKKs). The phosphorylated I $\kappa$ B $\alpha$  (pI $\kappa$ B $\alpha$ ) is a substrate of E3 ubiquitin ligase. pI $\kappa$ B $\alpha$  is poly-ubiquitinated for concomitant proteasome-dependent degradation (231-233). Therefore, by hydrolyzing I $\kappa$ B $\alpha$ , NF- $\kappa$ B is free from inhibition and exposes its nuclear localization signal peptide (234). NF- $\kappa$ B then migrates into the nucleus and acts as a transcription factor activating anti-apoptosis gene, primarily the gene encoding Bcl-2 (B-cell lymphoma-2) anti-apoptotic protein. By inhibiting the proteasome, I $\kappa$ B $\alpha$  eludes proteolysis, preventing NF- $\kappa$ B from being activated (235-237).

The site specific inhibitors of the proteasome have shown distinct cytotoxicity profile in cancer and normal cells, thus revealing the importance of the individual active site. CT-L activity was proposed to be rate-limiting for protein degradation, and its inactivation caused severe growth deficiency in a yeast model (238).

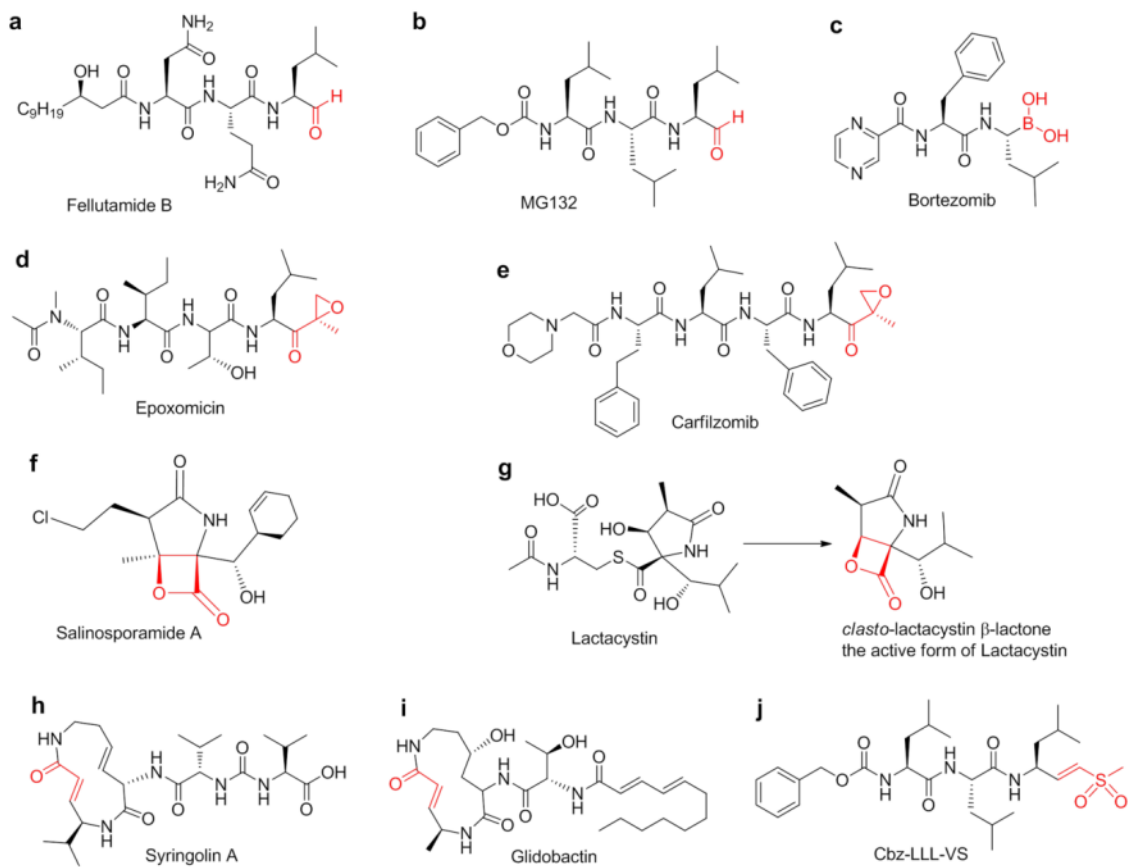
#### III.1.4 *Proteasome Inhibitors*

There have been numerous proteasome inhibitors identified, which were either extracted from natural sources or chemically synthesized in the laboratory. Most proteasome inhibitors form covalent adduct with the side chain hydroxyl group of the Thr1. The reactive moieties of the inhibitors are known as the warheads. The common warheads include aldehyde (i.e. MG132, fellutamide B), boronate (i.e. bortezomib),  $\alpha$ ,  $\beta$ -epoxyketone (i.e. epoxomicin, carfilzomib),  $\beta$ -lactone (i.e. lactacystin (needs to be activated), salinosporamide A), and Michael acceptors including vinyl sulfone and enone (also termed  $\alpha$ ,  $\beta$ -unsaturated carbonyl, i.e. Syringolin A, Glidobactin) (224) (Fig. 40, 41).

One of the early proteasome inhibitors, MG132, a tripeptide aldehyde synthesized by MyoGenics (MG, later renamed to ProScript, PS) (239), is now widely used in research. MG132 attacks all three active sites by mimicking the acyl-enzyme transition state. MG132 preferentially blocks the CT-L site with high affinity ( $K_i=4$  nM) (240). However, there are several drawbacks for these aldehyde containing inhibitors to become drugs. The aldehyde moiety can be easily oxidized to acid form and lose the activity. The hemiacetal formed with Thr 1 is readily reversible. The aldehyde can also be nucleophilically attacked by cysteine proteases and serine proteases, thus weak target specificity (241).



**Figure 40. Major categories of covalent inhibitors and the corresponding adducts**



**Figure 41. Examples of the covalent proteasome inhibitors a and b, aldehyde; c, boronate; d and e,  $\alpha', \beta'$ -epoxyketone; f and g,  $\beta$ -lactone; h and i,  $\alpha, \beta$ -unsaturated carbonyl; j, vinyl sulfone. Reactive moieties are colored red.**

A group of boronic acids was also synthesized (224), and one of these compounds, bortezomib (242) (PS-341, marketed as Velcade) is now widely used in cancer chemotherapy. Bortezomib has a target priority of CT-L > C-L >> T-L, with the CT-L most selective and T-L barely touched (240). Bortezomib targets the CT-L site with affinity comparable to MG132. The crystal structure of bortezomib with the yeast 20S proteasome showed that the boronic acid moiety of the compound covalently

attached to the hydroxyl group of the catalytic Thr 1 of all active sites, possibly due to the high concentration of the inhibitor used for crystal soaking (243). It has also been reported that bortezomib inhibits NF- $\kappa$ B activation and induces caspase-dependent apoptosis in a variety of cell lines. Bortezomib exhibits a remarkable efficacy for treating multiple myeloma (MM) (237, 244-247). As a result, bortezomib became the first-in-class therapeutic proteasome inhibitor tested in humans (248, 249), and the first FDA approved drug targeting proteasome for treating MM (250). However, the compound causes a number of toxic side effects including fatigue, peripheral neuropathy, thrombocytopenia, and heart problems.

(<http://www.velcade.com/Treatment-with-velcade/Possible-side-effects>) The toxicity was likely due to the promiscuous activity on other proteases.

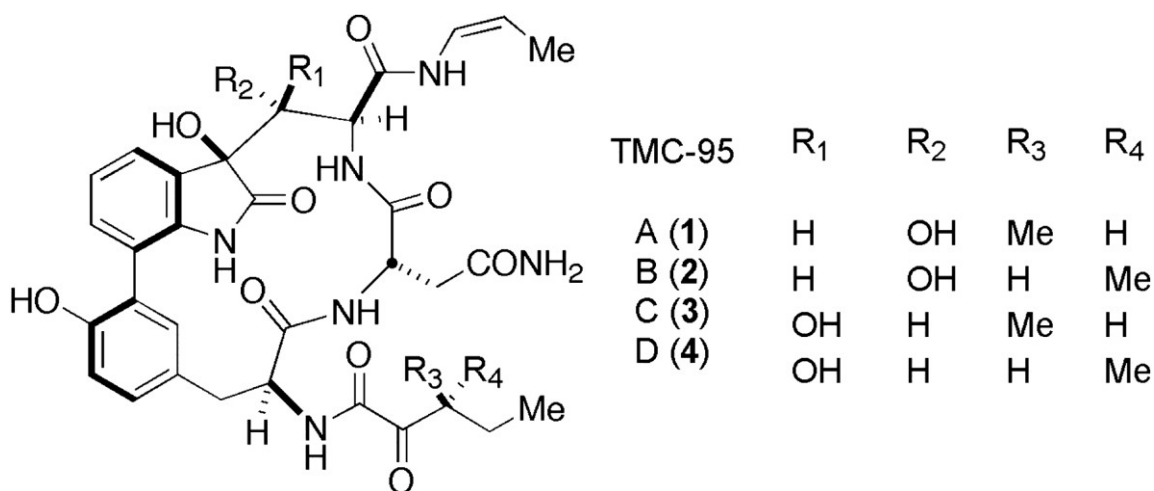
To reduce the toxicity, compounds displaying higher target specificity against the proteasome were exploited, and these include the  $\alpha'$ , $\beta'$ -epoxyketone peptides (241, 251). A natural product epoxomicin is a representative of this class, and it reacts primarily with the CT-L site. The crystal structure of the yeast 20S proteasome-epoxomicin complex revealed that besides forming an expected ether adduct between the Thr1 hydroxyl group and the  $\beta'$  carbon (Fig. 41), the epoxyketone moiety also unexpectedly formed a morpholino adduct with the hydroxyl and amino groups of Thr 1 (252) (Fig. 41). Carfilzomib, a synthetic analog of epoxomicin that is more potent than bortezomib (253), was approved by the FDA in 2012 to treat MM patients that are resistant to bortezomib (<http://www.cancer.gov/cancertopics/druginfo/fda-carfilzomib>). Despite of the great efficacy of bortezomib and carfilzomib, only 25% response was seen for both



drugs as a single agent treating relapsed and refractory MMs. These results indicated an innate or acquired resistance to the drugs has emerged (254). Since drug dosage currently is still empirically based on the patient population response, personalized, or precision, medicine is difficult to achieve (255), new drugs with higher therapeutic indices are desirable.

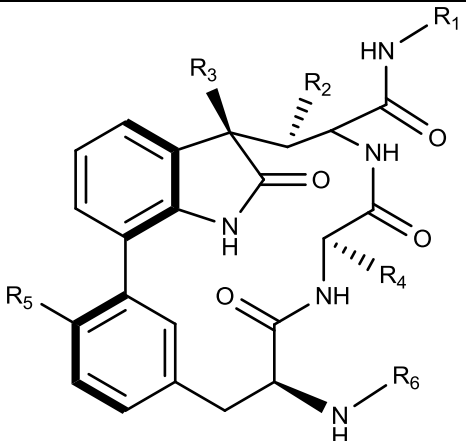

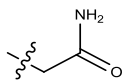
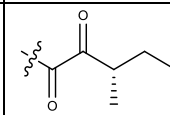
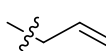
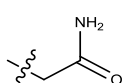
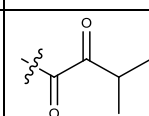
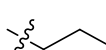
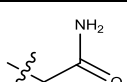
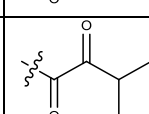
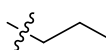
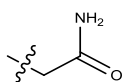

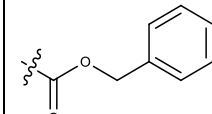
TMC-95A and its diastereomers TMC-95B, C and D are a series of natural cyclic tripeptides (Fig. 42) (256, 257). TMC-95 series specifically inhibit the proteasome rather than non-proteasomal proteases. The inhibition exhibited strong stereoselectivity. In particular, TMC-95A was more potent than TMC-95B, and TMC-95C and D had much weaker activity compared to the former two. TMC-95A showed high potency and versatility against all three active sites of the proteasome with  $IC_{50}$  values ranging from low to mid nanomolar (5.4 nM for CT-L site, 60 nM for C-L site, and 200 nM for T-L site) in the presence of 0.03% SDS (256, 257). The crystal structure of TMC-95A with the yeast 20S proteasome revealed that the compound is a non-covalent inhibitor residing in the pockets of all three active sites (258). The high affinity binding is likely achieved by the constrained conformation of the cyclic peptide backbone and the formation of a host of hydrogen bonds between the inhibitor and the main chain atoms of the protein. The binding is entropically favored since it neither rearranged the ligand nor significantly changed the conformation of the protein (258, 259). Currently, the total synthesis of TMC-95A/B is available but not commercially practicable (260-262). Since TMC-95A is such an informative lead compound and a promising anti-tumor candidate, modifications have been made based on the cyclic backbone scaffold. However, the total

synthesis of TMC-95A was so challenging that it hindered the structure-activity relationship (SAR) studies. As a result, simple chemical modifications were only incorporated during the final steps of the synthesis, particularly around the P1 position (R1 group in Table 4) (262-264). Therefore, the diversity of the compounds was very limited. To simplify the synthesis, especially the Suzuki coupling reaction to form the biaryl linkage, biphenyl ether linkage was instead used to cyclize the backbone (259, 265). Although in this case, the P1 and P3 sites were able to be modified with higher diversity (R1 and R2 in Table 5), the potency of the modified compounds dramatically decreased. This loss of potency could be attributed to the lower rigidity of the biphenyl ether bond than the biaryl clamp (265). So far, none of the TMC-95A synthetic derivatives improved the potency against the three catalytic sites (Tables 4 and 5) (262-268).

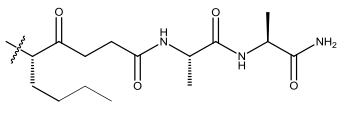
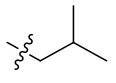

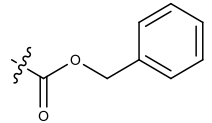
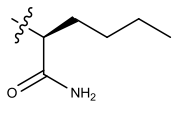
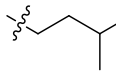

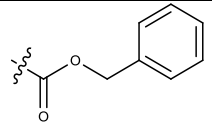


**Figure 42. Structures of TMC-95A—D, adapted from (269) with permission Copyright (2004) National Academy of Sciences.**

**Table 4. Structure-activity relationship of TMC-95A based on previous literatures**

							
Compound	R1	R2	R3	R4	R5	R6	Ki or IC <sub>50</sub> for CTL, TL, PGPH (nM)
TMC-95A		OH	OH		OH		5.4; 200; 60 (256)
							1.1; 800; 29 (262)
1		OH	OH		OH		1.9; 1,200; 23 (262)
2		OH	OH		OH		24; 13,000; 110 (262)
3		H	H				8,000; 106,000; 7,400 (263)
							2,400; 55,000; N/A (264)
							550; 150; 90,000 (266)

**Table 4. Continued**

Compound	R1	R2	R3	R4	R5	R6	Ki or IC <sub>50</sub> for CTL, TL, PGPH (nM)
4		H	H				9,100; 60,000; N/A (264)
5		H	H				2,400; 22,000; N/A (266)

N/A:  $\geq 2,000,000$

**Table 5. Structure-activity relationship of a backbone modified TMC-95A analog**

Compounds	R1	R2	R3	R4	Ki or IC <sub>50</sub> for CTL, TL, PGPH (nM)
1			NO <sub>2</sub>		5,500; 74,000; ≥ 2,000,000 (265)
2			NO <sub>2</sub>		65,000; ≥ 2,000,000; ≥ 2,000,000 (265)
3			NH <sub>2</sub>		N/R (267)
4			NO <sub>2</sub>		N/R (267)
5			NH <sub>2</sub>		N/R (267)

N/R: not reported.

In this study, we identified a natural product from the fungus *Auxarthron umbrinum* (originally named strain F01804) that could strongly inhibit the 20S proteasome. The compound has a  $[M+H]^+$   $m/z$  value of 721.33 in ESI positive mode mass spectrometry (tentatively named Mz721). The crystal structure of yeast 20S proteasome in complex with Mz721 showed that Mz721 possesses a peptide backbone that was cyclized via a biaryl linkage similar to TMC-95 series. The major difference between Mz721 and the TMC-95 series was at the P3 position, where Mz721 had a longer side chain. By comparing the proteasome-Mz721 complex structure with the published proteasome-TMC-95A structure, we found some alterations of hydrogen bonds between Mz721 and the protein active sites. In vitro bioactivity assays showed that Mz721 preferentially inhibits the T-L activity with below five nanomolar  $IC_{50}$ . Mz721 exerts markedly lower cytotoxicity than TMC-95A and bortezomib on human dermal fibroblasts (HDF). The discovery of this new natural product is highly significant in that the inhibitor has a potentially larger therapeutic window, and the unique modification at P3 position is informative in guiding the synthesis of more potent and active site-selective analogs.

## **III.2 Materials and methods**

### *III.2.1 Proteasome inhibitors and fungal strain acquisition*

A high-throughput screening against a library of 105,500 crude extracts to look for proteasome inhibitors was previously performed by the scientists in the biotechnology company Cetek Corporation in Marlborough, Massachusetts.

Fungal strain with the code name F01804 showed up as a hit, and was isolated by the scientists at the Cetek Corporation and had been identified to produce secondary metabolites that inhibited the 20S proteasome. When Cetek was closed, Ohr Pharmaceuticals acquired the assets of Cetek and licensed the intellectual property around some bioactive metabolites to Texas A&M University (AgriLife Research) for possible development. Small amount of the metabolites were semi-purified and were predicted to be structural analogs of the known proteasome inhibitor TMC-95 series (256, 257). Before that research was completed, Ohr Pharmaceuticals sold the natural product assets of Cetek to Dow AgroSciences LLC, and the fungal strain was acquired for metabolite production purposes under a material transfer agreement (MTA) negotiated with Dow AgroSciences.

Since a major metabolite showed a mass to charge ratio of 721, it was tentatively named Mz721 during the following research.

### III.2.2 *Taxonomy, strain identification*

The mycelial suspensions of the F01804 was inoculated into a 200 ml erlenmeyer flask containing 30 ml PYG medium (0.125 % (w/v) of soybean peptone, 0.125 % (w/v) of yeast extract, and 0.5 % (w/v) of dextrose) with or without 0.4 % of calcium carbonate to neutralize the produced acids during growth. The culture was incubated at 26 ° C for one week with gentle shaking at < 100 rpm.

An estimate of three grams of the fungus was collected by centrifuging at 2,000 rpm for 20 min. The medium was aspirated. The pellet was washed with 20 mM Tris-

HCl pH 7.5, transferred into a sterilized mortar, and flash frozen with liquid nitrogen. The fungus was ground into powder with a sterilized pestle to disrupt the cell wall and membrane. The powder was resuspended with 10 ml of the buffer containing 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.3 % SDS, and the cell debris was pelleted by centrifugation. The supernatant was transferred to clean tube for genomic DNA extraction.

The genomic DNA extraction of F01804 followed a general protein denaturation and DNA precipitation procedure (270). Basically, 600  $\mu$ l of the above supernatant was mixed with 200  $\mu$ l of 5 M NaCl. 160  $\mu$ l of CTAB solution was preheated at 65 °C and added to the above mixture. After that, one volume of 24:1 chloroform/isoamyl alcohol solution was added and vigorously mixed. The mixture was centrifuged at 13,000 rpm for 5 min to separate the phase. About 900  $\mu$ l of the upper aqueous phase was transferred to a fresh tube and the extraction was repeated again. About 800  $\mu$ l of the aqueous phase was transferred to a fresh tube and 560  $\mu$ l of isopropyl alcohol was added to precipitate the genomic DNA. The tube was placed at 4 °C for five minutes to accelerate the precipitation procedure. The DNA was pelleted by centrifugation at 13,000 rpm for 10 min. The solution was aspirated and the DNA pellet was washed with 1 ml of 70 % ethanol and centrifuged again. The supernatant was removed and the DNA was air dried. 50  $\mu$ l of DNase free pure water was used to dissolve the DNA pellet.

Around 40 ng/ $\mu$ l final concentration of the extracted genomic DNA was used as the template for PCR reactions. The fungal conserved primers were used (listed in Table 6) to amplify the ribosomal DNA (rDNA). Primers 5.8SR+LR7 and LROR+LR7 were



used to amplify the large subunit (LSU) 28S rDNA and the PCR products were sequenced with LROR, LR3R, LR6 and LR7. The sequencing results were compared against the sequences from the type material of the Targeted Loci Nucleotide BLAST ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=TargetedLociBlast](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=TargetedLociBlast)), a curated database, to find the genus information. To obtain more information of the species, the 28s rDNA sequence was also compared with the entire GenBank database ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)) to potentially find more hits with higher matching scores. Phylogenetic trees were generated in NCBI BLAST tree view based on the BLAST results. The 28S rDNA sequences of the representative organisms were retrieved from GenBank and a customized phylogenetic tree was generated using the Phylogeny.fr (271, 272) <http://www.phylogeny.fr/>.

**Table 6. Fungal 28S universal primers used in this research**

<b>Primer name</b>	<b>Sequence</b>
5.8SR	5'-TCGATGAAGAACGCAGCG-3'
LROR	5'-ACCCGCTGAACTTAAGC-3'
LR3R	5'-GTCTTGAAACACGGACC-3'
LR5	5'-TCCTGAGGGAAACTTCG-3'
LR16	5'-TTCCACCCAAACACTCG-3'
LR6	5'-CGCCAGTTCTGCTTACC-3'
LR7	5'-TACTACCACCAAGATCT-3'

### III.2.3 *Fermentation*

Secondary metabolites production from fungal strain F01804 was undertaken using solid state fermentation. Brown rice was used as the solid substrate, and a broth solution composed of yeast extract (0.2 % (w/v)), glucose (0.2 % (w/v)) and tryptonized casein (0.2 % (w/v)) was added into a 100 ml bottle prior to autoclaving at a ratio of 6 g brown rice and 20 mL of broth. Mycelial suspensions of the fungus were inoculated onto the rice medium and incubated at room temperature for 4 to 10 weeks.

### III.2.4 *Natural product extraction and purification*

Each bottle of the rice medium with F01804 growing was mixed with 15 ml of methanol, acetone or ethyl acetate and was left static for an overnight extraction. Since acetone gave the best extraction results with respect of the yield and initial purity, it was used for later batches of the extractions. The extracts were passed through a filter paper to remove the rice and insoluble materials in the solvents prior to evaporating the solvents in a centrifugal evaporator (Genevac Ltd.) or a rotary evaporator. The dried materials were redissolved in a smaller volume of dimethylformamide for preparative HPLC purification using (see below for details). The target compound in the HPLC fractions was tracked with both the analytical HPLC-mass spectrometry (LC-MS) signal (see below for details) and the yeast 20S proteasome bioactivity (see below for details). The fractions containing the bioactive metabolites were pooled and lyophilized.

### III.2.5 HPLC and mass spectrometry

For the preparative HPLC experiment, one milliliter of crude extract was injected into a Kromasil 100-5-C18 column  $150 \times 21.2$  mm and was separated on a PerSeptive Biosystem BioCad HPLC by a mobile phase gradient elution of water (A) and acetonitrile (B) both acidified with 0.1% formic acid at a flow rate of 20 ml/min. The gradient was performed starting from 5-10% B over the first 3 min, followed by 10-35% B from 3 to 15 min, and 35-100% from 15 to 18 min. 10 ml per fraction was collected using an ADVANTEC SF-3120 fraction collector.

For the analytical HPLC, 30  $\mu$ l of the sample was injected into a PrincetonSPHER-100 C18  $50 \times 4.6$  mm column and was separated on a Shimadzu Prominence HPLC System by a mobile phase gradient elution of water (A) and acetonitrile (B) both acidified with 0.1% formic acid at a flow rate of 1.2 ml/min. For initial analysis of the crude extraction, a steep gradient of 5% to 95% B over 5 min was performed (Method I). For further analysis, a slower gradient was used starting from 0% to 20% B over the first 2 min, and 20% to 50% B from 2 min to 8 min (Method II). For mass spectrometry analysis, the A Shimadzu LCMS-2010 EV with an electrospray ionization system was operated in positive mode.

### III.2.6 Yeast and human cell line

*Saccharomyces cerevisiae* strain MBY1202 was provided as a courtesy from Dr. Mary Bryk's lab. Human cell lines were purchased from ATCC, Manassas, Virginia. Chemical sources are indicated in the context.

### III.2.7 *Yeast 20S proteasome Cloning, Protein Purification and Crystallography*

The yeast 20S proteasome was chosen as the model to elucidate the structural basis of inhibition. The yeast strain MBY1327 expressing a protein-A (ProA) tag fused proteasome was cloned by a former lab member Michael Schneider. Briefly, a ProA tag was cloned to the C-terminus of the PUP1 ( $\beta$  2) subunit linked by a tobacco etch virus (TEV) protease cleavage site as described previously (273). The modified yeast strain was cultivated in YPDt media (1 % yeast extract, 2 % peptone, 2 % dextrose, 0.4 mM tryptophan) because the parental strain MBY1202 was tryptophan auxotroph.

The yeast 20S proteasome was purified with IgG resin (GE healthcare) and was polished with the size exclusion chromatography (SEC) HiPrep Sephacryl S-500 HR column (GE healthcare) as described previously (273). The purified recombinant 20S proteasome was concentrated for the biochemical assays (1 mg/ml stock) and the crystallographic studies (30 mg/ml). The crystallization conditions were slightly modified around the reported condition (243) 28 mM  $\text{Mg}(\text{CH}_3\text{COO})_2$ , 100 mM MES pH 7.2, and 10% 2-methyl-2,4-pentanediol (MPD). The proteasome and the crystallization buffer were mixed in a 1:1 or 2:1 ratio for the hanging drop vapor diffusion crystallization. Rod shaped crystals appeared after 2 days at 16 °C. The proteasome crystals were then soaked for 3 h with Mz721 dissolved in the crystallization condition to a final concentration of around 1 mM to obtain the protein-inhibitor complex. Crystals were picked up and soaked into the cryoprotectant (20 mM  $\text{Mg}(\text{CH}_3\text{COO})_2$ , 100 mM MES pH6.9, 30% MPD) before being flash frozen in the liquid nitrogen.

### III.2.8 *X-ray diffraction data collection and processing*

The crystals were shipped to the synchrotron beamlines 19-ID or 23-ID of the Advanced Photon Source (Argonne National Laboratory, Illinois). X-ray diffraction data were collected using the oscillation method at the wavelength of 0.97949 Å.

The 20S proteasome crystal was diffracted to an overall resolution of ~2.8Å. X-ray intensity evaluation and data reduction were performed with HKL2000 (155). The crystal unit cell was isomorphous to the reported yeast 20S proteasome structures (219, 243, 258). Therefore, an isomorphous replacement was performed by omitting the ligand from the template model, randomizing the b-factors of the protein atoms and performing the simulated annealing (163). The Mz721 compound geometry library file was generated by modifying the TMC-95A in JLigand (274) and in PHENIX.elbow (163) by providing a modified TMC-95A atomic coordinates. Mz721 was manually placed into the electron density in Coot. Structure refinement was performed in PHENIX.refine or Refmac5 and rounds of manual refinement were performed in Coot (157, 275, 276). Data collection and refinement statistics were listed in Table 2.

The structure was analyzed and the figures were generated using UCSF Chimera (166).

### III.2.9 *Yeast Proteasome Bioactivity Assay*

Bioactivity of Mz721 was tested with the purified recombinant yeast 20S proteasome. The substrates used for the three active sites were succinyl-Leu-Leu-Val-Tyr-AMC (Suc-LLVY-AMC) for CT-L activity, aceto-Arg-Leu-Arg-AMC (Ac-RLR-

AMC) for T-L activity, and carboxybenzyl- Leu-Leu-Glu-AMC (Z-LLE-AMC) for post-glutamyl peptidyl hydrolase activity (all substrate were purchased from Boston Biochem). The  $K_m$  values for the three substrates were determined in the buffer containing 50 mM Tris-HCl pH 7.4, 1 mM EDTA. Yeast 20S proteasome was used at a final concentration of 2 nM, and the substrates were serially diluted.  $K_m$  value for Suc-LLVY-AMC substrate was approximately 50  $\mu$ M, and for the other two substrates were around 100  $\mu$ M.

The inhibition assays were performed in 384-well microtiter plates (Greiner) at a total volume of 50  $\mu$ l/well containing 29  $\mu$ l assay buffer (50mM Tris-HCl pH 7.4, 1 mM EDTA), 10  $\mu$ l yeast 20S proteasome (2 nM final concentration), 10  $\mu$ L of the peptide mimic substrates (100  $\mu$ M final concentration), and 1  $\mu$ l DMSO or inhibitors (dissolved in 100% DMSO). Bortezomib was used as the control inhibitor if indicated in this assay at concentrations ranging from 100 nM to 0.78 nM by serial dilutions. DMSO was kept at 2 % throughout the assay. A time-course fluorescence intensity of the AMC cleavage was measured by exciting at 380 nm and recorded at 460 nm with a microtiter plate reader (Omega).

### III.2.10 *Cell Culture and Treatments*

Mz721 was tested on multiple tumor cell lines to assess the activity and specificity. Four tumor cell lines, MOLT-4 (leukemia), OVCAR-8 (ovarian cancer), Toledo (lymphoma), and SNU-398 (liver cancer), were selected for this batch of tests. The cells were cultivated at 37 °C in RPMI-1640 medium supplemented with 10% fetal

calf serum, 100 units/ml penicillin, and 100 µg/ml streptomycin. During treatment, the cells were grown to optimal drug-test density ( $9 \times 10^4$  cells/ml for MOLT4,  $2.5 \times 10^4$  cells/ml for OVCAR-8,  $6.4 \times 10^4$  cells for 2631, and  $3 \times 10^4$  cells/ml for SNU-398), and aliquoted into a 384-well microtiter plate in triplicates. Mz721 was added in 2-fold serial dilutions ranging from 20 µM to 2.4 nM. The cells were incubated with the compound at 37 °C. Resazurin, a dye measuring aerobic respiration and commonly used as an indicator of cell viability, was applied 2 days after the compound treatment. The plate was read on a microtiter plate reader 24 h after the addition of the dye. The intensity of the fluorescence at 590 nm is proportional to the number of live cells. IC<sub>50</sub> values were calculated accordingly. Doxorubicin, a DNA intercalating agent and an anti-cancer drug, was set in 2-fold serial dilutions from 10 µM to 0.3125 µM as a control compound throughout the experiment. Mz721 was also tested upon human dermal fibroblasts (HDF) to assess the cytotoxicity against normal cells.

### **III.3 Results**

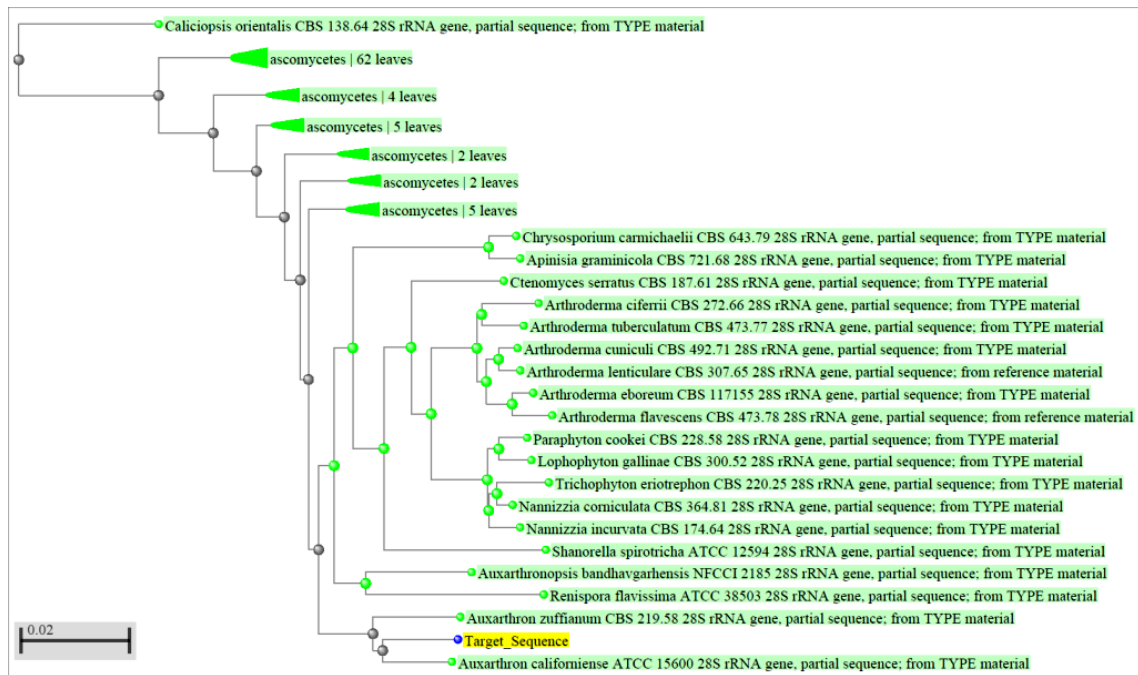
#### *III.3.1 Identification and taxonomy study of the fungal strain F01804*

Mz721 was a hit originated from a high throughput capillary electrophoresis screening against a library containing 105,500 crude extract samples. The crude extract was then tracked back to a fungal strain with the code name F01804. The genomic DNA of F01804 was extracted and its 28s ribosomal DNA (rDNA) was amplified and sequenced. A sequence alignment against the curated database of fungal 28s rDNA at the National Center for Biotechnology Information (NCBI) suggested that it was closest

to the genus of *Auxarthron* (Fig. 43). The hit with the highest score (*A. californiense*) showed a sequence identity of approximately 97%. In order to potentially obtain more taxonomic information, the sequence was tentatively searched against the GenBank at NCBI which includes many more entries although about 20% of the fungal sequences in GenBank were potentially misannotated (277). The query sequence was 100% covered and 100% matched to *A. umbrinum* (Fig. 44), in agreement in the genus annotation with the preceding search against the curated database. Therefore, we are more confident to conclude that the fungal strain F01804 belongs to *Auxarthron spp.*

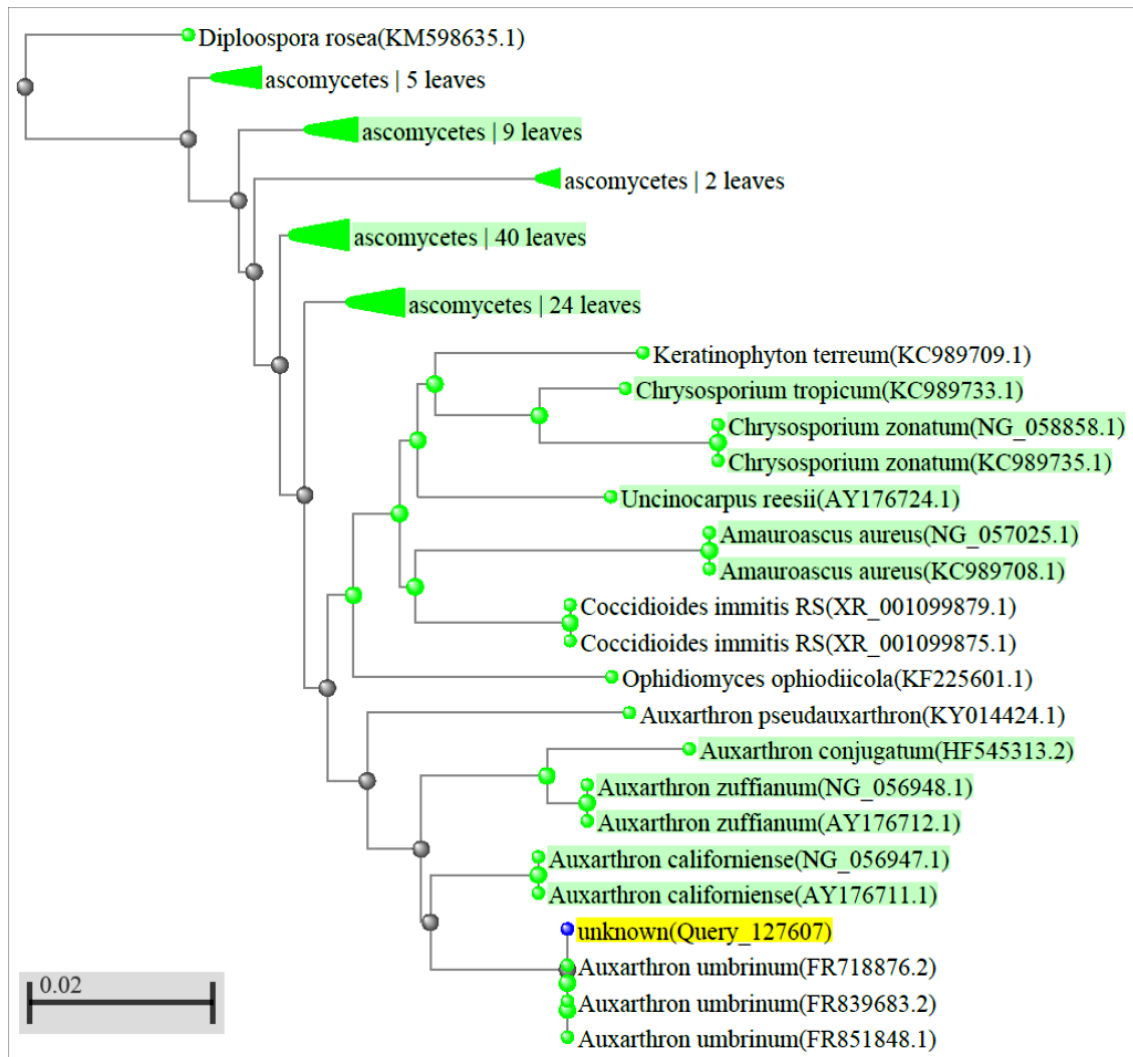
According to the previous owner of the fungus and metabolites assets, the novel proteasome inhibitors produced by F01804 were presumably analogs of the TMC-95 series. Recall that the TMC-95 series was produced by *Apiospora montagnei* (256). Although causing confusion, the same fungi species could have multiple names describing their sexual morph and asexual morph. This also applies to *A. montagnei*, whose asexual morph is *Arthrinium arundinis* according to (278) and the NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>). We compared the 28s rDNA sequences between *A. arundinis* and our target fungal strain *Auxarthron spp.* F01804. The identity between the partial 28s rDNA sequences was only 78 %. As references, some other fungal representatives from the previous searches were included in Fig. 45. Clearly, the genetic relationship between F01804 and *A. arundinis* was farther than that between F01804 and the farthest hit in the preceding searches in Fig. 43 and Fig. 44. The results indicated that the biosynthetic pathway of TMC-95 analogs may exist in many genera.



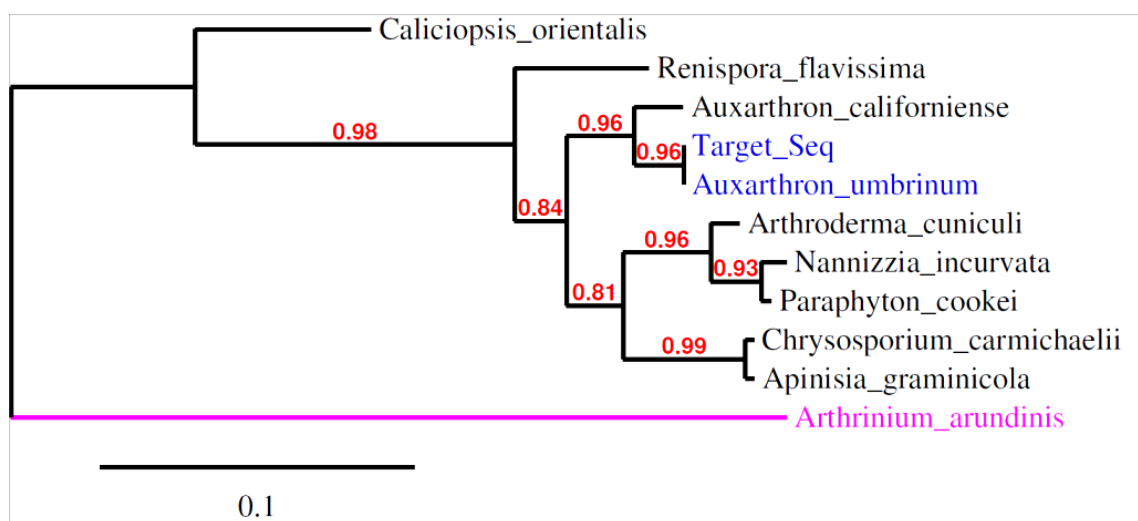


**Figure 43. Taxonomy study of fungal strain F01804 (Part 1)**

**A phylogenetic tree of the hits from a search against a curated database of the targeted loci of the fungal 28s rDNA in NCBI. Some farther related genera were collapsed. The target fungus was highlighted in yellow. Green shades represent the sequences from type material (type material is the taxonomic device that ties formal names to the physical specimens that serve as exemplars for the species) (279).**



**Figure 44. Taxonomy study of fungal strain F01804 (Part 2)**  
**A phylogenetic tree of the hits from a search against an uncurated database of the GenBank in NCBI. Some farther related genera were collapsed. The target fungus was highlighted in yellow. Green shades represent the sequences from type material.**

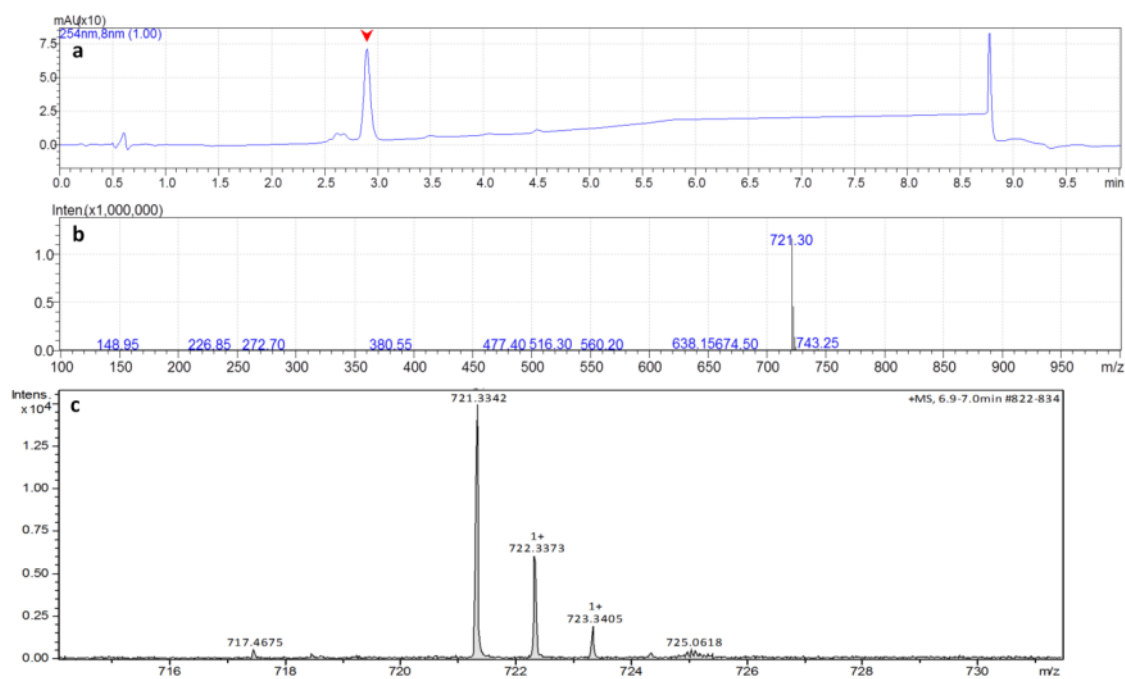


**Figure 45. Taxonomy study of fungal strain F01804 (Part 3)**  
**A phylogenetic tree of the target sequence (blue) and some representative sequences including *A. arundinis* (*A. montagnei*, magenta), the fungal strain that produced TMC-95 series of compounds.**

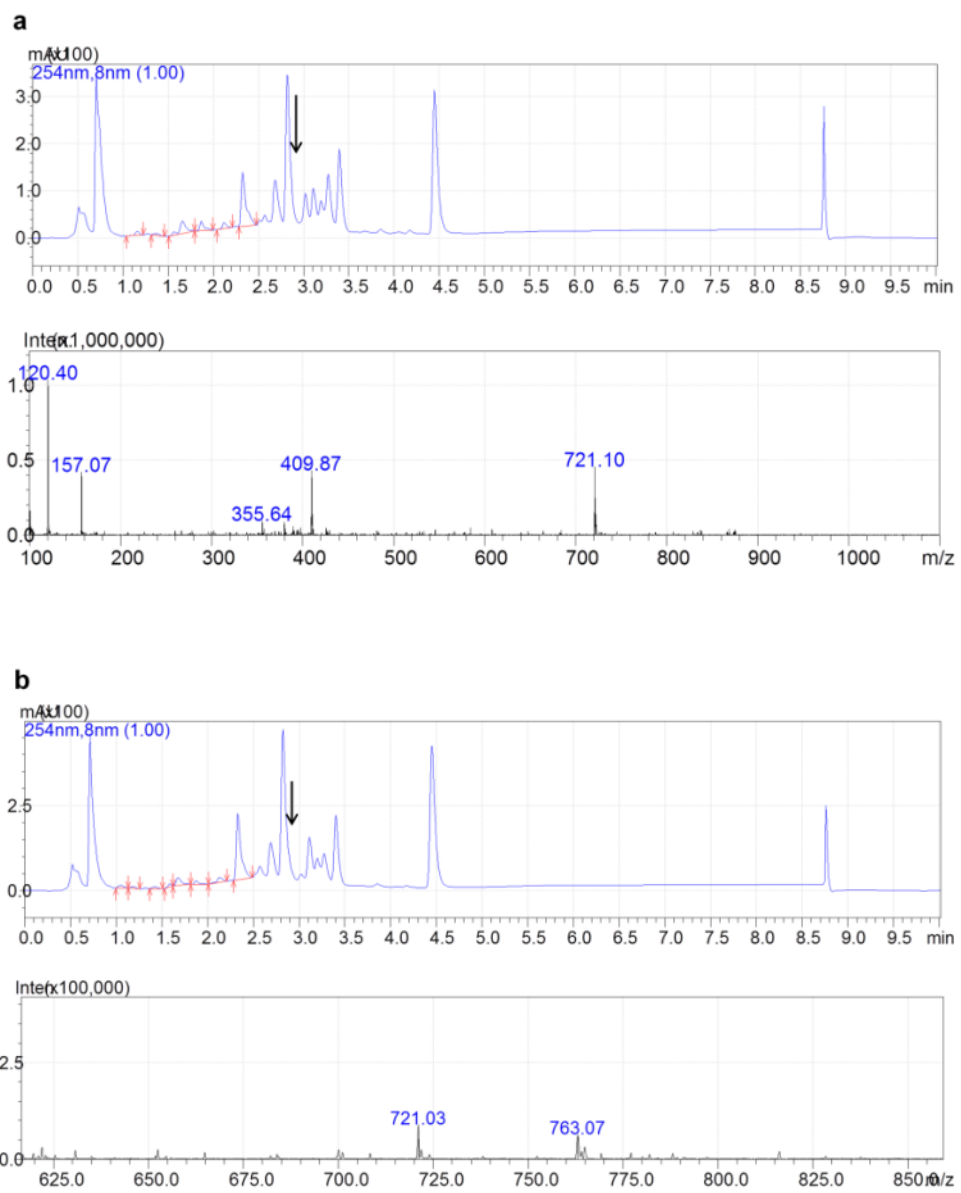
### III.3.2 Fermentation and extraction of the novel secondary metabolites

After identifying the fungal strain F01804 as *A. umbrinum*, we sought to confirm that it could produce the desired secondary metabolites which inhibit the 20S proteasome. Upon receiving the materials from Ohr Pharmaceuticals, we were informed that a particular metabolite of interest had a  $[M+H]^+$   $m/z$  value of 721 (designated Mz721), and a vial containing the purified fraction of Mz721 was included in the assets. We assessed the quality of Mz721. The purity was around 90 % assuming the major UV absorption peak was Mz721 (Fig. 46a), which was very probable because the peak showed a clean  $[M+H]^+$  molecular ion peak of 721.30 on the mass spectrometry (Fig. 46b). High resolution mass spectrometry showed that the accurate  $m/z$  value to the second decimal was 721.33 (Fig. 46c), and a clear isotopic pattern was observed. The

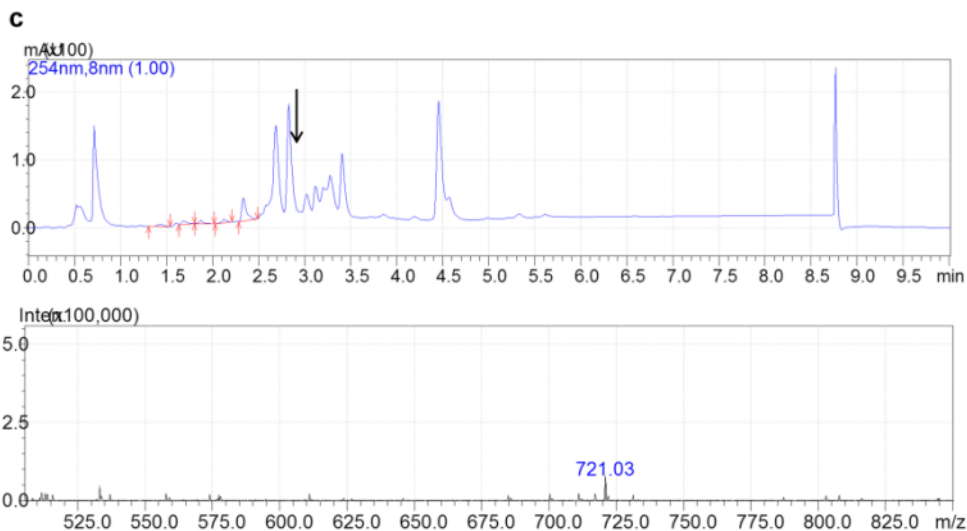
group of peaks centered at 725.0618 came from the residual contamination on the column, which was also observed in blank runs before sample injection (data not shown). Therefore, we used the purified fraction as the standard to track our fermentation extract using a combination of HPLC-MS and the bioactivity against the CT-L site of the yeast 20S proteasome. The production of the metabolites was assessed at different time points from four to ten weeks post inoculation. The production of Mz721 started to appear after about five weeks and accumulated. We initially used fungus grown for five weeks to evaluate three different extraction solvents, including methanol, acetone and ethyl acetate. While the total extraction efficiency of the Mz721 was similar, acetone seemed to extract less unwanted metabolites (Fig. 47). Therefore, we used acetone for a larger scale extraction of F01804 grown for ten weeks.



**Figure 46. HPLC-MS of the Mz721 standard fraction**  
**a, The UV chromatogram of Mz721 (280). The major peak with the retention time of 2.9 min (indicated by red arrowhead) corresponds to the Mz721. b, Mass spectrometry analysis of the center of the UV peak showed a clean molecular ion peak with the  $m/z$  value of 721. c, The high resolution mass spectrometry of Mz721 showed the accurate  $m/z$  was equal to 721.33.**



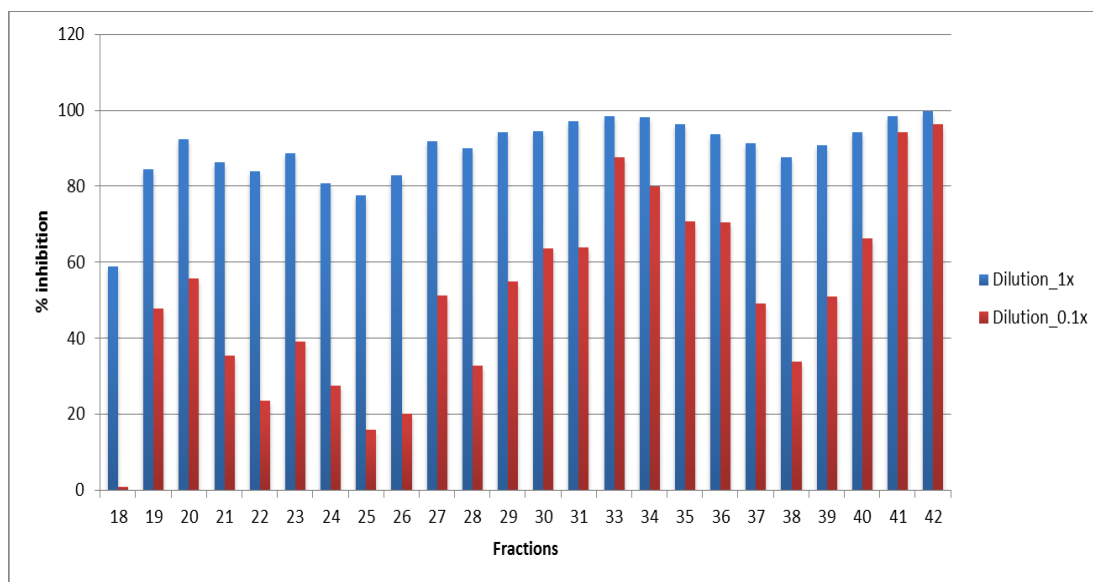
**Figure 47. HPLC-MS of extraction with different solvents**  
**a, Methanol; b, Acetone; c, Ethyl acetate. LC elution at 2.9 min was subjected to MS. (The highest peak at 2.8 min was not Mz721, but a mixture of  $[M/z]^+$  355 and 409)**



**Figure 47. Continued.**

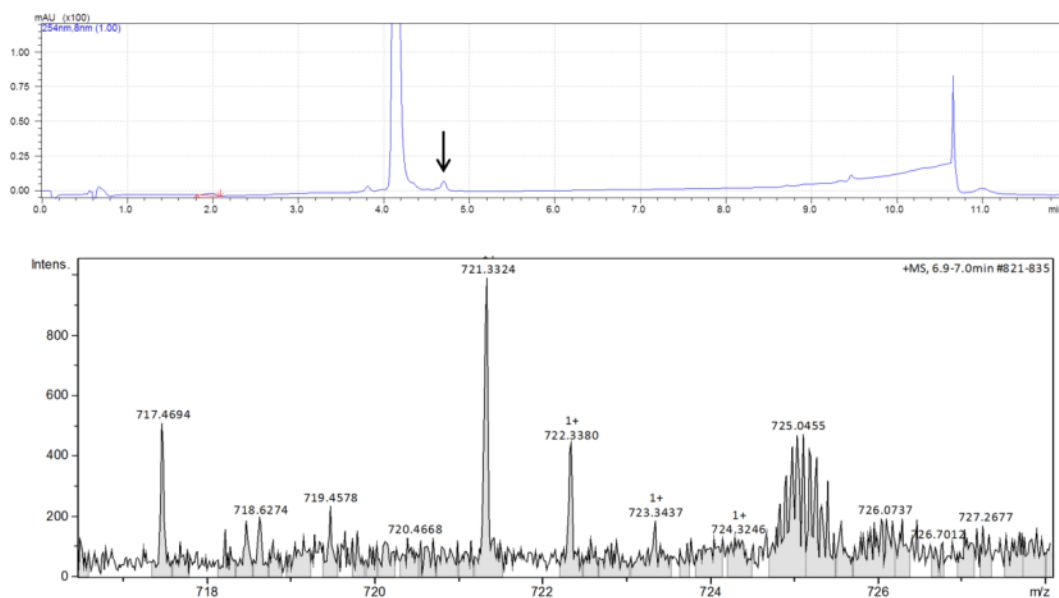
Since the total yield of Mz721 was still very low comparing to other extracted metabolites (much less than 1%), and was too dilute, the HPLC fractions had to be dried and redissolved in smaller volumes of DMSO for enzyme activity test. The bioactivity profile of the fractions indicated that there seemed to be more than one proteasome inhibitor (Fig. 48). The fractionation ended at Fraction 42, and the later elution was unfortunately not collected. High resolution mass spectrometry showed that the activity around Fraction 33 was likely because of Mz721 (Fig. 49), and Fraction 42 appears to contain another active compound probably with an  $[M/z]^+$  of 705 or 679 (Fig. 50). The mass spectrometry signal for Mz705 was stronger and more consistent than Mz679. We tentatively named it Mz705-mix for easier clarification throughout this research.

Nevertheless, we were not certain that the active ingredient was the Mz705, given that  $[M/z]^+$  of 679 happened to match TMC-95A.



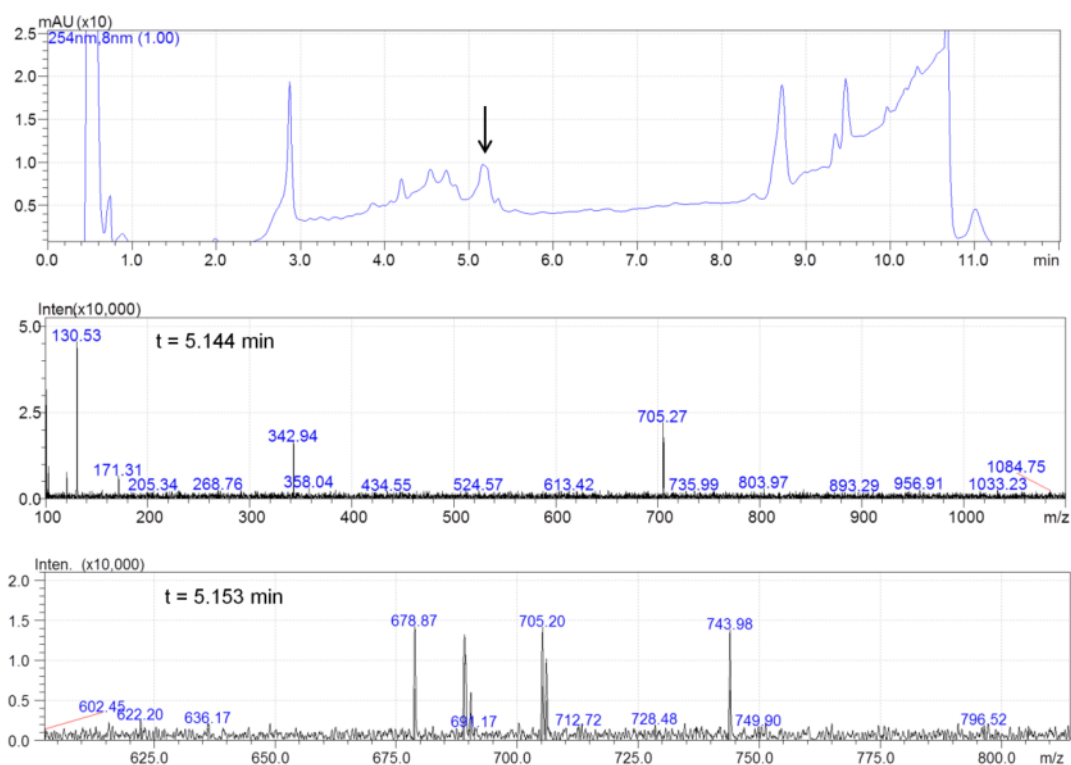
**Figure 48. Inhibition profile of the preparative HPLC fractions**  
The 1X dilution was too concentrated and all showed strong inhibition. 0.1X dilution suggested there were multiple proteasome inhibitory compounds which were not well-separated on the HPLC column.





**Figure 49. LC-MS of Fraction 33 in Fig. 48**  
**The analytical HPLC of the Fraction 33 (280), and the high resolution mass spectrometry of the peak (281).**

Due to the low fermentation scale and the low abundance of the metabolites, we could not further enrich Mz721 or Mz705-mix. Therefore, we conducted the following crystallography and biochemical assays using either the standard sample obtained from Ohr Pharmaceuticals or the mixture after the crude purification.



**Figure 50. LC-MS of Fraction 42 in Fig. 48**  
**Analytical HPLC (280) suggested the fraction still contains multiple components. The peak indicated by the arrow was subjected to MS. Mz705 was the most consistent peak in MS based on its detection in multiple time frames (represented by two here).**

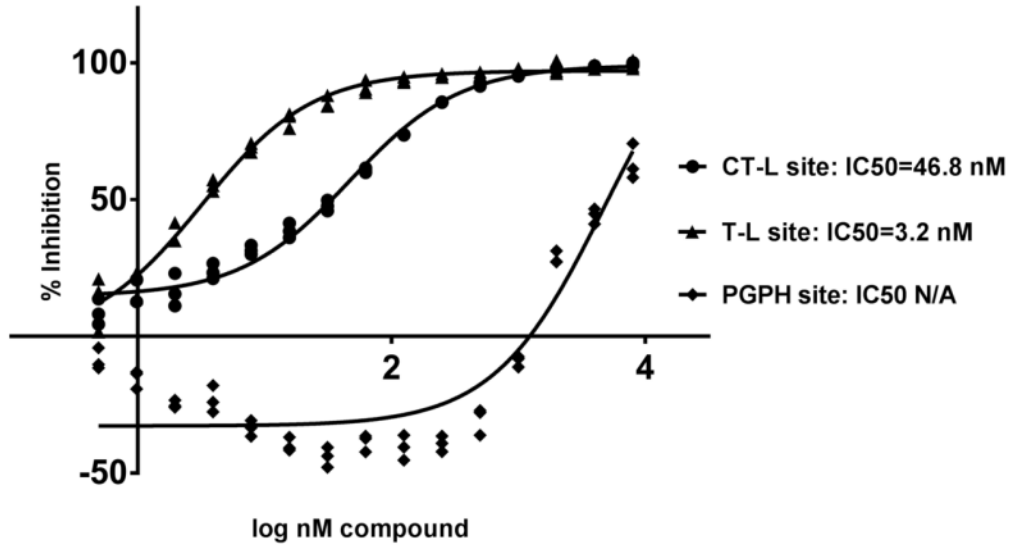
### III.3.3 *Mz721 preferentially inhibits the Trypsin-like activity of the yeast 20S proteasome*

Both Mz721 standard and Mz705-mix were tested against the three types of the catalytic sites of the yeast 20S proteasome. While most other known proteasome inhibitors, including TMC-95series, preferentially hit the CT-L activity, Mz721 showed more than ten-fold higher potency against the T-L activity than the CT-L activity (Fig. 51a). In particular, the IC<sub>50</sub> values for the T-L and CT-L activities were 3.2 and 46.8 nM, respectively. It had been reported previously that the addition of 0.03% SDS can activate the proteasome and also increase the apparent potency of an inhibitor. Therefore, the effect of SDS on the potency of Mz721 against CT-L activity was assessed, and the IC<sub>50</sub> decreased to 32.9 nM in the presence of SDS (Fig. 51b). The IC<sub>50</sub> values of Mz721 and the TMC-95 series were listed in Table 7.

The peak fraction containing Mz705-mix had an inhibition profile similar to the known inhibitors, and showed approximately three-fold higher potency for the CT-L sites than for the T-L sites (Fig. 52).

Both Mz721 and Mz705-mix showed weakest potency against the PGPH site. Interestingly, both compounds activated the PGPH activity by around 50 % at low concentrations. This was not reported in the TMC-95 series of compounds, we do not know if similar effect was also observed.

**a Mz721\_Yeast 20S Proteasome**



**b Mz721\_SDSEffect**

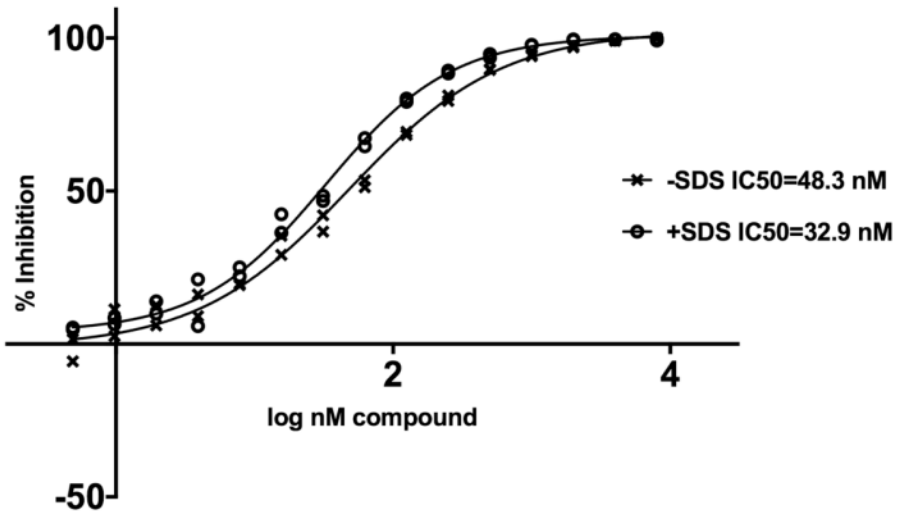


Figure 51. Yeast 20S proteasome inhibition test of Mz721

a. IC<sub>50</sub> on the three active sites of the yeast 20S proteasome. b. The effect of SDS on the CT-L site

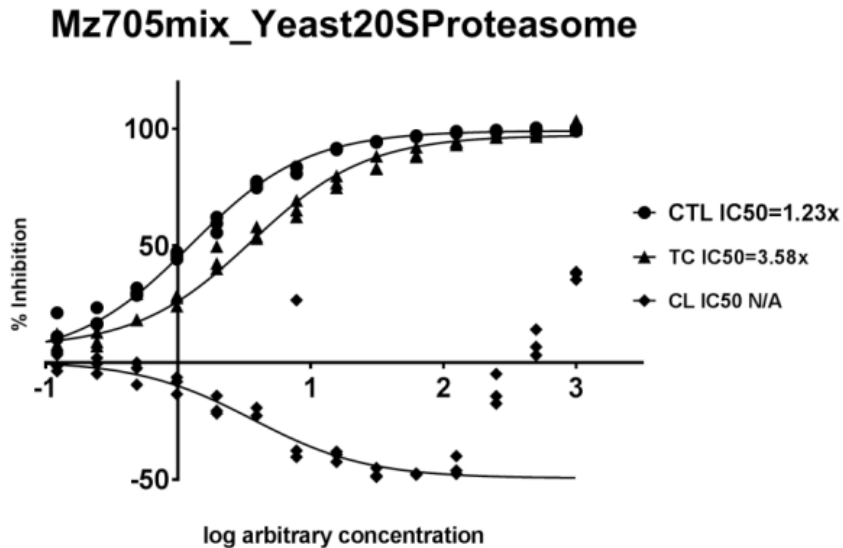
**Table 7. Inhibitory activities of Mz721 and the TMC-95 series against CT-L, T-L and PGPH active sites of the 20S proteasome.**  
**(The data for TMC-95 series were reported in (256).)**

Compounds	SDS	IC50 (nM)		
		CT-L	T-L	PGPH
Mz721	-	46.8	3.2	~ 4,800 <sup>A</sup>
	+	32.9	N/A	N/A
TMC-95A	-	12 <sup>B</sup>	1,500 <sup>B</sup>	6,700 <sup>B</sup>
	+	5.4 <sup>B</sup>	200 <sup>B</sup>	60 <sup>B</sup>
TMC-95B	+	8.7 <sup>B</sup>	490 <sup>B</sup>	60 <sup>B</sup>
TMC-95C	+	360 <sup>B</sup>	14,000 <sup>B</sup>	8,700 <sup>B</sup>
TMC-95D	+	270 <sup>B</sup>	9,300 <sup>B</sup>	3,300 <sup>B</sup>

N/A: the activity was not assessed.

A : the IC50 value was inaccurate because the data points could not be fit into a sigmoid curve.

B: tested on human 20S proteasome



**Figure 52. Yeast 20S proteasome inhibition profile of Mz705-mix**

### III.3.4 Mz721 binds to the active sites of the 20S proteasome

We employed X-ray crystallography to understand the high potency of Mz721 against the T-L activity. The yeast 20S proteasome crystal was soaked with Mz721 to obtain the complex crystal. The crystal diffracted X-ray to an overall resolution of approximately 2.8Å. The crystal belonged to P2<sub>1</sub> space group, and the unit cell was isomorphous to the available structures of yeast 20S proteasome with a = 135.92, b = 299.99, c = 143.63,  $\beta = 112.99^\circ$  (data collection and refinement statistics are listed in Table 8). The electron density was well-defined for the protein and the compound. Mz721 bound to all three active sites similar to the structurally elucidated TMC-95A (258). The electron density of the catalytic Thr1 O $\gamma$  of all three active sites was clearly disconnected from the electron density of the inhibitors, indicating Mz721 was a non-covalent inhibitor.

**Table 8. Crystallography data collection and refinement statistics**

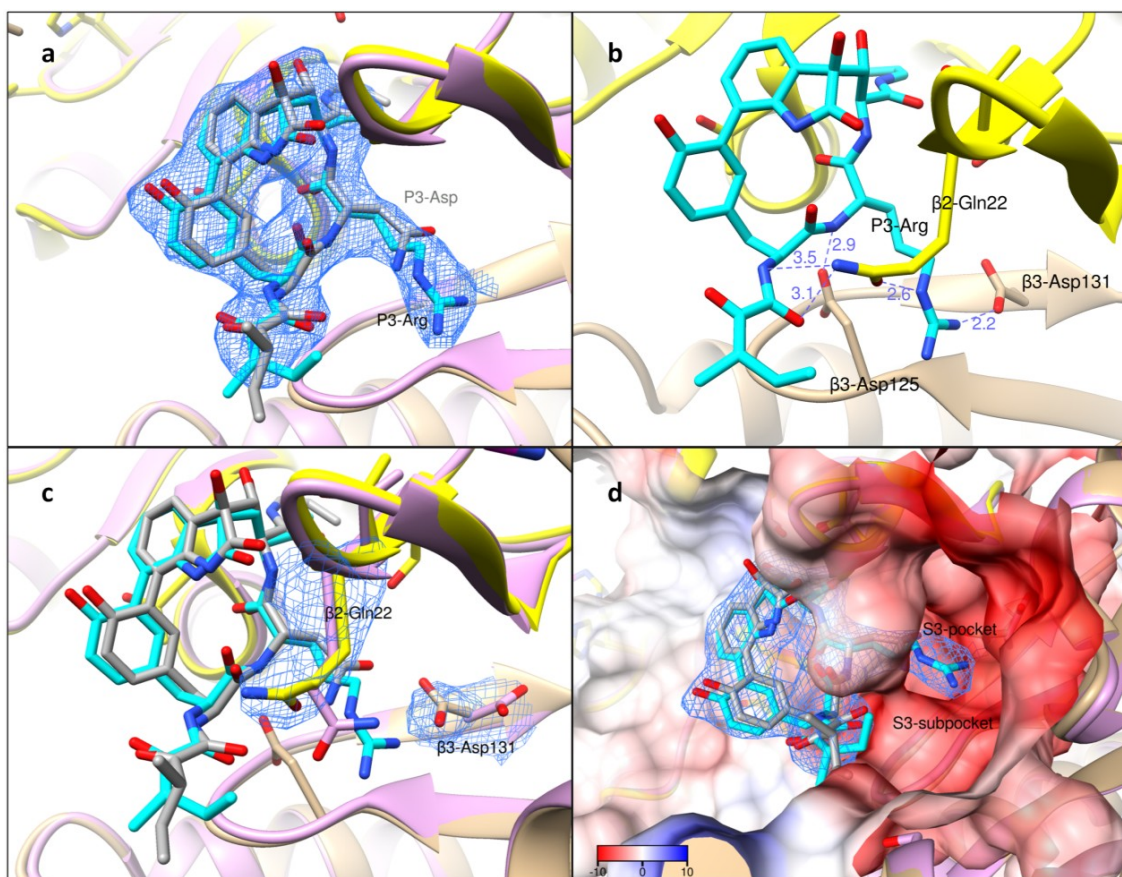
<b>Data Collection</b>	<b>Yeast 20S proteasome-Mz721</b>
Space Group	P2 <sub>1</sub>
Unit cell dimensions	
a, b, c (Å)	a = 135.92, b = 299.99, c = 143.63
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 112.99, 90.00
Wavelength (Å)	0.97949
Resolution range (Å) *	50.00-2.80 (2.85-2.80)
Unique reflections *	259,599 (13,021)
Multiplicity *	4.3 (4.3)
Completeness (%) *	99.8 (100.0)
R <sub>merge</sub> *	0.098 (0.634)
R <sub>pim</sub> *	0.054 (0.347)
I/ $\sigma$ (I) *	19.68 (2.76)
Wilson B value (Å <sup>2</sup> )	56.59
<b>Refinement</b>	
Resolution (Å) *	34.45-2.79 (2.86-2.79)
R <sub>work</sub> /R <sub>free</sub> *	0.175/0.232 (0.29/0.38)
No. of atoms	49,902
Protein	49,586
Ligand	313
Solvent	3
Mean B factors (Å <sup>2</sup> )	72.69
Protein	72.77
Ligand	60.71
Solvent	55.55
Rmsd bond length (Å)	0.009
Rmsd bond angle (°)	1.40
Clashscore	12.7
Number of TLS groups	1
Ramachandran plot (%) (favored/allowed/outliers)	93.31/5.66/1.03
Rotamer outliers (%)	0.11

\* Values in parentheses indicate highest-resolution bin.

When we tentatively built TMC-95A into the electron density of Mz721 at all three active sites, we noticed a major difference at the P3 position between the two inhibitors (Fig. 53a). Mz721 possesses a longer P3 side chain. Based on the 42 Da mass difference between TMC-95A ( $[M/z]^+ = 679$ , exact mass 678.26 Da) and Mz721, we suggest that the P3 site of Mz721 is an arginine substituting an asparagine in TMC-95A, and the arginine side chain could perfectly fit into the electron density. In addition, the calculated exact mass (721.33) also precisely matches the high-resolution mass spectrometry data (Fig. 46c, 54).

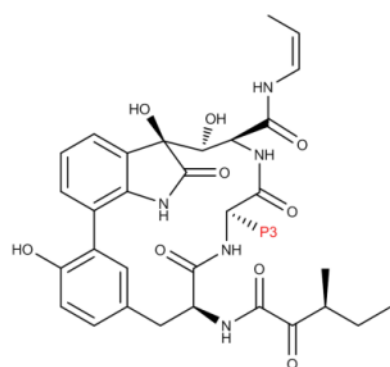
Although peptide inhibitors commonly contain two to four residues to occupy the S1 to S4 substrate binding pockets, only the S1 and S3 pockets are practically used for inhibitor binding. This is because the S2 and S4 pockets face the lumen of the proteasome and do not form deep and well-structured pockets. The S1 pockets of all three active sites are solely formed by the catalytic subunits, while the S3 pockets are cooperatively formed by the catalytic subunits and the neighboring subunits. For example, the S1 pocket of the T-L site is formed by  $\beta 2$  subunit, and the S3 pocket is formed by  $\beta 2$  and  $\beta 3$  subunits. Notably, most of the known proteasome inhibitors, including the covalent ones bortezomib and epoxomicin and the non-covalent TMC-95A, barely touch the  $\beta 3$  portion of the S3 sub-pocket (Fig. 53c) (243, 252, 258).



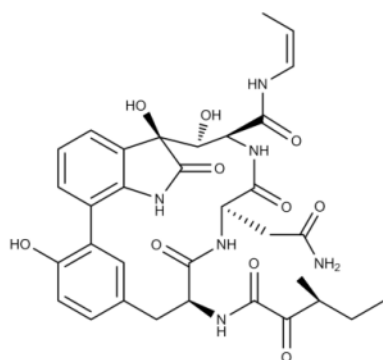


**Figure 53. T-L site bound by Mz721**

**a**, The electron density of the ligand showed that there was a major difference at the P3 position between Mz721 and TMC-95A. **b**, The longer P3-Arg side chain interacted with  $\beta 2$ -Gln22 and  $\beta 3$ -Asp131. **c**, Comparison with the TMC-95A bound T-L site clearly showed that the P3-Arg induced the rotamer conformational change of  $\beta 2$ -Gln22 and  $\beta 3$ -Asp131. **d**, P3-Arg inserted deep into the S3 pocket and contacted with the S3 sub-pocket.



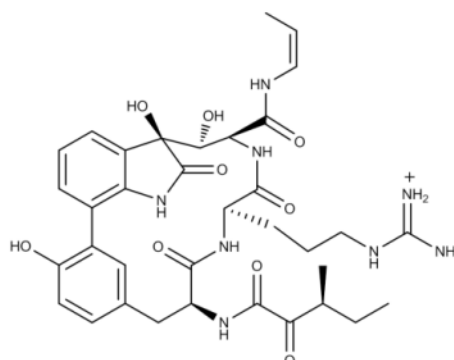
Compound	P3
TMC-95A	Asn
Mz721	Arg



**TMC-95A**

Chemical Formula: C<sub>33</sub>H<sub>38</sub>N<sub>6</sub>O<sub>10</sub>

Exact Mass: 678.26



**Mz721**

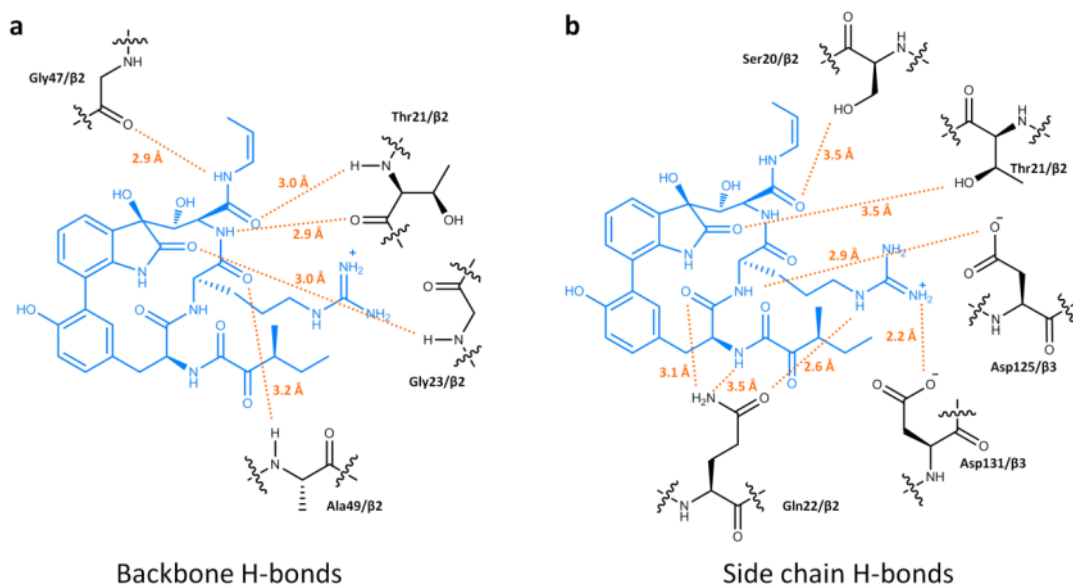
Chemical Formula: C<sub>35</sub>H<sub>45</sub>N<sub>8</sub>O<sub>9</sub><sup>+</sup>

Exact Mass: 721.33

**Figure 54. The comparison of TMC-95A with Mz721**

When scrutinizing the Mz721 binding mode, we found that the P3-arginine inserted deep into the S3 sub-pocket and interacted strongly with the acidic residues (Fig. 53c, d). Furthermore, the P3-arginine induced side chain conformational change of the  $\beta 3$  subunit (Fig. 53b, c). The  $\beta 3$ -Asp131 in other structures (243, 252, 258) faced away from the inhibitors and could not form hydrogen bond with the P3 side chain, while in the Mz721-bound structure,  $\beta 3$ -Asp131 rotated nearly 180 ° to interact with the

P3-arginine, and a very strong hydrogen bond with a bond length of around 2.2 Å was formed. In addition, the P3-arginine also forced β2-Gln22 to adopt another side-chain conformation to prevent steric clashes. The new rotamer resulted in the formation of an additional set of hydrogen bonds between β2-Gln22 and P3-arginine (2.6 Å), and between β2-Gln22 and the backbone ketoamide moiety (3.1 and 3.5 Å, respectively) of Mz721 (Fig. 53b). In the TMC-95A bound structure, the ketoamide moiety was proposed to have little contribution to the protein-inhibitor interactions (258). Since the other interactions between Mz721 and the proteasome T-L site were very similar to those between TMC-95A and the proteasome, we conclude that the unique interactions contributed by the P3-arginine gave rise to the high potency of Mz721 against the T-L activity. The interactions between the T-L site amino acid residues and Mz721 is illustrated in Fig. 55.



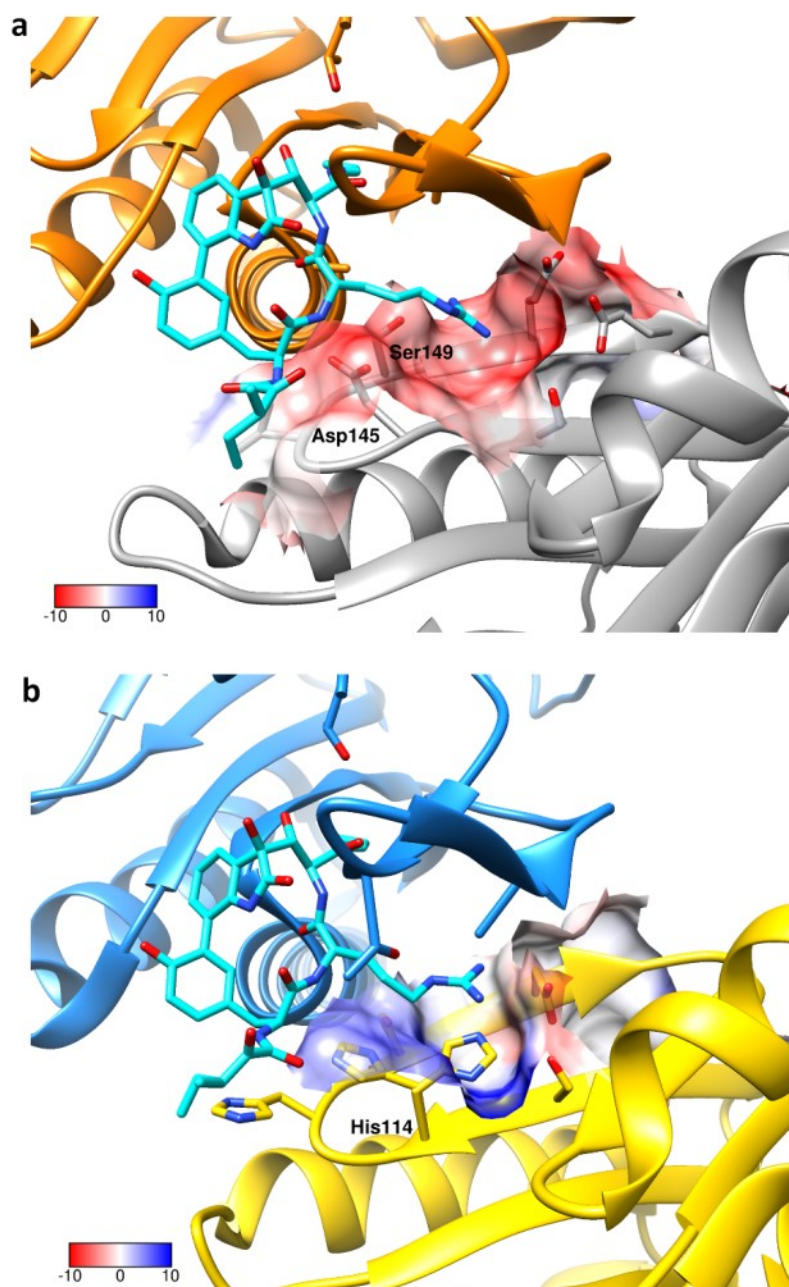
**Figure 55. Mz721 forms a host of hydrogen bonds with the proteasome T-L active site**

**The hydrogen bonds include those formed between the inhibitor and the protein main-chain atoms (a), and those formed between the inhibitor and the side-chain groups of the protein (b).**

In addition to the T-L activity, Mz721 also strongly inhibited the CT-L activity. Similar to that of the T-L site, CT-L site ligand binding pocket was also composed by two subunits, the  $\beta 5$  bearing the catalytic Thr1 and  $\beta 6$  without catalytic role. The P3-arginine did not form hydrogen bonds with the protein. Although the S3 sub-pocket of CT-L site was also very acidic, the distances between the protein residues and the P3-arginine were all above 4 Å, indicating the interactions were likely electrostatic. In the TMC-95A-bound structure (258), the P3-Asn interacted with Asp145 and Ser149 of  $\beta 6$  subunit with hydrogen bonds (Fig. 56a). However, in the Mz721-bound structure, the P3-Arg could not form these two hydrogen bonds. This potentially explains why

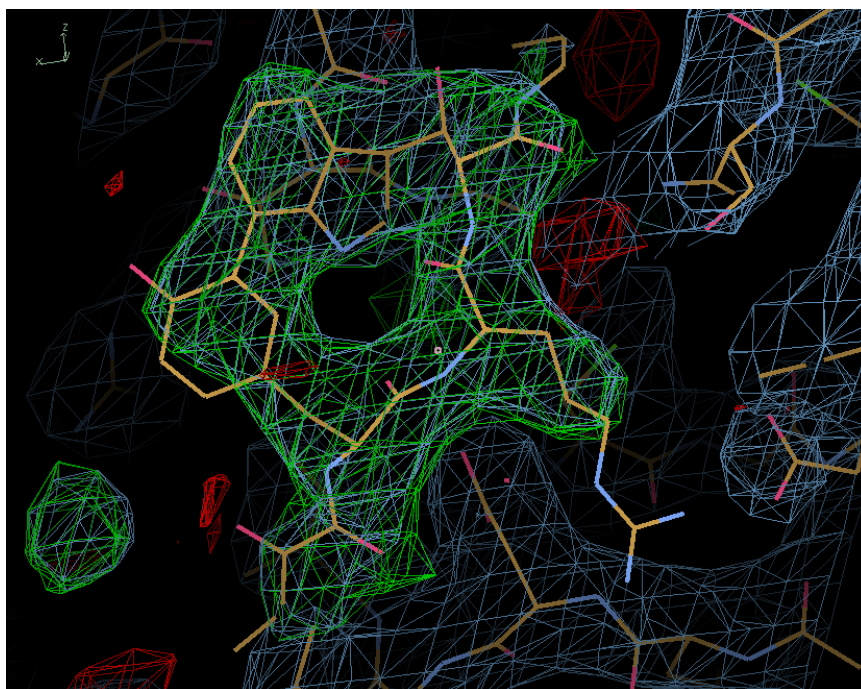
substitution of the P3 position of Mz721 decreased the potency of the inhibitor against CT-L activity by nearly one order of magnitude compared to TMC-95A.

The PGPH activity was inhibited by Mz721 with an  $IC_{50}$  value in the low micromolar range without SDS supplementation into the assay buffer. This was in the same concentration range as the TMC-95A. In both TMC-95A-bound and Mz721-bound proteasome structures, only the protein backbone of the PGPH site formed hydrogen bonds with the inhibitors. Because major interactions between TMC-95A and the protein are conserved among all three active sites, the weak inhibition of TMC-95A against PGPH site was previously ascribed to the P3 asparagine side chain which might not be optimal for PGPH site binding pocket (258). We noticed that the P3 arginine of Mz721 is also not optimal. P3-arginine induced a partial conformational change of  $\beta$ 2-His114. Around 50% of the  $\beta$ 2-His114 in the whole protein population adopted an alternative conformation (Fig. 56b) as analyzed by the electron density map (data not shown). Notably, the S3 sub-pocket of PGPH site is more basic than those of the CT-L and T-L sites. This was largely because of the His114 residue, and it can cause some level of electrostatic repulsion between the positively charged guanidino group of P3-arginine and the positively charged imidazole group of His114 (Fig.56b).



**Figure 56. Binding mode of Mz721 in CT-L and PGPH sites**  
**a, The Mz721 binding environment in CT-L site. The  $\beta 5$  subunit bearing the CT-L activity is colored in orange. The  $\beta 6$  subunit contributing the S3 sub-pocket is colored in grey. Electrostatic surface for the S3 sub-pocket is shown. b, The Mz721 binding environment in PGPH site. The  $\beta 1$  subunit bearing the PGPH activity is colored in blue, and the  $\beta 2$  subunit which both contributes the S3 sub-pocket to the PGPH site and bares the T-L activity is colored in yellow.**

We also soaked the crude Mz705-mix into the yeast 20S proteasome crystals and solved the structure. Electron density showed that Mz705 mix has a shorter P3 side chain when we placed Mz721 into the electron density map (Fig. 57). Combining its inhibition profile, it may possess an asparagine at the P3 position. More work is required to identify this potent compound.



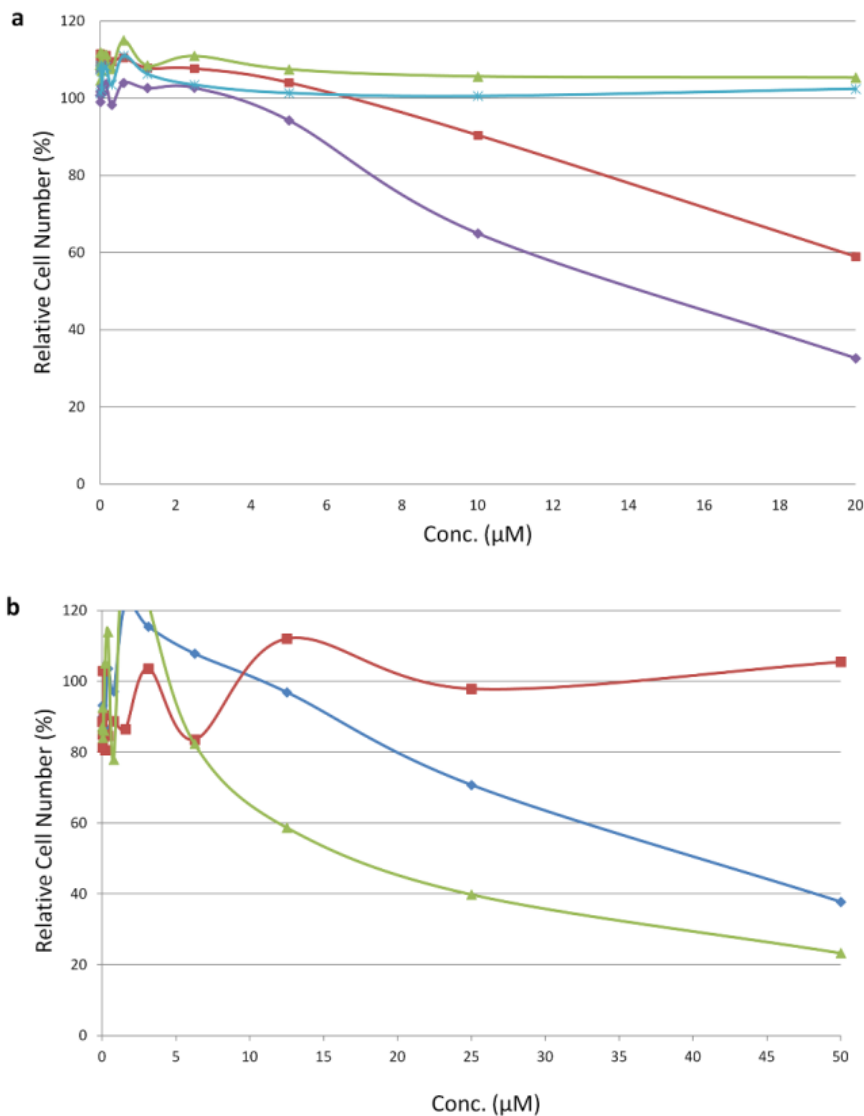
**Figure 57. Electron density of Mz705-mix overlaid with Mz721 coordinates at the T-L site.**

**The 2Fo-Fc (blue) and Fo-Fc (green) electron density maps of the Mz705-mix in complex with yeast 20S proteasome were shown in 1.5  $\sigma$  and 3.0  $\sigma$ , respectively. The coordinates of Mz721 (yellow and by elements) was overlaid with the electron density of Mz705-mix. Mz721 clearly possessed a longer P3 side chain.**

### III.3.5 *Mz721 showed cytotoxicity on several cancer cell lines*

Mz721 was tested on four different tumor cell lines to assess its whole cell activity. Growth of SNU-398, a liver cancer cell line, and MOLT-4, a leukemia cell line, was inhibited by Mz721 with  $IC_{50}$  values of around 15 and 22  $\mu\text{M}$ , respectively (Fig. 58a). While the growth of two other cell lines Toledo (also named CRL-2631), a lymphoma cell line, and OVCAR-8, an ovarian carcinoma cell line, were not affected by Mz721 under tested concentration range. These results suggest that certain cell lines are more susceptible to Mz721. The cytotoxicity of Mz721 against normal cells was tested on human dermal fibroblast. Mz721 was not toxic even treated at 50  $\mu\text{M}$  (Fig. 58b). TMC-95A, however, started to show toxicity above 10  $\mu\text{M}$ , and an  $IC_{50}$  of around 40  $\mu\text{M}$ . Although Mz721 showed less cytotoxicity in normal cells, its potency on the cancer cells, especially the drug-resistant cell lines, was not improved. We ascribe this mainly as the amino acid sequence differences of the  $\beta 2$  subunit between the human proteasome and the yeast proteasome. Although Mz721 showed high potency against the T-L site of the yeast proteasome, the amino acids  $\beta 2$ -Gln22 and  $\beta 3$ -Asp131 which formed unique interactions with the P3-arginine are not conserved in the human proteasome (Fig. 59). The amino acids at equivalent positions in human proteasome are glutamate and methionine, respectively. The latter may even clash with the P3-arginine. Another possibility for the reduced potency is that the positively charged guanidino group on the P3-arginine side chain may impair the compound uptake by the cells, resulting in low intracellular concentrations of the drug, and therefore lower cytotoxicity for both normal and cancer cells.





**Figure 58. Effects of Mz721 on tumor cell lines and human dermal fibroblasts**  
**a**, MOLT-4 (leukemia, red square), OVCAR-8 (ovarian cancer, green triangle), SNU-398 (liver cancer, purple diamond), Toledo (lymphoma, blue asterisk) were tested for Mz721 dose-response assay. MOLT-4 and SNU-398 showed response to Mz721, with the apparent  $\text{IC}_{50}$  values of approximately 25 and 15  $\mu\text{M}$  respectively.  
**b**, Mz721 (red square), TMC-95A (blue diamond), bortezomib (green triangle) were tested against the HDF. The bortezomib curve was expanded five times along the concentration axis. Therefore, the concentrations of the data points for bortezomib are virtually one-fifth of the indicated concentrations. Bortezomib and TMC-95A are cytotoxic against the HDF with apparent  $\text{IC}_{50}$  values of approximately 3 and 40  $\mu\text{M}$  respectively.



### III.4 Discussion

In this study we report the discovery of a non-covalent proteasome inhibitor Mz721 that is naturally produced by the fungus strain *A. umbrinum* F01804. Moreover, the fungus seems to produce a series of the analogs that all inhibit the 20S proteasome with different selectivity. At least two compounds were observed or semi-purified in our studies. Mz721 showed superb potency against the T-L activity when tested on the purified yeast 20S proteasome, while Mz705-mix preferentially inhibits the CT-L activity. The crystal structure of Mz721 in complex with the yeast 20S proteasome revealed that the inhibitor interacts with the active sites via a host of hydrogen bonds. Similar to TMC-95A, most of these hydrogen bonds are formed between the inhibitor and the main-chain atoms of the protein. Given that the overall scaffold of the proteasome from different organisms are more or less identical, these cyclic peptides may potentially inhibit proteasome regardless of the origin. A unique feature of Mz721 is that the molecule has a long P3 side chain, which protrudes deep into the S3 sub-pocket, particularly that of the T-L site, and specifically interacts with the amino acid side chains of the protein. Based on the electron density map and the precise molecular mass of Mz721, and the acidic environment of the S3 sub-pocket of T-L and CT-L sites, we propose that the amino acid in the P3 position of Mz721 is an arginine. The P3 position makes Mz721 a potent inhibitor of the T-L activity of the yeast 20S proteasome.

Although TMC-95A has been chemically synthesized, the challenging total synthesis and the limited choice of the building blocks of the P1 position restricted the structure-activity relationship (SAR) studies. For the P3 site, only asparagine, valine and

leucine were tried without altering the cyclic peptide backbone, and the potency for the T-L site was not improved (256, 262-264, 266). In another study, the P3 site was changed to arginine but the biaryl linkage was also substituted by a biphenyl-ether linkage. Although this linkage was aimed to mimic the strained backbone conformation of TMC-95A and meanwhile simplify the synthesis, the potency of the inhibitor was also dramatically compromised (259, 266). This was either because the loss of backbone rigidity cost more entropy during binding, or because substituting the oxindole ring also removed the critical hydrogen bond between the carbonyl on the oxindole ring and the backbone NH of Gly23. Taken together, none of the modifications to date have improved the potency of TMC-95A against any of the three activities (Table 4, 5). Our studies of Mz721 proved the concept that in order to further potentiate the inhibitor against a particular active site, modification of the P3 position is a practical route. Moreover, the S3 sub-pocket contributed by the adjacent subunit should be utilized but with caution during inhibitor development because of the amino acid sequence differences between different organisms. In the Mz721 example, the P3-arginine strongly interacted with the  $\beta$ 2-Gln22 and  $\beta$ 3-Asp131. However, in the human proteasome, these residues are substituted by glutamate and methionine. This might be the part of the reason why the whole cell activity of Mz721 was not so impressive, although the charged side-chain of P3-arginine could also be a liability in drug uptake.

Similar to the TMC-95 series compounds (258)(259)(258)(257), the (*Z*)-propyl-1-enyl of the P1 position barely interacts with the reactive nucleophile of Thr10<sup>y</sup> or other residues in the S1 pocket of all three active sites. So far, the (*Z*)-propyl-1-enyl

group has only been modified into propyl, (Z)-propyl-2-enyl, norleucine-NH<sub>2</sub> and a longer ketomethylene containing group (262-264). Apart from (Z)-propyl-2-enyl substitution which showed similar potency to TMC-95A, the others all lost hundreds to thousands-fold of the potency especially against the CT-L site. These suggest that the double bond which stabilizes the P1 group is essential and should be maintained in future SAR studies. Notably, in the proteasome-TMC-95A complex structure reported in (258), the amide of the P1 position of TMC-95A adopted a *cis* configuration, which is rare in the natural peptide bonds except for proline. This may need re-examination. In the Mz721 bound structure, the amide adopted a more common *trans* configuration.

In the enzyme activity assay, we observed that Mz721 below 1  $\mu$ M enhanced PGPH activity (Fig. 51). It has been shown that CT-L site substrates allosterically activated PGPH activity (282). Since Mz721 is also a peptide which can mimic a substrate, it is likely that by binding to the CT-L site, Mz721 allosterically activates the PGPH site. In agreement with this hypothesis, our data showed that Mz721 started to exhibit inhibitory effect on PGPH site after CT-L activity is fully inhibited (Fig. 51).

In summary, Mz721 is a potent proteasome inhibitor as well as an informative lead compound in guiding the design of cyclic peptide proteasome inhibitors.

CHAPTER IV  
CONCLUSIONS AND FUTURE DIRECTIONS

**IV.1 MtPrpR and the MCC transcriptional regulation**

MCC is a critical pathway of *Mtb*, even though its *in vivo* essentiality remains to be learned. We showed that MCC is transcriptionally regulated by MtPrpR through direct binding to CoA derivatives and an iron-sulfur cluster. There are several major questions to be answered in order to fully understand the regulatory mechanisms of MCC and potentially the glyoxylate shunt in *Mtb*. The questions and future directions are listed below.

*IV.1.1 Propionyl-CoA bound MtPrpR structure*

So far, we still do not know the actual structure of MtPrpR when it has propionyl-CoA bound, although our modeling suggested that the helix  $\alpha 1$  is expected to move by approximately  $15^\circ$ . It will be very desirable to confirm this potential conformational change and to see what follows, particularly the positions of the HTH DNA-binding domains. We have already tried heterologous expression of the recombinant MtPrpR<sub>81-486</sub> in *E. coli* using the M9-propionate media but did not see propionyl-CoA bound in the structure. One experiment that might be done in the future is to express MtPrpR in mycobacteria, i.e. *M. smegmatis*, a non-pathogenic fast growing surrogate of *M. tuberculosis*. However, there is another potential problem that may account for our current and future failure to visualize propionyl-CoA. The crystal

structure is a representation of the ensemble of proteins. If the vast majority of the protein molecules bind to CoA, the feature of propionyl-CoA will be averaged and thus may never be discerned. Moreover, we do not know whether all four chains in the MtPrpR tetramer bind to propionyl-CoA or only a subset bind, because physiologically it will be difficult for all four chains to bind propionyl-CoA simultaneously. Therefore, it is likely that the transcription can be activated when one or two subunits have propionyl-CoA bound. This can bring severe homogeneity problem in crystal packing. So far, the main stream structural biology techniques cannot overcome this problem. Single particle CryoEM, for example, uses similar concept of classifying and averaging particles. Unless the conformational change is significant enough, it may not be visualized. In this case, it may be helpful to include the HTH domain to potentially see a more remarkable conformational change, and also introduce the recognition DNA to stabilize a potential propionyl-CoA bound protein.

#### IV.1.2 *MtPrpR tetramer mystery*

If the recognition operator of MtPrpR is a 17 bp DNA, why would MtPrpR be a tetramer? Typically, the tetrameric transcriptional regulators can bend DNA operators. However, this requires the operators being separated by hundreds of base pairs in order not to generate too much strain. Footprinting experiments may be carried out to confirm the DNA binding region. Although DNase I footprinting of MtPrpR-DNA complex has been reported before, the result was very ambiguous (135). In a tetrameric photoreceptor transcriptional regulator CarH structure, the four HTH domains followed an unusual

organization with three involved in binding to a 26-bp DNA segment and the fourth one was disordered (283). It will be interesting to see how the four HTH domains in MtPrpR arrange. Single particle CryoEM may be the technique to resort to especially if longer DNA fragment is required for protein binding, because long piece of DNA will make crystallization much more difficult.

#### IV.1.3 *Signaling molecules that may induce CoA derivative exchange*

Based on our experiences of MtPrpR, the protein binds to CoA with very high affinity, and we were not able to release CoA unless we destroyed the protein. This does not seem to be the case *in vivo*, because it will be very uneconomic not to recycle transcriptional regulators, especially for bacteria that survive in nutrient restricted environment. Ideally, there should be certain signal that can open the lid of the CoA binding pocket and allow CoA exchange. The signal may be related to the iron-sulfur cluster or MCC intermediates. It will be very challenging but of great significance to identify the factor required for CoA release so that MtPrpR could be manipulated.

#### **IV.2 Mz721 purification, identification and characterization**

Mz721 potently inhibits the yeast 20S proteasome, particularly the T-L activity. The structure of the inhibitor has been proposed based on the high-resolution mass spectrometry data and its binding mode on the proteasome active sites. The structures of Mz721 and its analogs, including the TMC-95 series, are very unique, and it will be of



great value to identify its biosynthetic pathway and characterize the natural product assembly lines. The remaining questions and future directions are listed below.

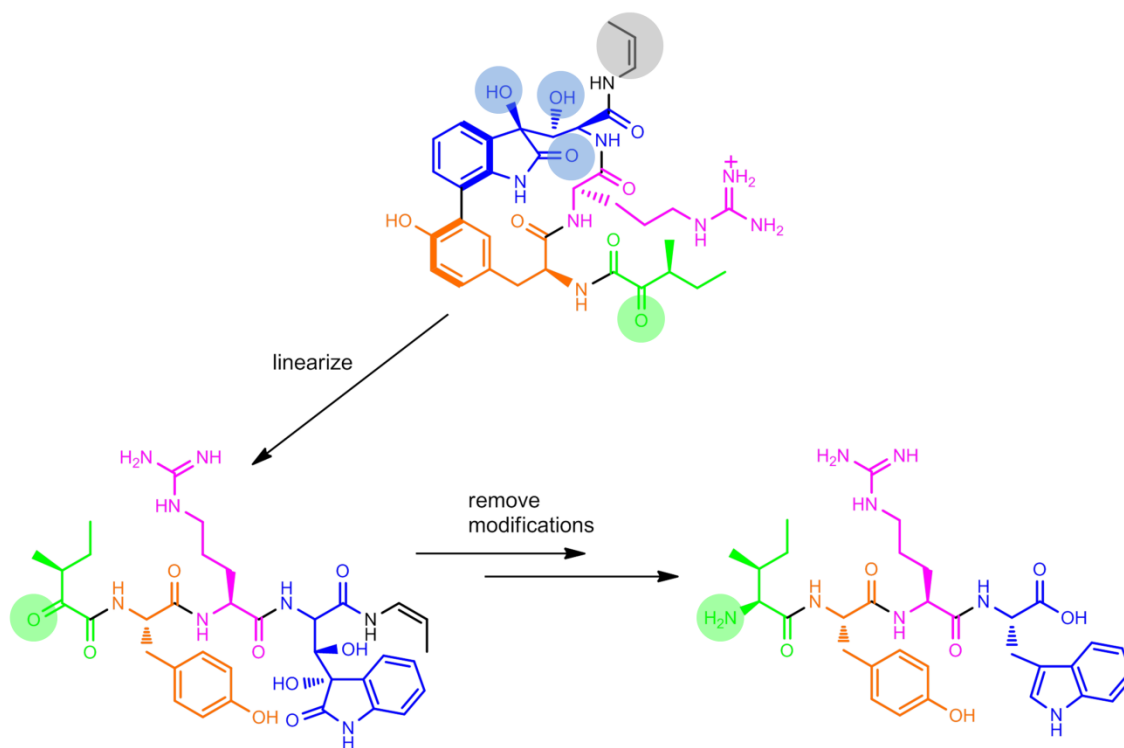
#### IV.2.1 *Genome sequencing and mining of the biosynthetic pathways of Mz721 and analogs*

So far, we have identified the genus and species information of the fungus as *A. umbrinum*. However, apart from the rDNA region, the genome information of *A. umbrinum* is barely available. This complicates further investigation of the biosynthetic pathway of Mz721 and analogs. Thereby, the whole genome sequencing and the gene annotation of *A. umbrinum* need to be conducted. Provided that the whole genome of *A. umbrinum* was sequenced and annotated, we need to determine which genes should be targeted. Clearly, Mz721 and its analogs are peptide based natural products, but it is not clear whether they are ribosomal peptides (RP) or non-ribosomal peptides (NRP) just based on the structure. Both RP and NRP synthetic systems can produce peptides with extensive “posttranslational” modifications, and the same type of peptides could be synthesized using either system in different organisms (284). Although it will be extremely challenging to unravel the biosynthetic pathway of the natural products, the structure of Mz721 provides some information that may guide the genome mining.

Typically, genes for both RP and NRP synthetic pathways are clustered. This prerequisite should simplify the search. The structure of Mz721 showed it is a tripeptide. However, it is likely that the precursor of Mz721 contains at least four amino acids Ile-Tyr-Arg-Trp (Fig. 60), and the isoleucine is deaminated during Mz721 biosynthesis

(green shade). Therefore, we should anticipate a deaminase-encoding gene in the cluster. Mz721 contains a unique C-C biaryl linkage between the oxindole and phenol. This type of reactions during natural product syntheses involves cytochrome P450 monooxygenase, such as OxyC in vancomycin biosynthesis (285) and ComI in complestatin biosynthesis (286). Furthermore, the tryptophan residue is hyperoxidized (blue shades). The 3-hydroxy 2-oxindole type of modification was also observed on several other natural products including maremycin A, welwitindolinone C, donaxaridine, etc (287). In the maremycin A, the 2-oxo group was added by MarE, a non-canonical tryptophan dioxygenase (TDO), which is a heme-containing enzyme (288). Although the modifications on the indole ring may be catalyzed by enzyme other than a dioxygenase, we may expect multiple metalloenzymes in the cluster such as cytochrome P450 oxygenases. Mz721 also contains an *n*-propylene group ((*Z*)-propyl-1-enyl group, grey shade), which is potentially reduced from *n*-propyl. So we may expect an alkyl transferase and a reductase in the gene cluster.

The NRP synthetic systems are pipelines containing different modules. A majority of NRPs contain D-amino acids, most of which are converted from L-amino acids by epimerase domains (E domain) in the synthetic modules (289). However, there is no D-amino acid in Mz721, which means if the compound is synthesized by NRP synthase, E domains are not expected in any modules of the synthetic pipeline.



**Figure 60. Retro-biosynthesis analysis of Mz721**

**The modification sites of Mz721 are shaded. Mz721 is first linearized and then the modifications are removed to potentially reveal its original amino acid sequence. The precursor is likely to contain Ile-Tyr-Arg-Trp tetrapeptide.**

#### IV.2.2 Mz721 extraction and purification

If the identification of the biosynthetic pathway eventually becomes possible, the synthetic gene cluster may be cloned and overexpressed heterologously to improve the yield of the natural products. If that is not practical, the fungal growth condition needs to be optimized to enhance Mz721 production. In addition, the purification methods also need to be improved. For large scale purification, preparative HPLC may not be suitable for the initial steps of crude extract separation, provided that the percentage of Mz721 in the crude extract is extremely low. Therefore, separating the metabolites based on their

partition in different solvents can be used in the preliminary purification before the chromatography-based polishing purification. In addition, since Mz721 is likely to contain a positively charged arginine side chain, ion exchange chromatography may be used.

#### *IV.2.3 Mz721 accurate structure identification*

Natural product structure identification has always been a challenging task. So far, we do not have pure enough compound to perform NMR. If we were able to obtain sufficient amount of Mz721, we may purify the compound by crystallization, or even solve the structure using small molecule X-ray crystallography. X-ray diffraction (XRD) is a powerful technique in elucidating the structures of natural products, and in determining the absolute configuration of the chiral centers when a reference chiral center is available to solve the X-ray phase ambiguity (290). Fortunately, since we already know from the structure of proteasome-Mz721 complex that the amino acids in Mz721 are in L- form, these C $\alpha$  chiral centers can serve as internal references for the rest of the chiral centers in the molecule.

## REFERENCES

1. Haberland ME & Reynolds JA (1973) Self-Association of Cholesterol in Aqueous-Solution. *Proceedings of the National Academy of Sciences of the United States of America* 70(8):2313-2316.
2. Goluszko P & Nowicki B (2005) Membrane cholesterol: a crucial molecule affecting interactions of microbial pathogens with mammalian cells. *Infection and immunity* 73(12):7791-7796.
3. Liu SL, *et al.* (2017) Orthogonal lipid sensors identify transbilayer asymmetry of plasma membrane cholesterol. *Nature chemical biology* 13(3):268-274.
4. Gatfield J & Pieters J (2000) Essential role for cholesterol in entry of mycobacteria into macrophages. *Science* 288(5471):1647-1650.
5. Ferrari G, Langen H, Naito M, & Pieters J (1999) A coat protein on phagosomes involved in the intracellular survival of mycobacteria. *Cell* 97(4):435-447.
6. Sohngen NL (1913) Oxidation of petroleum, parffin, paraftinoil and benzine by microbes. *P K Akad Wet-Amsterd* 15:1145-1151.
7. Tak JD (1942) On bacteria decomposing cholesterol. *Antonie van Leeuwenhoek* 8(1):32-40.
8. Turfitt GE (1944) Microbiological agencies in the degradation of steroids I The cholesterol-decomposing organisms of soils. *Journal of bacteriology* 47(6):487-493.
9. Turfitt GE (1947) Microbiological Agencies in the Degradation of Steroids: II. Steroid Utilization by the Microflora of Soils. *Journal of bacteriology* 54(5):557-562.
10. Turfitt GE (1948) The Microbiological Degradation of Steroids .4. Fission of the Steroid Molecule. *Biochem J* 42(3):376-383.
11. Dodson RM & Muir RD (1958) Microbiological Transformations .3. The Hydroxylation of Steroids at C-9. *J Am Chem Soc* 80(22):6148-6148.
12. Dodson RM & Muir RD (1961) Microbiological Transformations .6. Microbiological Aromatization of Steroids. *J Am Chem Soc* 83(22):4627-+.

13. Sih CJ, Wang KC, & Tai HH (1967) C-22 acid intermediates in the microbiological cleavage of the cholesterol side chain. *J Am Chem Soc* 89(8):1956-1957.
14. Horinouchi M, *et al.* (2004) Steroid degradation gene cluster of *Comamonas testosteroni* consisting of 18 putative genes from meta-cleavage enzyme gene *tesB* to regulator gene *tesR*. *Biochemical and biophysical research communications* 324(2):597-604.
15. Ehrt S, Schnappinger D, & Rhee KY (2018) Metabolic principles of persistence and pathogenicity in *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* 16(8):496-507.
16. Timm J, *et al.* (2003) Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proceedings of the National Academy of Sciences of the United States of America* 100(24):14321-14326.
17. Voskuil MI, Visconti KC, & Schoolnik GK (2004) *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis* 84(3-4):218-227.
18. Via LE, *et al.* (2008) Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infection and immunity* 76(6):2333-2340.
19. Harper J, *et al.* (2012) Mouse Model of Necrotic Tuberculosis Granulomas Develops Hypoxic Lesions. *J Infect Dis* 205(4):595-602.
20. Ehlers S & Schaible UE (2013) The granuloma in tuberculosis: dynamics of a host-pathogen collusion. *Front Immunol* 3.
21. Prosser G, *et al.* (2017) The bacillary and macrophage response to hypoxia in tuberculosis and the consequences for T cell antigen recognition. *Microbes Infect* 19(3):177-192.
22. Segal W & Bloch H (1956) Biochemical Differentiation of *Mycobacterium-Tuberculosis* Grown *In vivo* and *In vitro*. *Journal of bacteriology* 72(2):132-141.
23. Munoz-Elias EJ & McKinney JD (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. *Nature medicine* 11(6):638-644.
24. Boshoff HI & Barry CE (2005) A low-carb diet for a high-octane pathogen. *Nature medicine* 11(6):599-600.

25. Munoz-Elias EJ, Upton AM, Cherian J, & McKinney JD (2006) Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. *Mol Microbiol* 60(5):1109-1122.
26. Cole ST, *et al.* (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393(6685):537-+.
27. McKinney JD, *et al.* (2000) Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406(6797):735-738.
28. Rhee KY, *et al.* (2011) Central carbon metabolism in Mycobacterium tuberculosis: an unexpected frontier. *Trends Microbiol* 19(7):307-314.
29. Sasseti CM & Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America* 100(22):12989-12994.
30. Schnappinger D, *et al.* (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment. *J Exp Med* 198(5):693-704.
31. Van der Geize R, *et al.* (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 104(6):1947-1952.
32. Sasseti CM, Boyd DH, & Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 98(22):12712-12717.
33. Arruda S, Bomfim G, Knights R, Huimabyron T, & Riley LW (1993) Cloning of an Mycobacterium-Tuberculosis DNA Fragment Associated with Entry and Survival inside Cells. *Science* 261(5127):1454-1457.
34. Pandey AK & Sasseti CM (2008) Mycobacterial persistence requires the utilization of host cholesterol (vol 105, pg 4376, 2008). *Proceedings of the National Academy of Sciences of the United States of America* 105(26):9130-9130.
35. Joshi SM, *et al.* (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proceedings of the National Academy of Sciences of the United States of America* 103(31):11760-11765.

36. Rengarajan J, Bloom BR, & Rubin EJ (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 102(23):8327-8332.
37. Griffin JE, *et al.* (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *Plos Pathog* 7(9):e1002251.
38. Nesbitt NM, *et al.* (2010) A Thiolase of *Mycobacterium tuberculosis* Is Required for Virulence and Production of Androstenedione and Androstadienedione from Cholesterol. *Infection and immunity* 78(1):275-282.
39. Brzostek A, Pawelczyk J, Rumijowska-Galewicz A, Dziadek B, & Dziadek J (2009) *Mycobacterium tuberculosis* Is Able To Accumulate and Utilize Cholesterol. *Journal of bacteriology* 191(21):6584-6591.
40. Mohn WW, *et al.* (2008) The Actinobacterial *mce4* Locus Encodes a Steroid Transporter. *Journal of Biological Chemistry* 283(51):35368-35374.
41. Nazarova EV, *et al.* (2017) Rv3723/LucA coordinates fatty acid and cholesterol uptake in *Mycobacterium tuberculosis*. *Elife* 6.
42. Ramon-Garcia S, Stewart GR, Hui ZK, Mohn WW, & Thompson CJ (2015) The mycobacterial P55 efflux pump is required for optimal growth on cholesterol. *Virulence* 6(5):444-448.
43. Irving AA, Duchow EG, Plum LA, & DeLuca HF (2018) Vitamin D deficiency in the ApcPirc/+ rat does not exacerbate colonic tumorigenesis, while low dietary calcium might be protective. *Dis Model Mech* 11(3).
44. Kendall SL, Rison SCG, Movahedzadeh F, Frita R, & Stoker NG (2004) What do microarrays really tell us about M-tuberculosis? *Trends Microbiol* 12(12):537-544.
45. Talaat AM, Lyons R, Howard ST, & Johnston SA (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proceedings of the National Academy of Sciences of the United States of America* 101(13):4602-4607.
46. Kendall SL, *et al.* (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol Microbiol* 65(3):684-699.



47. Ho NAT, *et al.* (2016) The Structure of the Transcriptional Repressor KstR in Complex with CoA Thioester Cholesterol Metabolites Sheds Light on the Regulation of Cholesterol Catabolism in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 291(14):7256-7266.
48. Uhia I, Galan B, Medrano FJ, & Garcia JL (2011) Characterization of the KstR-dependent promoter of the gene for the first step of the cholesterol degradative pathway in *Mycobacterium smegmatis*. *Microbiol-Sgm* 157:2670-2680.
49. Uhia I, Galan B, Morales V, & Garcia JL (2011) Initial step in the catabolism of cholesterol by *Mycobacterium smegmatis* mc2155. *Environ Microbiol* 13(4):943-959.
50. Yang XX, Gao J, Smith I, Dubnau E, & Sampson NS (2011) Cholesterol Is Not an Essential Source of Nutrition for *Mycobacterium tuberculosis* during Infection. *Journal of bacteriology* 193(6):1473-1476.
51. Wipperman MF, Sampson NS, & Thomas ST (2014) Pathogen roid rage: Cholesterol utilization by *Mycobacterium tuberculosis*. *Critical reviews in biochemistry and molecular biology* 49(4):269-293.
52. Garcia-Fernandez E, Medrano FJ, Galan B, & Garcia JL (2014) Deciphering the Transcriptional Regulation of Cholesterol Catabolic Pathway in Mycobacteria IDENTIFICATION OF THE INDUCER OF KstR REPRESSOR. *Journal of Biological Chemistry* 289(25):17576-17588.
53. Kendall SL, *et al.* (2010) Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: *kstR* and *kstR2*. *Microbiol-Sgm* 156:1362-1371.
54. Casabon I, *et al.* (2013) Regulation of the KstR2 regulon of *Mycobacterium tuberculosis* by a cholesterol catabolite. *Mol Microbiol* 89(6):1201-1212.
55. Casabon I, Crowe AM, Liu J, & Eltis LD (2013) FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria. *Mol Microbiol* 87(2):269-283.
56. Capyk JK, Casabon I, Gruninger R, Strynadka NC, & Eltis LD (2011) Activity of 3-Ketosteroid 9 alpha-Hydroxylase (KshAB) Indicates Cholesterol Side Chain and Ring Degradation Occur Simultaneously in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 286(47):40717-40724.
57. Ouellet H, Johnston JB, & de Montellano PRO (2011) Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends Microbiol* 19(11):530-539.

58. Capyk JK, *et al.* (2009) Mycobacterial Cytochrome P450 125 (Cyp125) Catalyzes the Terminal Hydroxylation of C27 Steroids. *Journal of Biological Chemistry* 284(51):35534-35542.
59. Ouellet H, *et al.* (2010) Mycobacterium tuberculosis CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Mol Microbiol* 77(3):730-742.
60. Brzostek A, Dziadek B, Rumijowska-Galewicz A, Pawelczyk J, & Dziadek J (2007) Cholesterol oxidase is required for virulence of Mycobacterium tuberculosis. *Fems Microbiol Lett* 275(1):106-112.
61. Yang XX, Dubnau E, Smith I, & Sampson NS (2007) Rv1106c from Mycobacterium tuberculosis is a 3 beta-hydroxysteroid dehydrogenase. *Biochemistry* 46(31):9058-9067.
62. Chang JC, Harik NS, Liao RP, & Sherman DR (2007) Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *J Infect Dis* 196(5):788-795.
63. McLean KJ, *et al.* (2009) The Structure of Mycobacterium tuberculosis CYP125 MOLECULAR BASIS FOR CHOLESTEROL BINDING IN A P450 NEEDED FOR HOST INFECTION. *Journal of Biological Chemistry* 284(51):35524-35533.
64. Petrusma M, Dijkhuizen L, & van der Geize R (2009) Rhodococcus rhodochrous DSM 43269 3-Ketosteroid 9 alpha-Hydroxylase, a Two-Component Iron-Sulfur-Containing Monooxygenase with Subtle Steroid Substrate Specificity. *Appl Environ Microb* 75(16):5300-5307.
65. Rosloniec KZ, *et al.* (2009) Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in Rhodococcus jostii RHA1. *Mol Microbiol* 74(5):1031-1043.
66. Johnston JB, Ouellet H, & de Montellano PRO (2010) Functional Redundancy of Steroid C-26-monooxygenase Activity in Mycobacterium tuberculosis Revealed by Biochemical and Genetic Analyses. *Journal of Biological Chemistry* 285(47):36352-36360.
67. Wilbrink MH, Petrusma M, Dijkhuizen L, & 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 Microb* 77(13):4455-4464.

68. Casabon I, Swain K, Crowe AM, Eltis LD, & Mohn WW (2014) Actinobacterial Acyl Coenzyme A Synthetases Involved in Steroid Side-Chain Catabolism. *Journal of bacteriology* 196(3):579-587.
69. Lu R, Schmitz W, & Sampson NS (2015) alpha-Methyl Acyl CoA Racemase Provides Mycobacterium tuberculosis Catabolic Access to Cholesterol Esters. *Biochemistry* 54(37):5669-5672.
70. Yang M, *et al.* (2015) Unraveling Cholesterol Catabolism in Mycobacterium tuberculosis: ChsE4-ChsE5 alpha(2)beta(2) Acyl-CoA Dehydrogenase Initiates beta-Oxidation of 3-Oxo-cholest-4-en-26-oyl CoA. *Acs Infect Dis* 1(2):110-125.
71. Crowe AM, *et al.* (2017) Catabolism of the Last Two Steroid Rings in Mycobacterium tuberculosis and Other Bacteria. *Mbio* 8(2).
72. Crowe AM, *et al.* (2018) IpdAB, a virulence factor in Mycobacterium tuberculosis, is a cholesterol ring-cleaving hydrolase. *Proceedings of the National Academy of Sciences of the United States of America* 115(15):E3378-E3387.
73. Thomas ST, VanderVen BC, Sherman DR, Russell DG, & Sampson NS (2011) Pathway Profiling in Mycobacterium tuberculosis ELUCIDATION OF CHOLESTEROL-DERIVED CATABOLITE AND ENZYMES THAT CATALYZE ITS METABOLISM. *Journal of Biological Chemistry* 286(51):43668-43678.
74. Thomas ST & Sampson NS (2013) Mycobacterium tuberculosis Utilizes a Unique Heterotetrameric Structure for Dehydrogenation of the Cholesterol Side Chain. *Biochemistry* 52(17):2895-2904.
75. Yang M, Guja KE, Thomas ST, Garcia-Diaz M, & Sampson NS (2014) A Distinct MaoC-like Enoyl-CoA Hydratase Architecture Mediates Cholesterol Catabolism in Mycobacterium tuberculosis. *Acs Chem Biol* 9(11):2632-2645.
76. Gilbert S, Hood L, & Seah SYK (2018) Characterization of an Aldolase Involved in Cholesterol Side Chain Degradation in Mycobacterium tuberculosis. *Journal of bacteriology* 200(2).
77. Knol J, Bodewits K, Hessels GI, Dijkhuizen L, & Van der Geize R (2008) 3-Keto-5 alpha-steroid Delta'-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.
78. Rohman A, van Oosterwijk N, Thunnissen AMWH, & Dijkstra BW (2013) Crystal Structure and Site-directed Mutagenesis of 3-Ketosteroid Delta(1)-

- Dehydrogenase from *Rhodococcus erythropolis* SQ1 Explain Its Catalytic Mechanism. *Journal of Biological Chemistry* 288(49):35559-35568.
79. Capyk JK, D'Angelo I, Strynadka NC, & Eltis LD (2009) Characterization of 3-Ketosteroid 9 alpha-Hydroxylase, a Rieske Oxygenase in the Cholesterol Degradation Pathway of *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 284(15):9937-9946.
  80. Horinouchi M, Hayashi T, Yamamoto T, & 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 Microb* 69(8):4421-4430.
  81. Dresen C, *et al.* (2010) A Flavin-dependent Monooxygenase from *Mycobacterium tuberculosis* Involved in Cholesterol Catabolism. *Journal of Biological Chemistry* 285(29):22264-22275.
  82. Sucharitakul J, Tinikul R, & Chaiyen P (2014) Mechanisms of reduced flavin transfer in the two-component flavin-dependent monooxygenases. *Arch Biochem Biophys* 555:33-46.
  83. Ballou DP, Entsch B, & Cole LJ (2005) Dynamics involved in catalysis by single-component and two-component flavin-dependent aromatic hydroxylases. *Biochemical and biophysical research communications* 338(1):590-598.
  84. Dai SD, *et al.* (2002) Identification and analysis of a bottleneck in PCB biodegradation. *Nature structural biology* 9(12):934-939.
  85. Siegbahn PEM & Haeffner F (2004) Mechanism for catechol ring-cleavage by non-heme iron extradiol dioxygenases. *J Am Chem Soc* 126(29):8919-8932.
  86. Yam KC, *et al.* (2009) Studies of a Ring-Cleaving Dioxygenase Illuminate the Role of Cholesterol Metabolism in the Pathogenesis of *Mycobacterium tuberculosis*. *Plos Pathog* 5(3).
  87. Lipscomb JD (2008) Mechanism of extradiol aromatic ring-cleaving dioxygenases. *Curr Opin Struct Biol* 18(6):644-649.
  88. Lack NA, *et al.* (2010) Characterization of a Carbon-Carbon Hydrolase from *Mycobacterium tuberculosis* Involved in Cholesterol Metabolism. *Journal of Biological Chemistry* 285(1):434-443.
  89. Ruzzini AC, *et al.* (2012) Identification of an Acyl-Enzyme Intermediate in a meta-Cleavage Product Hydrolase Reveals the Versatility of the Catalytic Triad. *J Am Chem Soc* 134(10):4615-4624.

90. Kuatsjah E, Chan ACK, Kobylarz MJ, Murphy MEP, & Eltis LD (2017) The bacterial meta-cleavage hydrolase LigY belongs to the amidohydrolase superfamily, not to the alpha/beta-hydrolase superfamily. *Journal of Biological Chemistry* 292(44):18290-18302.
91. Dong LH, Zhang SJ, & Liu YJ (2017) A water-assisted nucleophilic mechanism utilized by BphD, the meta-cleavage product hydrolase in biphenyl degradation. *J Mol Graph Model* 76:448-455.
92. Carere J, McKenna SE, Kimber MS, & Seah SYK (2013) Characterization of an Aldolase-Dehydrogenase Complex from the Cholesterol Degradation Pathway of *Mycobacterium tuberculosis*. *Biochemistry* 52(20):3502-3511.
93. Manjasetty BA, Powlowski J, & Vrielink A (2003) Crvstal structure of a bifunctional aldolase-dehydrogenase: Sequestering a reactive and volatile intermediate. *Proceedings of the National Academy of Sciences of the United States of America* 100(12):6992-6997.
94. Lengyel P, Mazumder R, & Ochoa S (1960) Mammalian Methylmalonyl Isomerase and Vitamin B(12) Coenzymes. *Proceedings of the National Academy of Sciences of the United States of America* 46(10):1312-1318.
95. Banerjee R (2006) B12 trafficking in mammals: A for coenzyme escort service. *Acs Chem Biol* 1(3):149-159.
96. Savvi S, *et al.* (2008) Functional characterization of a vitamin B(12)-dependent methylmalonyl pathway in *Mycobacterium tuberculosis*: Implications for propionate metabolism during growth on fatty acids. *Journal of bacteriology* 190(11):3886-3895.
97. Han Y, Hawkins AS, Adams MW, & Kelly RM (2012) Epimerase (Msed\_0639) and mutase (Msed\_0638 and Msed\_2055) convert (S)-methylmalonyl-coenzyme A (CoA) to succinyl-CoA in the *Metallosphaera sedula* 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ Microbiol* 78(17):6194-6202.
98. Muller M, *et al.* (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76(2):444-495.
99. Gago G, Kurth D, Diacovich L, Tsai SC, & Gramajo H (2006) Biochemical and structural characterization of an essential acyl coenzyme A carboxylase from *Mycobacterium tuberculosis*. *Journal of bacteriology* 188(2):477-486.
100. Lin TW, *et al.* (2006) Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium*

- tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* 103(9):3072-3077.
101. Mancia F, *et al.* (1996) How coenzyme B12 radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. *Structure* 4(3):339-350.
  102. Korotkova N & Lidstrom ME (2004) MeaB is a component of the methylmalonyl-CoA mutase complex required for protection of the enzyme from inactivation. *Journal of Biological Chemistry* 279(14):13652-13658.
  103. Edwards TE, *et al.* (2015) Crystal structures of Mycobacterial MeaB and MMAA-like GTPases. *J Struct Funct Genomics* 16(2):91-99.
  104. Griffin JE, *et al.* (2012) Cholesterol Catabolism by Mycobacterium tuberculosis Requires Transcriptional and Metabolic Adaptations. *Chemistry & biology* 19(2):218-227.
  105. Lee W, VanderVen BC, Fahey RJ, & Russell DG (2013) Intracellular Mycobacterium tuberculosis Exploits Host-derived Fatty Acids to Limit Metabolic Stress. *Journal of Biological Chemistry* 288(10):6788-6800.
  106. Raux E, Schubert HL, & Warren MJ (2000) Biosynthesis of cobalamin (vitamin B-12): a bacterial conundrum. *Cell Mol Life Sci* 57(13-14):1880-1893.
  107. Rodionov DA, Vitreschak AG, Mironov AA, & Gelfand MS (2003) Comparative Genomics of the vitamin B-12 metabolism and regulation in prokaryotes. *Journal of Biological Chemistry* 278(42):41148-41159.
  108. Young DB, Comas I, & de Carvalho LP (2015) Phylogenetic analysis of vitamin B12-related metabolism in Mycobacterium tuberculosis. *Front Mol Biosci* 2:6.
  109. Min CH, *et al.* (1993) Isolation, Structure, and Genetically-Engineered Synthesis of Precorrin-5, the Pentamethylated Intermediate of Vitamin-B12 Biosynthesis. *J Am Chem Soc* 115(22):10380-10381.
  110. Gopinath K, *et al.* (2013) A vitamin B-12 transporter in Mycobacterium tuberculosis. *Open Biol* 3.
  111. Warner DF, Savvi S, Mizrahi V, & Dawes SS (2007) A riboswitch regulates expression of the coenzyme B-12-independent methionine synthase in Mycobacterium tuberculosis: Implications for differential methionine synthase function in strains H37Rv and CDC1551. *Journal of bacteriology* 189(9):3655-3659.

112. Mowa MB, Warner DF, Kaplan G, Kana BD, & Mizrahi V (2009) Function and Regulation of Class I Ribonucleotide Reductase-Encoding Genes in Mycobacteria. *Journal of bacteriology* 191(3):985-995.
113. Jain M, *et al.* (2007) Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. *Proceedings of the National Academy of Sciences of the United States of America* 104(12):5133-5138.
114. Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis* 83(1-3):91-97.
115. Jackson M, Stadthagen G, & Gicquel B (2007) Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: Biosynthesis, transport, regulation and biological activities. *Tuberculosis* 87(2):78-86.
116. Domenech P & Reed MB (2009) Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from Mycobacterium tuberculosis grown in vitro: implications for virulence studies. *Microbiol-Sgm* 155:3532-3543.
117. Rousseau C, *et al.* (2003) Sulfolipid deficiency does not affect the virulence of Mycobacterium tuberculosis H37Rv in mice and guinea pigs. *Infection and immunity* 71(8):4684-4690.
118. Cox JS, Chen B, McNeil M, & Jacobs WR (1999) Complex lipid determine tissue specific replication of Mycobacterium tuberculosis in mice. *Nature* 402(6757):79-83.
119. Rainwater DL & Kolattukudy PE (1985) Fatty-Acid Biosynthesis in Mycobacterium-Tuberculosis Var Bovis-Bacillus-Calmette-Guerin - Purification and Characterization of a Novel Fatty-Acid Synthase, Mycocerosic Acid Synthase, Which Elongates N-Fatty Acyl-Coa with Methylmalonyl-Coa. *Journal of Biological Chemistry* 260(1):616-623.
120. Maskens K & Polgar N (1973) Absolute-Configuration of Ch(Ome)-Chme - System in Phthiocerols. *J Chem Soc Perk T 1* (17):1909-1912.
121. Daffe M (1991) Further Stereochemical Studies of Phthiocerol and Phenol Phthiocerol in Mycobacteria. *Res Microbiol* 142(4):405-410.
122. Rainwater DL & Kolattukudy PE (1982) Isolation and Characterization of Acyl Coenzyme-a Carboxylases from Mycobacterium-Tuberculosis and Mycobacterium-Bovis, Which Produce Multiple Methyl-Branched Mycocerosic Acids. *Journal of bacteriology* 151(2):905-911.

123. Trivedi OA, *et al.* (2005) Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. *Molecular cell* 17(5):631-643.
124. Sirakova TD, Thirumala AK, Dubey VS, Sprecher H, & Kolattukudy PE (2001) The Mycobacterium tuberculosis pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *Journal of Biological Chemistry* 276(20):16833-16839.
125. Converse SE, *et al.* (2003) MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. *Proceedings of the National Academy of Sciences of the United States of America* 100(10):6121-6126.
126. Dolan SK, *et al.* (2018) Loving the poison: the methylcitrate cycle and bacterial pathogenesis. *Microbiol-Sgm* 164(3):251-259.
127. Gould TA, de Langemheen HV, Munoz-Elias EJ, McKinney JD, & Sacchettini JC (2006) Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis. *Mol Microbiol* 61(4):940-947.
128. Upton AM & McKinney JD (2007) Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. *Microbiol-Sgm* 153:3973-3982.
129. VanderVen BC, *et al.* (2015) Novel Inhibitors of Cholesterol Degradation in Mycobacterium tuberculosis Reveal How the Bacterium's Metabolism Is Constrained by the Intracellular Environment. *Plos Pathog* 11(2).
130. Eoh H & Rhee KY (2014) Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty acids. *Proceedings of the National Academy of Sciences of the United States of America* 111(13):4976-4981.
131. Man WJ, Li Y, Oconnor CD, & Wilton DC (1995) The Binding of Propionyl-Coa and Carboxymethyl-Coa to Escherichia-Coli Citrate Synthase. *Bba-Protein Struct M* 1250(1):69-75.
132. Brock M & Buckel W (2004) On the mechanism of action of the antifungal agent propionate - Propionyl-CoA inhibits glucose metabolism in Aspergillus nidulans. *European Journal of Biochemistry* 271(15):3227-3241.
133. Horswill AR, Dudding AR, & Escalante-Semerena JC (2001) Studies of propionate toxicity in Salmonella enterica identify 2-methylcitrate as a potent inhibitor of cell growth. *Journal of Biological Chemistry* 276(22):19094-19101.



134. Rocco CJ & Escalante-Semerena JC (2010) In *Salmonella enterica*, 2-Methylcitrate Blocks Gluconeogenesis. *Journal of bacteriology* 192(3):771-778.
135. Masiewicz P, Brzostek A, Wolanski M, Dziadek J, & Zakrzewska-Czerwinska J (2014) A Novel Role of the PrpR as a Transcription Factor Involved in the Regulation of Methylcitrate Pathway in *Mycobacterium tuberculosis* (vol 7, e43651, 2012). *PloS one* 9(11).
136. Anonymous (2015) World Health Organization releases 2015 global report on tuberculosis. *Breathe* 11(4):244-244.
137. Gagneux S (2018) Ecology and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 16(4):202-213.
138. Datta P, Shi LB, Bibi N, Balazsi G, & Gennaro ML (2011) Regulation of Central Metabolism Genes of *Mycobacterium tuberculosis* by Parallel Feed-Forward Loops Controlled by Sigma Factor E (sigma(E)). *Journal of bacteriology* 193(5):1154-1160.
139. Zhang HT, *et al.* (2013) Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nature genetics* 45(10):1255-U1217.
140. Hicks ND, *et al.* (2018) Clinically prevalent mutations in *Mycobacterium tuberculosis* alter propionate metabolism and mediate multidrug tolerance. *Nat Microbiol.*
141. Tsang AW, Horswill AR, & Escalante-Semerena JC (1998) Studies of regulation of expression of the propionate (prpBCDE) operon provide insights into how *Salmonella typhimurium* LT2 integrates its 1,2-propanediol and propionate catabolic pathways. *Journal of bacteriology* 180(24):6511-6518.
142. Palacios S & Escalante-Semerena JC (2004) 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of *Salmonella enterica* by the PrpR protein. *Microbiol-Sgm* 150:3877-3887.
143. Plassmeier J, *et al.* (2012) Molecular characterization of PrpR, the transcriptional activator of propionate catabolism in *Corynebacterium glutamicum*. *J Biotechnol* 159(1-2):1-11.
144. Lee SK, Newman JD, & Keasling JD (2005) Catabolite repression of the propionate catabolic genes in *Escherichia coli* and *Salmonella enterica*: Evidence for involvement of the cyclic AMP receptor protein. *Journal of bacteriology* 187(8):2793-2800.

145. Finn RD, *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic acids research* 45(D1):D190-D199.
146. Vujicic-Zagar A, *et al.* (2009) Crystal Structure of the IrrE Protein, a Central Regulator of DNA Damage Repair in Deinococcaceae. *Journal of molecular biology* 386(3):704-716.
147. Ludanyi M, *et al.* (2014) Radiation response in *Deinococcus deserti*: IrrE is a metalloprotease that cleaves repressor protein DdrO. *Mol Microbiol* 94(2):434-449.
148. Micklinghoff JC, *et al.* (2009) Role of the Transcriptional Regulator RamB (Rv0465c) in the Control of the Glyoxylate Cycle in *Mycobacterium tuberculosis*. *Journal of bacteriology* 191(23):7260-7269.
149. Galagan JE, *et al.* (2013) The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* 499(7457):178-183.
150. Carter MS & Alber BE (2015) Transcriptional Regulation by the Short-Chain Fatty Acyl Coenzyme A Regulator (ScfR) PccR Controls Propionyl Coenzyme A Assimilation by *Rhodobacter sphaeroides*. *Journal of bacteriology* 197(19):3048-3056.
151. Rock JM, *et al.* (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* 2(4).
152. DeJesus MA, Sacchettini JC, & Ioerger TR (2013) Reannotation of translational start sites in the genome of *Mycobacterium tuberculosis*. *Tuberculosis* 93(1):18-25.
153. Buchan DWA, Minneci F, Nugent TCO, Bryson K, & Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* 41(W1):W349-W357.
154. Stols L, *et al.* (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Express Purif* 25(1):8-15.
155. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.
156. Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D* 58:1772-1779.

157. Winn MD, *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235-242.
158. Sheldrick GM (2002) Macromolecular phasing with SHELXE. *Z Kristallogr* 217(12):644-650.
159. Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D* 66:470-478.
160. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D* 62:1002-1011.
161. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.
162. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
163. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.
164. Afonine PV, *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* 68:352-367.
165. Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic acids research* 38:W545-W549.
166. Pettersen EF, *et al.* (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.
167. Crack JC, Stapleton MR, Green J, Thomson AJ, & Le Brun NE (2013) Mechanism of [4Fe-4S](Cys)<sub>4</sub> Cluster Nitrosylation Is Conserved among NO-responsive Regulators. *Journal of Biological Chemistry* 288(16):11492-11502.
168. Cruz-Ramos H, *et al.* (2002) NO sensing by FNR: regulation of the Escherichia coli NO-detoxifying flavohaemoglobin, Hmp. *Embo Journal* 21(13):3235-3244.
169. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7.
170. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* 42(W1):W320-W324.

171. Wang BC (1985) Resolution of Phase Ambiguity in Macromolecular Crystallography. *Methods in enzymology* 115:90-112.
172. Hendrickson WA, Horton JR, & Lemaster DM (1990) Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (Mad) - a Vehicle for Direct Determination of 3-Dimensional Structure. *Embo Journal* 9(5):1665-1672.
173. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* 372(3):774-797.
174. Aravind L & Ponting CP (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* 22(12):458-459.
175. Henry JT & Crosson S (2011) Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu Rev Microbiol* 65:261-286.
176. Ho YSJ, Burden LM, & Hurley JH (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *Embo Journal* 19(20):5288-5299.
177. Martinez SE, *et al.* (2005) Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: Modes of ligand binding and dimerization. *Proceedings of the National Academy of Sciences of the United States of America* 102(8):3082-3087.
178. Martinez SE, *et al.* (2002) The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *Proceedings of the National Academy of Sciences of the United States of America* 99(20):13260-13265.
179. Boatright KM & Salvesen GS (2003) Mechanisms of caspase activation. *Curr Opin Cell Biol* 15(6):725-731.
180. Patel R, *et al.* (2018) MMR Deficiency Does Not Sensitize or Compromise the Function of Hematopoietic Stem Cells to Low and High LET Radiation. *Stem Cell Transl Med* 7(7):513-520.
181. Suvorova IA, Ravcheev DA, & Gelfand MS (2012) Regulation and Evolution of Malonate and Propionate Catabolism in Proteobacteria. *Journal of bacteriology* 194(12):3234-3240.
182. Parvez M, *et al.* (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci Rep-Uk* 6.

183. Wagner T, Bellinzoni M, Wehenkel A, O'Hare HM, & Alzari PM (2011) Functional Plasticity and Allosteric Regulation of alpha-Ketoglutarate Decarboxylase in Central Mycobacterial Metabolism. *Chemistry & biology* 18(8):1011-1020.
184. Baughn AD, Garforth SJ, Vilcheze C, & Jacobs WR (2009) An Anaerobic-Type alpha-Ketoglutarate Ferredoxin Oxidoreductase Completes the Oxidative Tricarboxylic Acid Cycle of Mycobacterium tuberculosis. *Plos Pathog* 5(11).
185. Xie LX, *et al.* (2015) First Succinyl-Proteome Profiling of Extensively Drug-Resistant Mycobacterium tuberculosis Revealed Involvement of Succinylation in Cellular Physiology. *J Proteome Res* 14(1):107-119.
186. Zhou ML, Xie LX, Yang ZZ, Zhou JH, & Xie JP (2017) Lysine succinylation of Mycobacterium tuberculosis isocitrate lyase (ICL) fine-tunes the microbial resistance to antibiotics. *J Biomol Struct Dyn* 35(5):1030-1041.
187. Kurthkoti K, *et al.* (2017) The Capacity of Mycobacterium tuberculosis To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas. *Mbio* 8(4).
188. Beinert H, Holm RH, & Munck E (1997) Iron-sulfur clusters: Nature's modular, multipurpose structures. *Science* 277(5326):653-659.
189. Cunningham RP, *et al.* (1989) Endonuclease-Iii Is an Iron Sulfur Protein. *Biochemistry* 28(10):4450-4455.
190. Fu WG, Ohandley S, Cunningham RP, & Johnson MK (1992) The Role of the Iron-Sulfur Cluster in Escherichia-Coli Endonuclease-Iii - a Resonance Raman-Study. *Journal of Biological Chemistry* 267(23):16135-16137.
191. Boon EM, Livingston AL, Chmiel NH, David SS, & Barton JK (2003) DNA-mediated charge transport for DNA repair. *Proceedings of the National Academy of Sciences of the United States of America* 100(22):12543-12547.
192. Earl AM, Mohundro MM, Mian IS, & Battista JR (2002) The IrrE protein of Deinococcus radiodurans R1 is a novel regulator of recA expression. *Journal of bacteriology* 184(22):6216-6224.
193. Ciechanover A (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nature reviews. Molecular cell biology* 6(1):79-87.
194. Varshavsky AV (2006) The early history of the ubiquitin field. *Protein Sci* 15(3):647-654.

195. Etlinger JD & Goldberg AL (1977) A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proceedings of the National Academy of Sciences of the United States of America* 74(1):54-58.
196. Ciechanover A, Hod Y, & Hershko A (1978) A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochemical and biophysical research communications* 81(4):1100-1105.
197. Wilkinson KD & Urban MK (1980) Ubiquitin Is the Atp-Dependent Proteolysis Factor-I of Rabbit Reticulocytes. *Faseb J* 39(6):1682-1682.
198. Goldknopf IL & Busch H (1977) Isopeptide Linkage between Nonhistone and Histone-2a Polypeptides of Chromosomal Conjugate-Protein-A24. *Proceedings of the National Academy of Sciences of the United States of America* 74(3):864-868.
199. Hershko A, Ciechanover A, & Rose IA (1981) Identification of the Active Amino-Acid Residue of the Polypeptide of Atp-Dependent Protein Breakdown. *Journal of Biological Chemistry* 256(4):1525-1528.
200. Ciechanover A, Heller H, Katetzion R, & Hershko A (1981) Activation of the Heat-Stable Polypeptide of the Atp-Dependent Proteolytic System. *P Natl Acad Sci-Biol* 78(2):761-765.
201. Ciechanover A, Elias S, Heller H, & Hershko A (1982) "Covalent affinity" purification of ubiquitin-activating enzyme. *The Journal of biological chemistry* 257(5):2537-2542.
202. Haas AL, Warms JVB, Hershko A, & Rose IA (1982) Ubiquitin-Activating Enzyme - Mechanism and Role in Protein-Ubiquitin Conjugation. *Journal of Biological Chemistry* 257(5):2543-2548.
203. Hershko A, Heller H, Elias S, & Ciechanover A (1983) Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *The Journal of biological chemistry* 258(13):8206-8214.
204. Orłowski M & Wilk S (1981) A multicatalytic protease complex from pituitary that forms enkephalin and enkephalin containing peptides. *Biochemical and biophysical research communications* 101(3):814-822.
205. Hough R & Rechsteiner M (1986) Ubiquitin-Lysozyme Conjugates - Purification and Susceptibility to Proteolysis. *Journal of Biological Chemistry* 261(5):2391-2399.

206. Tanaka K, Ii K, Ichihara A, Waxman L, & Goldberg AL (1986) A high molecular weight protease in the cytosol of rat liver. I. Purification, enzymological properties, and tissue distribution. *The Journal of biological chemistry* 261(32):15197-15203.
207. Tanaka K, Yoshimura T, Ichihara A, Kameyama K, & Takagi T (1986) A high molecular weight protease in the cytosol of rat liver. II. Properties of the purified enzyme. *The Journal of biological chemistry* 261(32):15204-15207.
208. Hough R, Pratt G, & Rechsteiner M (1987) Purification of two high molecular weight proteases from rabbit reticulocyte lysate. *The Journal of biological chemistry* 262(17):8303-8313.
209. Tanaka K, *et al.* (1988) Proteasomes (Multi-Protease Complexes) as 20 S Ring-Shaped Particles in a Variety of Eukaryotic Cells. *Journal of Biological Chemistry* 263(31):16209-16217.
210. Eytan E, Ganoth D, Armon T, & Hershko A (1989) Atp-Dependent Incorporation of 20s Protease into the 26s Complex That Degrades Proteins Conjugated to Ubiquitin - (Protein Breakdown Multicatalytic Proteinase Complex). *Proceedings of the National Academy of Sciences of the United States of America* 86(20):7751-7755.
211. Chau V, *et al.* (1989) A Multiubiquitin Chain Is Confined to Specific Lysine in a Targeted Short-Lived Protein. *Science* 243(4898):1576-1583.
212. Thrower JS, Hoffman L, Rechsteiner M, & Pickart CM (2000) Recognition of the polyubiquitin proteolytic signal. *Embo Journal* 19(1):94-102.
213. Finley D (2009) Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annual review of biochemistry* 78:477-513.
214. Nussbaum AK, *et al.* (1998) Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proceedings of the National Academy of Sciences of the United States of America* 95(21):12504-12509.
215. Matyskiela ME, Lander GC, & Martin A (2013) Conformational switching of the 26S proteasome enables substrate degradation. *Nature structural & molecular biology* 20(7):781-788.
216. Glickman MH, *et al.* (1998) A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94(5):615-623.

217. Kohler A, *et al.* (2001) The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Molecular cell* 7(6):1143-1152.
218. Lowe J, *et al.* (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268(5210):533-539.
219. Groll M, *et al.* (1997) Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386(6624):463-471.
220. Fenteany G, *et al.* (1995) Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* 268(5211):726-731.
221. Wilk S, Pearce S, & Orłowski M (1979) Identification and partial purification of a cation-sensitive neutral endopeptidase from bovine pituitaries. *Life sciences* 24(5):457-464.
222. Wilk S & Orłowski M (1980) Cation-sensitive neutral endopeptidase: isolation and specificity of the bovine pituitary enzyme. *Journal of neurochemistry* 35(5):1172-1182.
223. Kisselev AF, Songyang Z, & Goldberg AL (2000) Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *Journal of Biological Chemistry* 275(20):14831-14837.
224. Adams J (2004) The development of proteasome inhibitors as anticancer drugs. *Cancer cell* 5(5):417-421.
225. Glotzer M, Murray AW, & Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. *Nature* 349(6305):132-138.
226. Pagano M, *et al.* (1995) Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. *Science* 269(5224):682-685.
227. Baeuerle PA & Baltimore D (1988) I kappa B: a specific inhibitor of the NF-kappa B transcription factor. *Science* 242(4878):540-546.
228. Baeuerle PA & Baltimore D (1988) Activation of DNA-Binding Activity in an Apparently Cytoplasmic Precursor of the Nf-Kappa-B Transcription Factor. *Cell* 53(2):211-217.



229. Baeuerle PA & Baltimore D (1989) A 65-Kd Subunit of Active Nf-Kappa-B Is Required for Inhibition of Nf-Kappa-B by I-Kappa-B. *Gene Dev* 3(11):1689-1698.
230. Baeuerle PA & Henkel T (1994) Function and activation of NF-kappa B in the immune system. *Annual review of immunology* 12:141-179.
231. Alkalay I, *et al.* (1995) Stimulation-Dependent I-Kappa-B-Alpha Phosphorylation Marks the Nf-Kappa-B Inhibitor for Degradation Via the Ubiquitin-Proteasome Pathway. *Proceedings of the National Academy of Sciences of the United States of America* 92(23):10599-10603.
232. DiDonato JA, Mercurio F, & Karin M (1995) Phosphorylation of I-Kappa-B-Alpha Precedes but Is Not Sufficient for Its Dissociation from Nf-Kappa-B. *Molecular and cellular biology* 15(3):1302-1311.
233. DiDonato J, *et al.* (1996) Mapping of the inducible I kappa B phosphorylation sites that signal its ubiquitination and degradation. *Molecular and cellular biology* 16(4):1295-1304.
234. Hayden MS & Ghosh S (2004) Signaling to NF-kappaB. *Genes Dev* 18(18):2195-2224.
235. Traenckner EBM, Wilk S, & Baeuerle PA (1994) A Proteasome Inhibitor Prevents Activation of Nf-Kappa-B and Stabilizes a Newly Phosphorylated Form of I-Kappa-B-Alpha That Is Still Bound to Nf-Kappa-B. *Embo Journal* 13(22):5433-5441.
236. Meyer S, Kohler NG, & Joly A (1997) Cyclosporine A is an uncompetitive inhibitor of proteasome activity and prevents NF-kappaB activation. *FEBS letters* 413(2):354-358.
237. Sunwoo JB, *et al.* (2001) Novel proteasome inhibitor PS-341 inhibits activation of nuclear factor-kappa B, cell survival, tumor growth, and angiogenesis in squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 7(5):1419-1428.
238. Kisselev AF, Callard A, & Goldberg AL (2006) Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *The Journal of biological chemistry* 281(13):8582-8590.
239. Lee DH & Goldberg AL (1996) Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 271(44):27280-27284.

240. Berkers CR, *et al.* (2005) Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. *Nature methods* 2(5):357-362.
241. Kisselev AF & Goldberg AL (2001) Proteasome inhibitors: from research tools to drug candidates. *Chemistry & biology* 8(8):739-758.
242. Adams J, *et al.* (1998) Boronic ester and acid compounds. (Google Patents).
243. Groll M, Berkers CR, Ploegh HL, & Ovaas H (2006) Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome. *Structure* 14(3):451-456.
244. Hideshima T, *et al.* (2001) The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells. *Cancer research* 61(7):3071-3076.
245. Adams J (2002) Development of the proteasome inhibitor PS-341. *The oncologist* 7(1):9-16.
246. LeBlanc R, *et al.* (2002) Proteasome inhibitor PS-341 inhibits human myeloma cell growth in vivo and prolongs survival in a murine model. *Cancer research* 62(17):4996-5000.
247. Nawrocki ST, *et al.* (2002) Effects of the proteasome inhibitor PS-341 on apoptosis and angiogenesis in orthotopic human pancreatic tumor xenografts. *Molecular cancer therapeutics* 1(14):1243-1253.
248. Aghajanian C, *et al.* (2002) A Phase I trial of the novel proteasome inhibitor PS341 in advanced solid tumor malignancies. *Clinical Cancer Research* 8(8):2505-2511.
249. Orlovski RZ, *et al.* (2002) Phase I trial of the proteasome inhibitor PS-341 in patients with refractory hematologic malignancies. *J Clin Oncol* 20(22):4420-4427.
250. Kane RC, Bross PF, Farrell AT, & Pazdur R (2003) Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. *The oncologist* 8(6):508-513.
251. Meng L, *et al.* (1999) Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proceedings of the National Academy of Sciences of the United States of America* 96(18):10403-10408.

252. Groll M, Kim KB, Kairies N, Huber R, & Crews CM (2000) Crystal structure of epoxomicin : 20S proteasome reveals a molecular basis for selectivity of alpha',beta '-epoxyketone proteasome inhibitors. *J Am Chem Soc* 122(6):1237-1238.
253. Jain S, Diefenbach C, Zain J, & O'Connor OA (2011) Emerging role of carfilzomib in treatment of relapsed and refractory lymphoid neoplasms and multiple myeloma. *Core evidence* 6:43-57.
254. Orłowski RZ (2013) Why Proteasome Inhibitors Cannot ERADicate Multiple Myeloma. *Cancer cell* 24(3):275-277.
255. Garraway LA, Verweij J, & Ballman KV (2013) Precision Oncology: An Overview. *J Clin Oncol* 31(15):1803-1805.
256. Koguchi Y, *et al.* (2000) TMC-95A, B, C, and D, novel proteasome inhibitors produced by *Apiospora montagnei* Sacc. TC 1093. Taxonomy, production, isolation, and biological activities. *The Journal of antibiotics* 53(2):105-109.
257. Kohno J, *et al.* (2000) Structures of TMC-95A-D: novel proteasome inhibitors from *Apiospora montagnei* sacc. TC 1093. *The Journal of organic chemistry* 65(4):990-995.
258. Groll M, Koguchi Y, Huber R, & Kohno J (2001) Crystal structure of the 20 S proteasome:TMC-95A complex: a non-covalent proteasome inhibitor. *Journal of molecular biology* 311(3):543-548.
259. Groll M, Gotz M, Kaiser M, Weyher E, & Moroder L (2006) TMC-95-based inhibitor design provides evidence for the catalytic versatility of the proteasome. *Chemistry & biology* 13(6):607-614.
260. Lin SN & Danishefsky SJ (2002) The total synthesis of proteasome inhibitors TMC-95A and TMC-95B: Discovery of a new method to generate cis-propenyl amides. *Angew Chem Int Edit* 41(3):512-515.
261. Albrecht BK & Williams RM (2003) A concise formal total synthesis of TMC-95A/B proteasome inhibitors. *Org Lett* 5(2):197-200.
262. Lin SN, *et al.* (2004) Total synthesis of TMC-95A and -B via a new reaction leading to Z-enamides. Some preliminary findings as to SAR. *J Am Chem Soc* 126(20):6347-6355.
263. Kaiser M, Groll M, Renner C, Huber R, & Moroder L (2002) The core structure of TMC-95A is a promising lead for reversible proteasome inhibition. *Angew Chem Int Edit* 41(5):780-+.

264. Kaiser M, *et al.* (2003) Synthesis of a TMC-95A ketomethylene analogue by cyclization via intramolecular Suzuki coupling. *Org Lett* 5(19):3435-3437.
265. Kaiser M, *et al.* (2004) TMC-95A analogues with endocyclic biphenyl ether group as proteasome inhibitors. *Chem Biodivers* 1(1):161-173.
266. Kaiser M, *et al.* (2004) Binding mode of TMC-95A analogues to eukaryotic 20S proteasome. *Chembiochem* 5(9):1256-1266.
267. Groll M, Gotz M, Kaiser M, Weyher E, & Moroder L (2006) TMC-95-based inhibitor design provides evidence for the catalytic versatility of the proteasome. *Chemistry & biology* 13(6):607-614.
268. Basse N, *et al.* (2007) Linear TMC-95-based proteasome inhibitors. *J Med Chem* 50(12):2842-2850.
269. Albrecht BK & Williams RM (2004) A concise, total synthesis of the TMC-95A/B proteasome inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 101(33):11949-11954.
270. Larsen MH, Biermann K, Tandberg S, Hsu T, & Jacobs WR, Jr. (2007) Genetic Manipulation of Mycobacterium tuberculosis. *Curr Protoc Microbiol* Chapter 10:Unit 10A 12.
271. Dereeper A, *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* 36:W465-W469.
272. Dereeper A, Audic S, Claverie JM, & Blanc G (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *Bmc Evol Biol* 10.
273. Leggett DS, Glickman MH, & Finley D (2005) Purification of proteasomes, proteasome subcomplexes, and proteasome-associated proteins from budding yeast. *Methods in molecular biology* 301:57-70.
274. Lebedev AA, *et al.* (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D* 68:431-440.
275. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.
276. Vagin AA, *et al.* (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D* 60:2184-2195.

277. Raja HA, Miller AN, Pearce CJ, & Oberlies NH (2017) Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *J Nat Prod* 80(3):756-770.
278. Crous PW & Groenewald JZ (2013) A phylogenetic re-evaluation of *Arthrinium*. *IMA Fungus* 4(1):133-154.
279. Federhen S (2015) Type material in the NCBI Taxonomy Database. *Nucleic acids research* 43(D1):D1086-D1098.
280. Consortium EP, *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799-816.
281. Kennedy DG, *et al.* (1994) Cobalt Vitamin-B-12 Deficiency Causes Accumulation of Odd-Numbered, Branched-Chain Fatty-Acids in the Tissues of Sheep. *Brit J Nutr* 71(1):67-76.
282. Kisselev AF, Akopian TN, Castillo V, & Goldberg AL (1999) Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Molecular cell* 4(3):395-402.
283. Jost M, *et al.* (2015) Structural basis for gene regulation by a B-12-dependent photoreceptor. *Nature* 526(7574):536-U167.
284. McIntosh JA, Donia MS, & Schmidt EW (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Natural product reports* 26(4):537-559.
285. Forneris CC & Seyedsayamdost MR (2018) In Vitro Reconstitution of OxyC Activity Enables Total Chemoenzymatic Syntheses of Vancomycin Aglycone Variants. *Angew Chem Int Edit* 57(27):8048-8052.
286. Mollo A, *et al.* (2017) P450 monooxygenase ComJ catalyses side chain phenolic cross-coupling during complestatin biosynthesis. *Rsc Adv* 7(56):35376-35384.
287. Thakur PB & Meshram HM (2014) "On water" catalyst-free, column chromatography-free and atom economical protocol for highly diastereoselective synthesis of novel class of 3-substituted, 3-hydroxy-2-oxindole scaffolds at room temperature. *Rsc Adv* 4(11):5343-5350.
288. Zhang YY, *et al.* (2017) Characterization of 2-Oxindole Forming Heme Enzyme MarE, Expanding the Functional Diversity of the Tryptophan Dioxygenase Superfamily. *J Am Chem Soc* 139(34):11887-11894.

289. Sussmuth RD & Mainz A (2017) Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Edit* 56(14):3770-3821.
290. Flack HD & Bernardinelli G (2008) The use of X-ray crystallography to determine absolute configuration. *Chirality* 20(5):681-690.