A genus-level comparative study of mycobacterial isocitrate metabolising enzymes



Stephanie Rhiannon Lovell-Read

Mycobacterial Metabolism and Antibiotic Research Laboratory

The Francis Crick Institute

&

Keble College

University of Oxford

Thesis for the degree of Doctor of Philosophy

Oxford-GSK-Crick Doctoral Training Programme in Chemical Biology

Michaelmas Term 2022

.

Sit with me and tell me how far you have travelled.

I am here.

Rachel Macy Stafford

Declaration of Authorship

I confirm that this thesis is entirely my own work, except where otherwise indicated, and that all material derived from other sources has been appropriately attributed. This work has not been submitted, either partially or in full, for any other degree of this University, nor for a qualification at any other institution.

Supervision

The work described in this thesis was carried out in the Mycobacterial Metabolism and Antibiotic Research Laboratory at the Francis Crick Institute, under the supervision of Dr Luiz Pedro Sorio de Carvalho. Additional supervision was provided by Professor Christopher J. Schofield (University of Oxford), Professor Akane Kawamura (University of Newcastle), and Dr Peter D. Craggs (GSK).

Funding

This project was funded by EPSRC, GSK, and the Francis Crick Institute, through the Oxford-GSK-Crick Doctoral Training Programme in Chemical Biology. Work in the Mycobacterial Metabolism and Antibiotic Research Laboratory is funded by the Francis Crick Institute, which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust.

Table of Contents

Declaration of Authorship ..... **i**

Supervision ..... **i**

Funding ..... **i**

Table of Contents ..... **ii**

Abstract ..... **vi**

Acknowledgements ..... **viii**

Intellectual and Material Contributions ..... **x**

Abbreviations ..... **xi**

Chapter 1. Introduction **1**

1.1. Tuberculosis: ancient disease, modern emergency 1

1.2. The classification of mycobacteria 14

1.3. Mycobacterial ecology and host adaptation 25

1.4. Genus-level comparison as a tool to understand mycobacterial phenotypes 32

1.5. The importance of isocitrate metabolism in Mtb 36

1.6. Isocitrate Lyases 46

1.7. Isocitrate Dehydrogenases 54

1.8. Control of flux through the isocitrate branch point 60

1.9. Host and non-host niches are nutritionally distinct 68

1.10. A genus-level comparative study of mycobacterial isocitrate metabolising enzymes 74

Chapter 2. Distribution of isocitrate metabolising enzymes in the genus *Mycobacterium* 78

2.1. Chapter introduction 78

2.2. Distribution of ICL1 and ICL2 in the genus *Mycobacterium* 79

2.3. Distribution of IDH1 and IDH2 in the genus *Mycobacterium* 80

2.4. Isocitrate metabolising enzymes are phylogenetically distributed in the genus *Mycobacterium* 81

2.5. Chapter conclusion 84

Chapter 3. Exploring the origin of the frameshift mutation that creates *rv1915—1916* from *aceA* 85

3.1. Chapter introduction 85

3.2. Location and predicted consequences of the mutation 89

3.3. The mutation is not a consequence of laboratory adaptation 91

3.4. How are *aceA* genotypes distributed in the MTBC? 92

3.5. Chapter conclusion 101

Chapter 4. Preparation and kinetic characterisation of the isocitrate metabolising enzymes from *Mycobacterium tuberculosis* CDC 1551 103

4.1. Chapter introduction 103

4.2. Isocitrate lyase 1 104

4.3. Isocitrate lyase 2 128

4.4. Isocitrate dehydrogenase 2 142

4.5. Isocitrate dehydrogenase 1 149

4.6. Kinetic comparison of the isocitrate branch point enzymes from Mtb CDC 1551 156

4.7. Chapter conclusion 158

Chapter 5. A multispecies comparative enzymology study of the mycobacterial isocitrate branch point 160

5.1. Chapter introduction 160

5.2. Selection of mycobacterial species for comparison to Mtb 161

5.3. Preparation of mycobacterial isocitrate metabolising enzymes 165

5.4. Kinetic characterisation and comparison of ICL1s 172

5.5. Kinetic characterisation and comparison of ICL2s 177

5.6. Kinetic characterisation and comparison of IDH2s 183

5.7. Kinetic characterisation and comparison of IDH1s 190

5.8. Comparison of the kinetic parameters of the isocitrate branch point in eleven mycobacterial species 193

5.9. Variation in isocitrate metabolism kinetics is not related to mycobacterial ecology 197

5.10. Chapter conclusion 201

Chapter 6. Conclusions and future work 202

6.1. Summary and conclusions....202

6.2. Questions arising and future directions....205

6.3. Final remarks....208

Chapter 7. Computational Methods 209

7.1. Identification of mycobacterial isocitrate lyases and isocitrate dehydrogenases 209

7.2. Identification of Mtb strains encoding frameshifted and single frame *aceA* using web-based BLAST searches 213

7.3. BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment of Mtb strains from the NCBI Microbial Genomes RefSeq database....216

Chapter 8. Experimental Procedures: Biochemistry 223

8.1. Mycobacterial Species and Strains....223

8.2. Construct Preparation....224

8.3. List of constructs 238

8.4. Protein Expression....275

8.5. Protein Purification 282

8.6. Identification of contaminant proteins by proteomics....294

8.7. Enzyme Kinetics....295

Chapter 9. References 306

Abstract

Mycobacterial diseases, most notably tuberculosis, have had a significant impact on human health and society throughout recorded history. Accordingly, mycobacteriologists have long sought to determine why some mycobacterial species (such as *Mycobacterium tuberculosis*) cause disease while others remain innocuous. Pathogenic mycobacteria have evolved molecular mechanisms to engage, exploit and elude their hosts. We propose that these mechanisms can be elucidated by studying mycobacteria at the genus level, comparing species adapted to host and non-host niches. To illustrate this approach, we present a genus-level comparative study of mycobacterial isocitrate metabolising enzymes, including the first detailed examination of isocitrate metabolism in non-tuberculous mycobacteria.

Isocitrate sits at a branch point in mycobacterial central carbon metabolism. Control of flux through this branch point is critical for *M. tuberculosis* during infection. The metabolic demand on the isocitrate branch point is carbon source dependent, meaning that mycobacteria adapted to nutritionally distinct ecological niches might exhibit differences in isocitrate metabolism. Herein, we combine bioinformatics and genus-level comparative enzymology to explore how the structure and kinetic function of the isocitrate branch point varies between mycobacterial species.

First, we establish that the composition of the isocitrate branch point varies phylogenetically across the genus *Mycobacterium*, and identify four mycobacterial isocitrate branch point structures. We then examine a fifth isocitrate branch point structure created by a frameshift mutation in some strains of *M. tuberculosis*. Next, we

prepare and kinetically characterise isocitrate metabolising enzymes from a representative subset of the genus *Mycobacterium*: *M. tuberculosis* CDC 1551 and eleven non-tuberculous mycobacterial species. Comparison of 36 enzymes reveals that mycobacterial species adapted to different ecological niches do not exhibit biologically significant differences in isocitrate metabolism kinetics. We therefore conclude that changes in the structure and kinetic function of the isocitrate branch point do not occur as part of mycobacterial niche adaptation.

Acknowledgements

Firstly, thank you to Luiz for welcoming me into the lab and trusting me with this project. It has been a privilege to work towards realising your scientific vision. I am particularly grateful for the flexibility which has enabled the completion of this thesis in often challenging circumstances. Wishing you and your family all the very best in Florida!

Chris, thank you for always encouraging me to consider the bigger picture. Pete, thank you for sharing your experiences and for always being there to talk things through, especially during lockdowns and the writing process. I have been incredibly fortunate to have you on my team. Akane, thank you for your commitment and support as both supervisor and programme director. Your unfailing warmth, patience, compassion and encouragement have meant more than I can say.

Words cannot express the depth of my gratitude to Acely, whose guidance, companionship and strength have been invaluable. Sincere thanks to the other past and present members of the Mycobacterial Metabolism and Antibiotic Research Laboratory: Agnese, Aleksandra, Angela, Ariana, Cathy, Cesira, Dan, Debbie, Dimitri, Fernanda, Holly, Hua, Jon, Magd, Mariana, Miha, Parise, Rob, Sam, Shohei, Suzanne, and Teresa. It has been an honour to laugh with and learn from you all. From the Crick, I would also like to thank Andy, Brenda, David, Frank, Louise, Mathew, Matthew, Pam, Rosangela, and the Gutierrez and Wilkinson labs.

From Oxford, I would particularly like to thank Nina, who has been a star throughout. Thank you also to Adam, Annette, Catherine, David, Gail, Grace, Katherine, Lorraine, Raphael, Sam, Stuart, and Toby. From GSK, I would like to thank David, Jake, and Rob.

In a list this long, I am certain to have left someone out; please know that I am no less grateful. Thank you to those whose names I do not know, but whose work has been appreciated every day – the Crick's Glasswash team come particularly to mind.

These pages recount one part of an unexpected journey. I acknowledge all that has been left unsaid. Finally, I thank you, the reader, for sitting with me now: I am here.

To my mother and sister, my eternal love.

Intellectual and Material Contributions

Thank you to Luiz Carvalho, Acely Garza-Garcia, Catherine Hubert, and Fernanda Subtil for the discussion of mycobacterial host adaptation which led to the development of **Figure 5** and the accompanying narrative. I gratefully acknowledge the contributions of the following individuals to this work, which have been referenced throughout:

Mycobacterial Metabolism and Antibiotic Research Laboratory
The Francis Crick Institute

Acely Garza-Garcia	Mycobacterial phylogenetic tree used in the preparation of **Figure 3**, **Figure 4**, **Figure 16**, **Figure 17**, and **Figure 18**.
	MTBC phylogenetic tree used in the preparation of **Figure 19**.
	Literature review of mycobacterial host adaptation and phylogenetic tree of the Mycobacterial Species Library, both used in the preparation of **Figure 65**.
	Glycerol stocks of pNIC-NTH and pNIC-CTHF in *E. coli* NEB5α
	Chemically competent *E. coli* BL21(DE3)Star
Fernanda Subtil	Mycobacterial colonies and cell lysates
Deborah Hunt	Mtb CDC 1551 genomic DNA
Miha Homsak	*M. smegmatis* colonies
Daniel Groom	TEV protease

Mass Spectrometry Proteomics Science Technology Platform
The Francis Crick Institute

David Frith	Identification of contaminant proteins by proteomics: trypsin digestion, mass spectrometry and initial data processing (as discussed in **Sections 4.2.1.7 and 8.6**).

University of Oxford

Raphael Reinbold	Constructs encoding *icd1* and *icd2* from Mtb

Abbreviations

3-NP	3-Nitropropionate
5-NIC	(2*R*,3*S*)-2-Hydroxy-3-(nitromethyl)succinic acid
A_{250}	Absorbance at 250 nm
A_{280}	Absorbance at 280 nm
aa	Amino acids
Ac-CoA	Acetyl coenzyme A
aceA	Gene encoding ICL2: *rv1915—1916* and its mycobacterial homologues
AceK	Isocitrate dehydrogenase kinase/phosphatase
Acn	Aconitase
ADP	Adenosine diphosphate
AF/ANI	Genome Alignment Fraction/Average Nucleotide Identity
AIEX	Anion exchange chromatography
AMP	Adenosine monophosphate
ArnA	Bifunctional polymyxin resistance protein
Asp	Aspartic acid
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BCG	Bacille Calmette-Guérin
BLAST	Basic Local Alignment Search Tool
blastn	Nucleotide-nucleotide BLAST
blastp	Protein-protein BLAST
cAMP	Cyclic adenosine monophosphate
CBD	Chitin binding domain
CBN	Conformal Bayesian Network
CIPP	Capture, Intermediate Purification and Polishing
CL	Containment level

Clp	Caseinolytic protease
CoA	Coenzyme A
COVID-19	Coronavirus Disease 2019
Crp	cAMP receptor protein, the cAMP-activated global transcriptional regulator
CV	Column volume
Cys	Cysteine
del	Deletion
DeoR	Deoxyribose operon repressor
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DnaK	Heat shock protein 70
DR	Direct Repeat
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EC	Enzyme Commission
EDTA	Ethylenediaminetetraacetic acid
FF	Fast Flow
FLAG	FLAG tag, DYKDDDDK
FPLC	Fast protein liquid chromatography
fs	Frameshift
GFP	Green fluorescent protein
GlmS	Glucosamine-fructose-6-phosphate aminotransferase
GlnR	Global nitrogen regulator
Gly	Glycine
GNAT	Gcn5-related *N*-acetyltransferase
GntR	Gluconate utilisation system GNT-I transcriptional repressor
GroL	Heat shock protein 60

HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	Hexahistidine tag, HHHHHH
HisB	Imidazole glycerol-phosphate dehydrogenase
HIV	Human immunodeficiency virus
HP	High Performance
HPLC	High performance liquid chromatography
icd1	Gene encoding IDH1: *rv3339c* and its mycobacterial homologues
icd2	Gene encoding IDH2: *rv0066c* and its mycobacterial homologues
ICL	Isocitrate lyase
icl	Gene encoding ICL1: *rv0467* and its mycobacterial homologues
ICL1	Isocitrate lyase 1
ICL1-GFP	ICL1 fused to GFP
ICL2	Isocitrate lyase 2
ICL2(T605)-His_6FLAG	Protein comprising aa 1—605 of Mtb CDC 1551 ICL2 and a *C*-terminal His_6FLAG tag
IDH	Isocitrate dehydrogenase
IDH1	Isocitrate dehydrogenase 1
IDH2	Isocitrate dehydrogenase 2
IDT	Integrated DNA Technologies
IEX	Ion exchange chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k_{cat}	Turnover number
k_{cat}/K_M	Catalytic efficiency
K_M	Michaelis constant
LAM	Latin American Mediterranean
LB	Luria-Bertani

LB-kanamycin	LB medium containing 60 µg/mL kanamycin
LDH	Lactate dehydrogenase
LDS-PAGE	Lithium dodecyl sulfate polyacrylamide gel electrophoresis
LpxA	UDP-*N*-acetylglucosamine acyltransferase
LrpI	Local regulatory protein of *icl* (Rv0465c, RamB)
MAC	*Mycobacterium avium* complex
MCL	Methylisocitrate lyase
MES	2-(*N*-Morpholino)ethanesulfonic acid
MOPS	3-(*N*-Morpholino)propanesulfonic acid
mRNA	Messenger RNA
MS	Malate synthase (GlcB)
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
Mycobacterium	*Mycobacterium* Lehmann and Neumann 1896, unless otherwise specified
n	Number of times purified
n.d.	Not determined
NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide, reduced
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced
NanK	*N*-Acetylmannosamine kinase
NCBI	National Center for Biotechnology Information
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
ns	Not specified
OD_{600}	Optical density at 600 nm
OmpA	Outer membrane protein A
OmpF	Outer membrane protein F

PAS	*para*-Aminosalicylic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pI	Isoelectric point
pICL1	Complemented with a plasmid containing *icl*
pICL2	Complemented with a plasmid containing *aceA*
PknG	Protein kinase G
POCP	Percentage of Conserved Proteins
PPD	Purified protein derivative of tuberculin
prpC	Methylcitrate synthase
prpD	Methylcitrate dehydratase
PrpR	Propionate regulator, Rv1129c
PTM	Post-translational modification
PurU	Formyltetrahydrofolate deformylase
qPCR	Quantitative PCR
qRT-PCR	Quantitative reverse transcription PCR
Q-TOF	Quadrupole time of flight
RamB	Regulator of acetate metabolism B (Rv0465c, LrpI)
RefSeq	Reference Sequence
RegX3	Sensory transduction protein RegX3
RNA	Ribonucleic acid
Rob	Right origin-binding protein
RplQ	50S Ribosomal protein
rRNA	Ribosomal RNA
RsuA	Pseudouridine synthase
SdhA	Succinate dehydrogenase flavoprotein subunit
SEC	Size exclusion chromatography
SenX3	Sensor-like histidine kinase SenX3

SIT	Spoligotype International Type
SNP	Single nucleotide polymorphism
SOC	Super Optimal broth with Catabolite repression
Spoligotyping	Spacer oligonucleotide typing
STPK	Serine/threonine phosphatase/kinase
TAE	Tris-Acetate-EDTA buffer
TB	Tuberculosis
tblastn	Protein-nucleotide 6-frame translation
TCA	Tricarboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
Ter	Termination
TEV	Tobacco etch virus
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
Tween 20	Polyoxyethylene (20) sorbitan monolaurate, Polysorbate 20
Tyr	Tyrosine
UDP	Uridine diphosphate
UniProtKB	UniProt Knowledgebase
UV	Ultraviolet
v_{max}	Maximal velocity
WHO	World Health Organisation
WT	Wild type
YraL	Ribosomal RNA small subunit methyltransferase I
Δ	Mutant in which the specified gene has been deleted
ε_{280}	Extinction coefficient at 280 nm
ε_{340}	Extinction coefficient at 340 nm
σ^B	Sigma factor B, Rv2710, SigB
σ^E	Sigma factor E, Rv1221, SigE

Chapter 1. Introduction

1.1. Tuberculosis: ancient disease, modern emergency

The infectious disease tuberculosis (TB) predates recorded history, yet remains a significant global health burden to this day.[1–4] TB is primarily a pulmonary disease, but a wide range of organs and body systems can be affected, including the lymph nodes, skeleton, central nervous system, adrenal glands, and peritoneum.[5–8] TB has killed more people than any other infectious disease, claiming over one billion lives in the last two hundred years alone.[9] The usual causative agent of human TB is *Mycobacterium tuberculosis* (Mtb), a highly host-adapted bacterial species for which humans are the only known reservoir.[10,11] More generally, 'TB' may refer to human or animal disease caused by any member of the *Mycobacterium tuberculosis* complex (MTBC), a set of closely related species to which Mtb belongs.[10]

1.1.1. TB and the ancient world: from spirituality to rational medicine

TB was present in ancient civilisations across five continents, and its effects on the lungs and skeleton have consequently been preserved in the palaeopathological record.[12,13] Medical documents from India and China provide the earliest surviving descriptions of TB, dated 3300 and 2300 years ago respectively.[14,15] Perhaps better known are the two references to TB in the Old Testament (Leviticus 26:16 and Deuteronomy 28:22), where it is referred to as *schachepheth*, a wasting disease to be sent as a plague by God.[16] Interestingly, although the Israelites were living in Egypt during the period recounted in these books, there are no clear reports of TB in ancient Egyptian papyri.[14,16] However, it is believed that only a small proportion of papyri have been recovered, and many translations are by necessity tentative.[17] The indigenous American civilisations,

meanwhile, did not keep written medical records, but evidence of TB can be found in their art.[18] Spinal deformities consistent with skeletal TB are depicted in many ancient Egyptian and indigenous American artworks, from figurines to paintings.[14,15,18] Of course, one cannot definitively diagnose the subjects of these artworks, as skeletal TB is not the only cause of such deformities.[18–21] Additionally, it is not known how accurately these artworks represent the human form – naturally, their artists would have been influenced by the stylistic conventions of their time.[18] While open to linguistic and cultural interpretation, the early written and artistic depictions of TB enrich our understanding of early attitudes to the illness. In particular, they highlight the entwining of ancient disease with spirituality and, in the case of ancient Egypt, with magic.[16,17]

A stronger pathological perspective on ancient TB can be obtained through the examination of human remains. Traditionally, the diagnosis of TB in skeletons is based on macroscopic observations of bone lesions, usually on vertebrae.[19,21–29] As less than 5% of untreated TB cases involve the skeleton, diagnostic criteria based solely on visual inspection significantly underestimate the incidence of TB in ancient populations.[20] Consequently, the high prevalence of ancient TB was not apparent until the 1990s, when the introduction of molecular diagnostic methods such as PCR made it possible to look for the presence of ancient pathogens directly.[13,20,23,27] Since then, ancient Mtb DNA has been amplified from many samples of bone and mummified tissue, both with and without visible pathology, offering insights into Mtb evolution and the epidemiological history of TB.[12,13,26,28–30]

The symptoms of pulmonary TB – cough, fever, night sweats, anorexia, and haemoptysis – were well recognised by the ancient Greeks, who named the disease *phthisis*.[5,15,31–33]

In contrast to many other ancient civilisations, the Greeks did not consider TB an act of divine retribution, instead identifying it as a disease that could be passed from one person to another.[15] There was, however, no consensus on the mechanism by which transmission could occur – Hippocrates believed that TB was hereditary, while Isocrates and Aristotle thought TB was an infectious disease.[15,34] The aetiology of TB was not elucidated until the 19th century, by which time it had divided scholars for over a thousand years.[14]

1.1.2. TB in Mediaeval Europe: scrofula and the royal touch

Infectious diseases such as leprosy, bubonic plague and syphilis were rife in mediaeval Europe.[35] Although the archaeological record indicates that pulmonary TB was widespread during this time, there are relatively few accounts of the disease in written medical records.[14,36] Tuberculous cervical lymphadenitis, the most common form of extrapulmonary TB, was also prevalent in the Middle Ages. It is usually associated with mediaeval Britain and France, where it was known as scrofula or 'king's evil' and thought to be a consequence of the sins of humanity.[37]

For over a thousand years, it was believed that English and French monarchs could cure scrofula through the laying on of hands.[14,38] Healing by touch was regarded as a supernatural power afforded to those with the Divine Right to rule, who could perform miracles and forgive sins as God's earthly representatives.[37,38] The practice of the 'royal touch' was started in 496 by King Clovis of France, and arrived in England during the reign of Edward the Confessor.[14,36] Over several hundred years, the original ritual evolved into a formal religious ceremony involving prayers and, in England, the presentation of a gold medallion ('touch-piece') to the patient.[14,35,36,38] During this time,

hundreds of thousands of patients were touched by monarchs.[14,37,38] Those patients who later went into spontaneous remission were believed miraculously cured, while those who did not nonetheless reported positive psychological effects from the experience.[37,38]

Use of the 'royal touch' began to dwindle following the Reformation, together with the belief in the Divine Right.[38] In 1712, Queen Anne became the last English monarch to treat scrofula in this manner.[14,15] The practice lingered in France until 1825, when Charles X performed the last healing touches during his coronation.[15,38] By this time, pulmonary TB and scrofula were recognised as different presentations of the same disease. The modern understanding of TB had begun to emerge.[14]

1.1.3. TB in the Age of Reason: romance and revelations

During the Scientific Renaissance, the revival of the ancient Greek theories of medicine provided a foundation for vast progress in the scientific understanding of disease.[14] On the subject of TB aetiology, Europe was divided, just as the Greek scholars had been over a thousand years before. The Northern European countries, where TB was endemic, subscribed to the Hippocratic view that TB was hereditary. In Southern Europe, where the incidence of TB was lower, TB was believed to be an infectious disease, and public health measures such as isolation were employed accordingly.[14,15,35,39] In 1735, the Italian Republic of Lucca (in present-day Tuscany) became the first state to declare TB a notifiable disease, with Spain following closely behind.[15,39]

By this time, doctors were developing a detailed clinical understanding of TB.[40] Published posthumously in 1679, Franciscus Sylvius' *Opera Medica* provided a precise anatomical and pathological description of the disease.[15] In 1720, the English physician Benjamin Marten became the first to propose that TB, or *consumption*, as it was known in Britain, was microbiological in nature.[14,15,34] Pott's disease (skeletal TB) and disseminated TB were described in detail in 1779 and 1810 respectively.[15] In 1819, René Laennec recognised pulmonary and extrapulmonary TB as different presentations of the same disease.[14,15] The unified disease was named *tuberculosis* by Johann Schönlein in 1834 and proven to be infectious by Jean-Antoine Villemin in 1865.[14,15,34,35]

As medical science advanced, so did the TB epidemic, fuelled by rapid urbanisation during the Industrial Revolution.[41] In the city slums, overcrowded housing with poor sanitation facilitated the spread of disease, such that by the 1830s a third of working-class adults in England died from TB.[15] At the peak of the epidemic, it is estimated that half the population suffered from active TB, a disease that was considered incurable until the 1850s.[15,39] Society responded to this overwhelming and seemingly hopeless epidemic by romanticising its victims.[14] In opera and literature, the physical features of TB patients – their thin, pale faces and sunken eyes burning with fever – were transformed into romantic beauty and passion.[40] It was believed that such passion, together with creativity, genius and euphoria, was born from the slow wasting of the body.[40] Termed *spes phthisica*, this perceived side effect of TB became intimately associated with artists, poets and composers, to the point where suffering from TB was almost a prerequisite for genius.[39,40]

The human understanding of TB entered a new chapter on 24th March 1882, when Robert Koch stood before the Berlin Physiological Society to give his now famous lecture, *Die Aetiologie der Tuberculose*.[14] He had isolated *Tuberkelvirus*, the causative agent of TB, and in so doing had demonstrated conclusively the infectious nature of the disease.[34,35] *Tuberkelvirus* was known by various binary and colloquial names until 1896, when it received the name we still use today: *Mycobacterium tuberculosis*.[42]

1.1.4. Controlling TB: public health interventions

The discovery of Mtb earned Koch the 1905 Nobel Prize in Physiology or Medicine and paved the way for new strategies for the prevention and treatment of TB.[14,34] Despite the advances in pathological understanding, the fundamentals of TB treatment had not changed substantially since the Roman era, with a rich diet, bed rest and a change of climate being consistently recommended.[35,39] These principles formed the basis of the sanatorium cure for TB, which was pioneered by Hermann Brehmer in the 1850s.[15,35,43] By the end of the century, surgical interventions had become widespread, with common procedures including collapse of the affected lung to give it time to heal and excision of diseased parts of the organ.[14,35,40]

Meanwhile, new pharmacotherapies were being developed for the treatment of TB. In 1890, Koch proposed that *old tuberculin*, the concentrated culture filtrate of Mtb grown in medium containing glycerol, could be used as a cure.[35,44,45] While Koch's results indicated that old tuberculin inhibited Mtb growth *in vitro* and in guinea pigs, it was quickly found to be ineffective in clinical practice.[35,44] Another experimental therapy was the inhalation of antiseptics and disinfectants such as igazol, a mixture of formaldehyde, trioxane (a cyclic trimer of formaldehyde), and iodine-containing compounds.[35]

Although anecdotal evidence attributed positive benefits to igazol, large-scale trials indicated 'no favourable results of any kind'.[46,47]

By 1900, scientific acceptance of the infectious nature of TB had caused the disease to lose its romantic connotations, instead becoming a distasteful attribute of poverty.[40] During the following decades, hospitals and sanatoria continued to provide care for TB patients, while a range of public health measures were employed to reduce transmission in the community.[40,46,48] Spitting was banned in many public places and the use of Japanese paper handkerchiefs – the forerunners of modern facial tissues – was encouraged.[40,48] In 1912, TB became a notifiable disease in the UK, more than 150 years after it had been declared as such in Italy and Spain.[15,39,41] Public health efforts to control TB were assisted by the development of a diagnostic test for pre-symptomatic TB – the tuberculin skin test.[35] While old tuberculin had failed as a treatment for TB, it was known that intra-dermal injection produced a skin reaction in TB patients.[34] This reaction was the basis for the first tuberculin skin tests, introduced by Clemens von Pirquet and Charles Mantoux in 1907.[35,44] In the 1930s, the reliability of the tuberculin skin test was improved through the use of Florence Siebert's purified protein derivative of tuberculin (PPD), variants of which remain in use today.[44]

Another important step towards controlling TB was taken by Albert Calmette and Camille Guérin, who developed the Bacille Calmette-Guérin (BCG) vaccine.[34,35] BCG is an attenuated strain of *Mycobacterium bovis*, which causes TB in cattle.[49] While researching the pathogenesis of bovine TB, Calmette and Guérin found that *M. bovis* cultured on potato slices soaked in glycerol and ox bile became less virulent in guinea pigs and cattle after only a few months.[49–51] This observation inspired them to attenuate

the strain further in the hope of developing a human vaccine.[49] Between 1908 and 1921, Calmette and Guérin repeatedly sub-cultured *M. bovis* on their potato-based medium. After thirteen years and 231 passages, they had produced the BCG strain.[49,51] BCG was first tested in humans in 1921, but mass vaccination did not begin until after World War II, when the World Health Organisation (WHO) began to organise BCG vaccination campaigns.[35,51] By 1974, approximately 1.5 billion people had received the BCG vaccine.[51]

1.1.5. Conquering TB: the antibiotic era

The isolation of mycophenolic acid, the first antibiotic, in 1893 marked the beginning of a new era in the fight against bacterial infections, although it would take another twenty years for the impact of this discovery to be realised.[52,53] Meanwhile, Paul Ehrlich formulated the concept of antimicrobial chemotherapy and began his now famous search for antibacterial compounds among the anilines produced by the German dye industry.[53,54] The result was arsphenamine, the first synthetic antibiotic, which entered mass production in 1910.[53] Due to the success of arsphenamine, dyes continued to be investigated for antimicrobial properties, resulting in the discovery of the sulfonamide antibiotics in the 1930s.[53,55]

Alexander Fleming's serendipitous observation of the bactericidal activity of a *Penicillium* mould (now believed to have been *P. rubens*) in 1928 went on to revolutionise antimicrobial drug discovery.[53,55–57] Penicillin, a fungal secondary metabolite, was eventually purified by Ernst Chain, Howard Florey, and Norman Heatley in 1940 and went on to save hundreds of millions of lives.[55,56,58,59] Importantly, the work of Fleming, Chain, Florey, and Heatley inspired other scientists to investigate soil

microorganisms as sources of antibiotics, most notably Selman Waksman.[58,60] Having established the genus *Streptomyces* as a promising source of antimicrobial compounds, Waksman's laboratory began to search for natural products that were active against Mtb.[53,55,61] In 1944, they reported the isolation of streptomycin, a compound they had identified as active against a range of pathogenic bacteria including Mtb.[62] Streptomycin rapidly progressed to human studies, and its efficacy as a treatment for pulmonary TB became the subject of the first published randomised controlled clinical trial, carried out by the Medical Research Council in 1946.[55,60]

The first synthetic TB drug, *para*-aminosalicylic acid (PAS), was also discovered in 1944, just a few months after streptomycin.[35] One of around fifty benzoic acid derivatives synthesised by Jörgen Lehmann, PAS displayed bacteriostatic activity against Mtb *in vitro* and in animal models.[60,63] PAS was marketed as a treatment for human TB from 1946, and from 1948 was used in combination with streptomycin to reduce the incidence of streptomycin resistance.[35,63,64] In 1952, a third drug was added to the combination: isoniazid.[35,55] Although isoniazid (isonicotinic acid hydrazide) was first synthesised in 1912, it was not tested against Mtb until 1951, when three laboratories independently identified it as highly active in animal models.[35,65] With the introduction of isoniazid, TB could finally be considered a treatable disease – around 90% of patients were cured after 24 months of combination therapy.[35,55,66]

Over the next decade, a range of other compounds were found to be active against TB, including pyrazinamide, ethambutol, cycloserine, kanamycin, and ethionamide.[35,67–69] The new drugs were tested in various combinations and adopted as second line treatments for TB.[68,69] Isoniazid-streptomycin-PAS 'triple therapy' remained the

preferred first line treatment for TB until the late 1960s, when PAS was replaced with ethambutol, reducing the duration of standard TB chemotherapy to 18 months.[35,67,68,70] In the early 1970s, the inclusion of rifampicin in TB drug regimens reduced the length of treatment further, to nine months.[35,67,70] First prepared in 1965, rifampicin is a derivative of the natural product rifamycin SV which displays greater antituberculosis activity and oral bioavailability than the parent compound.[35,60,71] The development of rifampicin brought about the most significant change in TB treatment since the introduction of isoniazid some twenty years earlier.[60] Later in the 1970s, the addition of pyrazinamide shortened the duration of treatment to six months, creating the 'short-course' TB chemotherapy that is still used today.[60,72] Fundamentally unchanged since 1980, modern first line chemotherapy for drug-susceptible TB comprises six months of treatment with four antibiotics: rifampicin, isoniazid, ethambutol and pyrazinamide. All four antibiotics are taken during the first two months, with rifampicin and isoniazid being continued for a further four months.[73,74]

The success of short-course chemotherapy in Europe and the United States contributed to a decline in funding for TB drug research during the 1970s and 1980s.[35,75] As a consequence, scientists turned their attention to neglected tropical diseases and non-communicable conditions.[76] Despite the fact that TB remained a significant problem in developing countries, the WHO was confident that its goal of 'alleviating human suffering and preventing death and disability from tuberculosis' could be achieved by the year 2000.[77,78] In the eyes of many, the battle against TB had already been won.[79]

1.1.6. TB today: a global health emergency

The attention of economically developed nations was drawn back to TB in the late 1980s, when countries where TB had been in decline began to report increases in the incidence of the disease.[80] Two significant problems quickly became apparent: a TB-HIV co-epidemic was in progress, and antibiotic resistance was on the rise.[81,82] The extent of the global TB epidemic was revealed in 1989, when a WHO study estimated that one third of the world's population were infected with Mtb.[82] In response, the WHO TB Programme was revived and new goals were established for TB research and control.[79,81,82] On 23rd April 1993, TB became the first infectious disease to be declared a global health emergency by the WHO, making headline news across the globe.[79]

Almost thirty years later, although progress has been made, the TB emergency has yet to be resolved.[74] In 2019, 1.2 million HIV-negative people died from TB, making it the thirteenth most common cause of death worldwide and the leading cause of death due to an infectious agent.[83] The COVID-19 pandemic triggered a rise in TB deaths for the first time since 2005; 1.3 and 1.4 million HIV-negative people died from TB in 2020 and 2021 respectively.[84,85] This rise has been attributed to the disruption of services for TB diagnosis and treatment. A longer-term rise in TB incidence and mortality may result from the negative impact of the pandemic on TB determinants such as undernourishment and poverty.[84,85] While TB is presently the second most common cause of death due to an infectious disease (behind COVID-19), TB deaths are expected to overtake those due to COVID-19 in the near future.[84,85]

Currently, an estimated 10 million people become ill with TB each year, with pulmonary TB accounting for 85% of cases.[4,74,84] Men are disproportionately affected by TB – adult

males (aged 15+) comprise 37% of the world's population, but 58% of TB patients.[4,86–88] Although the lower incidence of TB among women has historically been attributed to sociocultural factors, it is likely that sex-related differences in immune function also contribute to the observed imbalance in TB cases.[88,89] The prevalence of TB also varies widely between countries. In 2020, 86% of all TB cases occurred in the 30 countries designated as having a high TB burden by the WHO, the majority of which are located in the WHO regions of South-East Asia, Africa and the Western Pacific.[84] Geographic variation is also seen in the incidence of drug resistant TB, which is most common in countries of the former Soviet Union. The global incidence of drug resistant TB has remained relatively constant over the last ten years, with 3—4% of new TB cases and 18—21% of relapsed TB cases exhibiting rifampicin or multidrug resistance.[84]

The WHO recommendations for the treatment of rifampicin- and multidrug-resistant TB were updated in 2022.[90] This change followed the conclusion of various trials aimed at identifying shorter, safer, and more efficacious regimens for the treatment of drug-resistant TB.[90–93] It is now suggested that a 6-month course of bedaquiline, pretomanid, linezolid, and moxifloxacin (known as the BPaLM regimen) should be administered in preference to the older 9-month and 18-month regimens.[90] Linezolid and moxifloxacin have been repurposed for the treatment of drug-resistant TB, while bedaquiline and pretomanid are two of just three new antituberculosis drugs to have been approved in the last forty years.[94,95] Acquired resistance to each of the drugs in the BPaLM regimen has been observed in the clinic, and the first instance of acquired resistance to all four drugs was recently reported. The reported case became resistant to all known antituberculosis drugs and the patient ultimately died from another bacterial

infection.[96] Currently, drug-resistant TB accounts for around one quarter of deaths associated with antimicrobial resistance; it is projected that this will still be the case in 2050.[97]

Ending the global TB epidemic by 2030 is part of Target 3.3 in the United Nations Sustainable Development Goals. The WHO's End TB Strategy quantifies this goal as 'a 90% reduction in the number of TB deaths and an 80% reduction in the TB incidence rate (new and relapse cases per 100,000 population per year) compared with levels in 2015'.[4,98,99] Currently, most of the world is not on track to reach these targets, and the COVID-19 pandemic has already set back progress towards ending TB by several years.[84] To move towards ending TB, we must increase the accurate diagnosis and reporting of TB cases, improve access and adherence to treatment, and develop new interventions for drug resistant TB.[4,100] We must also find new ways to identify and treat the estimated 1.7 billion people who are already infected with Mtb, and are therefore at risk.[4,66,101]

Although infected, most of these people are not currently ill with TB, and cannot pass the disease on to others.[4,101] They have latent TB – a persistent Mtb infection under immune-mediated control with no clinical signs of active disease.[66,101,102] Since Mtb is still present, active disease can develop if the immune system becomes unable to keep the infection in check.[102] The progression from latent to active disease, known as reactivation TB, most commonly occurs between two and five years after the initial infection.[103] Over the course of a lifetime, 5—15% of latently infected individuals will develop reactivation TB, a risk that rises to 5—10% per year for people living with HIV.[101–103] People with latent TB are therefore a reservoir for future active TB cases.[104]

Consequently, the use of preventive treatment to reduce the probability of disease progression will be an essential part of ending the TB epidemic.[4,104]

Importantly, the prevalence of latent TB highlights the age and complexity of the interactions that occur between Mtb and the human host.[11,22,66,105] A major reason for the success of Mtb as a pathogen is its ability to evade elimination by the immune system and persist in the human body for decades without causing disease.[66,102,106,107] The high level of host adaptation exhibited by Mtb has resulted in a profound impact on human health and society throughout recorded history. We know, however, that our relationship with Mtb considerably predates the written word – studies of human remains from the Neolithic period indicate that humans and Mtb had already undergone extensive host-pathogen coevolution by this time.[105] The origins of Mtb host adaptation lie somewhere between 35,000 and 70,000 years in the past, with a speciation event in the genus *Mycobacterium*.[108,109]

1.2. The classification of mycobacteria

1.2.1. The taxonomic position of mycobacteria

Mycobacterium is a genus in Actinobacteria, one of the most ancient bacterial phyla.[110,111] Arising in the anoxic ocean approximately 2.7 billion years ago, actinobacteria were among the first organisms to colonise land.[112,113] Modern actinobacteria are phylogenetically and physiologically diverse, and members of the phylum can be found in almost all ecosystems on Earth.[112] The taxonomic position of mycobacteria within Actinobacteria is shown in **Figure 1**. *Mycobacterium* Lehmann and Neumann 1896 is the nomenclatural type of the order Mycobacteriales Janke 1924, and

is the only genus in the family *Mycobacteriaceae* Chester 1897.[a] The most closely related genus is *Hoyosella* Jurado *et al.* 2009, which belongs to the family *Hoyosellaceae* Salam *et al.* 2020.[110,114,115]

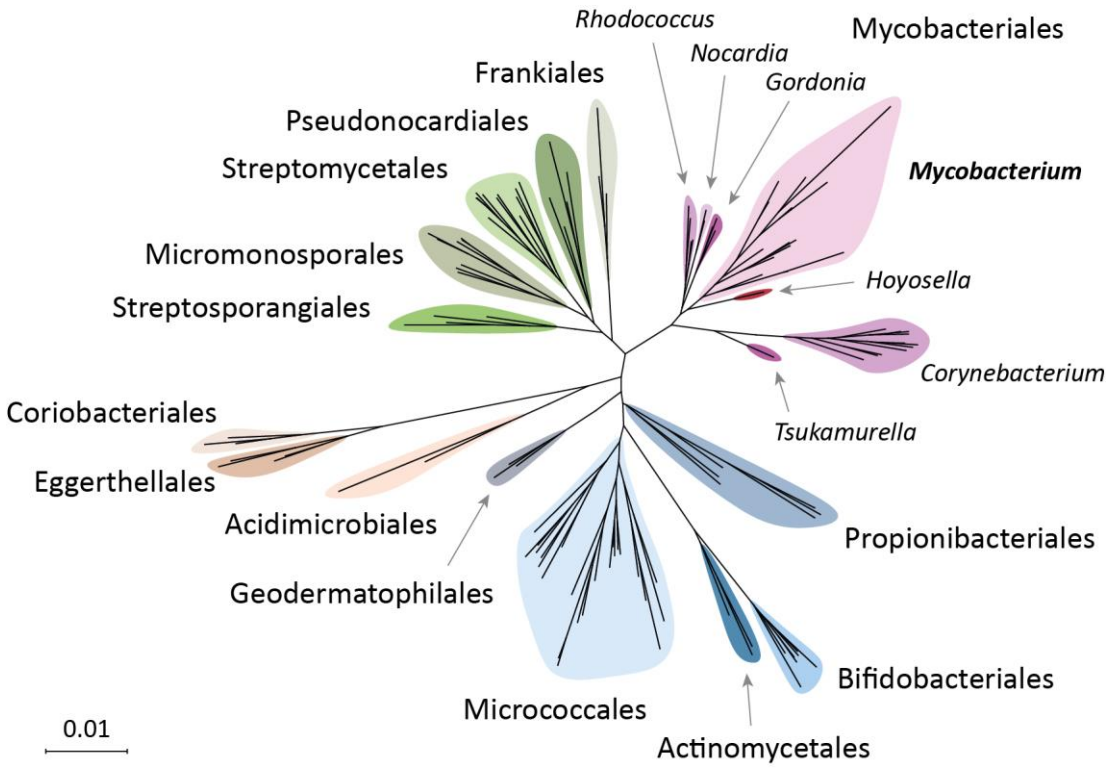


Figure 1: Actinobacterial phylogenetic tree.
Mycobacterium is a genus in Actinobacteria, an ancient and diverse bacterial phylum. The type genus of the order Mycobacteriales, *Mycobacterium* is most closely related to *Hoyosella*. Each clade is labelled with the corresponding actinobacterial order. Genera within Mycobacteriales are given in italics. The phylogenetic tree was generated using the TreeVector function on Superfamily 1.75 and displays the actinobacterial clades of the maximum likelihood tree calculated using protein domain architecture data for all genomes in Superfamily. The scale indicates the mean number of nucleotide substitutions per site. Annotations and colours were added in Adobe Illustrator to aid visual separation of the major clades.

[a] *Mycobacterium* Lehmann and Neumann 1896 was included in Corynebacteriales Goodfellow and Jones 2015, an order validly published by inclusion in Validation List 164. Consequently, recent literature (including the 2015 edition of Bergey's Manual of Systematics of Archaea and Bacteria) gives *Mycobacterium* Lehmann and Neumann 1896 as belonging to the order Corynebacteriales Goodfellow and Jones 2015.[122,545] However, since *Mycobacterium* Lehmann and Neumann 1896 is the nomenclatural type of Mycobacteriales Janke 1924, Corynebacteriales Goodfellow and Jones 2015 is taxonomically illegitimate under Rule 51b (1) of the International Code of Nomenclature of Prokaryotes (the Prokaryotic Code). Under Rule 54, the order containing *Mycobacterium* Lehmann and Neumann 1896 should be designated Mycobacteriales Janke 1924, as this is the oldest taxonomically legitimate name.[546,547] Accordingly, an emended description of Mycobacteriales Janke 1924 has been published that accommodates those taxa assigned to Corynebacteriales Goodfellow and Jones 2015.[548]

When first proposed, the genus *Mycobacterium* had three members: Mtb, *M. leprae* (the causative agent of leprosy), and *M. avium* (the causative agent of avian TB).[42] Over the next thirty years, a further nine species were added to the genus, including *M. bovis*, from which the BCG vaccine strain was derived (**Section 1.1.4**), and *M. marinum*, which causes TB in saltwater fish.[116,117] Mycobacterial discovery began to accelerate in the 1960s as new biochemical techniques were developed for species identification, and by 1990, a total of 64 mycobacterial species had been described.[116–118] The introduction of mycolic acid profiling (**Section 1.2.2.2**) and 16S ribosomal RNA sequencing (**Section 1.2.3.1**) in the early 1990s triggered an explosion in mycobacterial taxonomy, with the genus doubling in size by 2007.[116,119] As of August 2022, the names of 195 mycobacterial species have been validly published; around 50 additional species do not have formal standing in nomenclature.[120,121] Mtb is the type species of the genus and, due to the historical and continuing impact of TB on human health (**Section 1.1**), is undoubtedly its best-known member.[122]

1.2.2. Phenotypic approaches to mycobacterial classification

1.2.2.1. Runyon classification of mycobacteria: growth rate and pigmentation

Until the 1930s, Mtb, *M. leprae* and *M. bovis* were the only mycobacterial species thought to infect humans. This view changed when *M. marinum* was identified as the causative agent of swimming pool granuloma – the development of granulomatous lesions at the site of a skin abrasion acquired in water. Contemporaneously, cases of pulmonary disease resembling TB, but caused by species such as *M. avium* and *M. kansasii*, were reported. It therefore became desirable to classify mycobacterial species on the basis of their clinical relevance.[116]

Ernest Runyon proposed such a classification in 1959, dividing the non-tuberculous mycobacteria into four groups based on their growth rate and pigmentation (**Table 1**).[123] Mycobacteria are termed fast or slow growers based on the time taken to form visible colonies on solid medium.[b] Slow growers (Runyon groups I—III) require more than seven days to produce colonies, while fast growers (Runyon group IV) form clearly visible colonies within seven days. Runyon further categorised the slow growers based on their ability to produce carotenoid pigments and the response of this pigmentation to light exposure. Scotochromogenic mycobacteria (Runyon group II) are always pigmented, while photochromogenic species (Runyon group I) produce pigment only after exposure to light. Nonchromogenic slow growers (Runyon group III) do not produce carotenoid pigments, and so form non-pigmented, pale yellow, buff or tan colonies.[117,124–126]

Table 1: Examples of mycobacterial species belonging to each Runyon group.[117,123–125]

Runyon group	Growth rate	Example species
I: Photochromogens	Slow	*M. kansasii*, *M. marinum*, *M. simiae*
II: Scotochromogens	Slow	*M. scrofulaceum*, *M. gordonae*, *M. xenopi*, *M. flavescens*
III: Nonchromogens	Slow	*M. avium*, *M. intracellulare*, *M. paratuberculosis*, *M. malmoense*, *M. nonchromogenicum*, *M. terrae*
IV: Fast growers	Fast	*M. abscessus*, *M. chelonae*, *M. fortuitum*, *M. smegmatis*

[b] Mycobacterial growth rates vary between media.[549] The traditional media for the classification of mycobacterial growth rates include Löwenstein-Jensen medium and Sauton agar.[550–552]

When first developed, Runyon classification was of some use when considering the likely clinical relevance of mycobacterial species, as most of the opportunistic mycobacteria known at the time fell into Groups I and III.[118] Today, however, species from all four Runyon groups are known to cause disease in humans, and the system is no longer used for this purpose.[117,127] Over the years, at least fifty other phenotypic features of mycobacteria have been used to identify, characterise and taxonomically describe members of the genus.[116] As mentioned above (**Section 1.2.1**), an important method in modern mycobacterial chemotaxonomy is the characterisation of mycolic acids, lipids found in the cell wall.[128]

1.2.2.2. Mycolic acid profiling: characterisation of cell wall lipids

Mycobacterial cells are aerobic to microaerophilic rods, 0.2—0.6 µm wide and 1—10 µm long, which are non-motile and do not form spores.[122] Perhaps their most notable feature is the complex and distinctive cell envelope.[125,129,130] Rich in lipids, the mycobacterial cell wall is thick and hydrophobic, conferring resistance to many antibiotics and tolerance to chemical stressors.[125,131–133] The cell wall composition also causes mycobacteria to be acid-fast, meaning that once mycobacterial cells have been stained, they are resistant to decolourization by acids.[125,134]

Mycolic acids (2*R*-alkyl, 3*R*-hydroxy long chain fatty acids) are a major constituent of the cell walls of mycobacteria and other species in the order Mycobacteriales. Mycobacterial mycolic acids are among the largest, containing between 60 and 90 carbon atoms in total, and are unique in that their 3*R*-hydroxy chains usually incorporate oxygen-containing functional groups, such as ketones, methoxy groups, hydroxyl groups, epoxides, and wax esters.[135] As the mycolic acid composition of the cell wall

varies between species, characterisation of mycolic acid profiles using techniques such as high-performance liquid chromatography (HPLC) can be used for the species-level identification of mycobacteria.[119,128] To delineate the evolutionary relationships between mycobacterial species, however, we must turn to genome-based methods.[136]

1.2.3. Genome-based approaches to mycobacterial classification

1.2.3.1. 16S ribosomal RNA: the first genetic classification of mycobacteria

First proposed as a taxonomic marker in the 1970s, ribosomal RNA (rRNA) became fundamental to bacterial classification in the 1990s, when it became possible to sequence the genes encoding rRNAs in their entireties.[137–139] Today, the most sequenced rRNA is 16S, part of the small (30S) subunit of bacterial and archaeal ribosomes. Different areas of the 16S sequence evolve at different rates, meaning that it is phylogenetically informative across the majority of taxonomic levels, from domains to genera.[140] Phylogenetic resolution of species is not always possible, however, as closely related species can have identical 16S sequences.[141]

16S sequencing was first applied to mycobacterial phylogenetics in 1990, when a study of 19 species identified that those categorised as fast growers (Runyon group IV) clustered together. The slow-growing mycobacteria (Runyon groups I—III) formed a second cluster, but there was no further correlation between Runyon grouping and phylogenetic position.[142] Today, the All-Species Living Tree Project database contains over 200 mycobacterial 16S sequences.[143] **Figure 2** shows a phylogenetic tree constructed using a subset of these sequences. As in the early trees, mycobacteria cluster into two distinct groups, which can be broadly labelled as fast and slow

growers.[118,144] However, the position of a species in the 16S phylogeny is not always indicative of its growth rate under experimental conditions.[145]

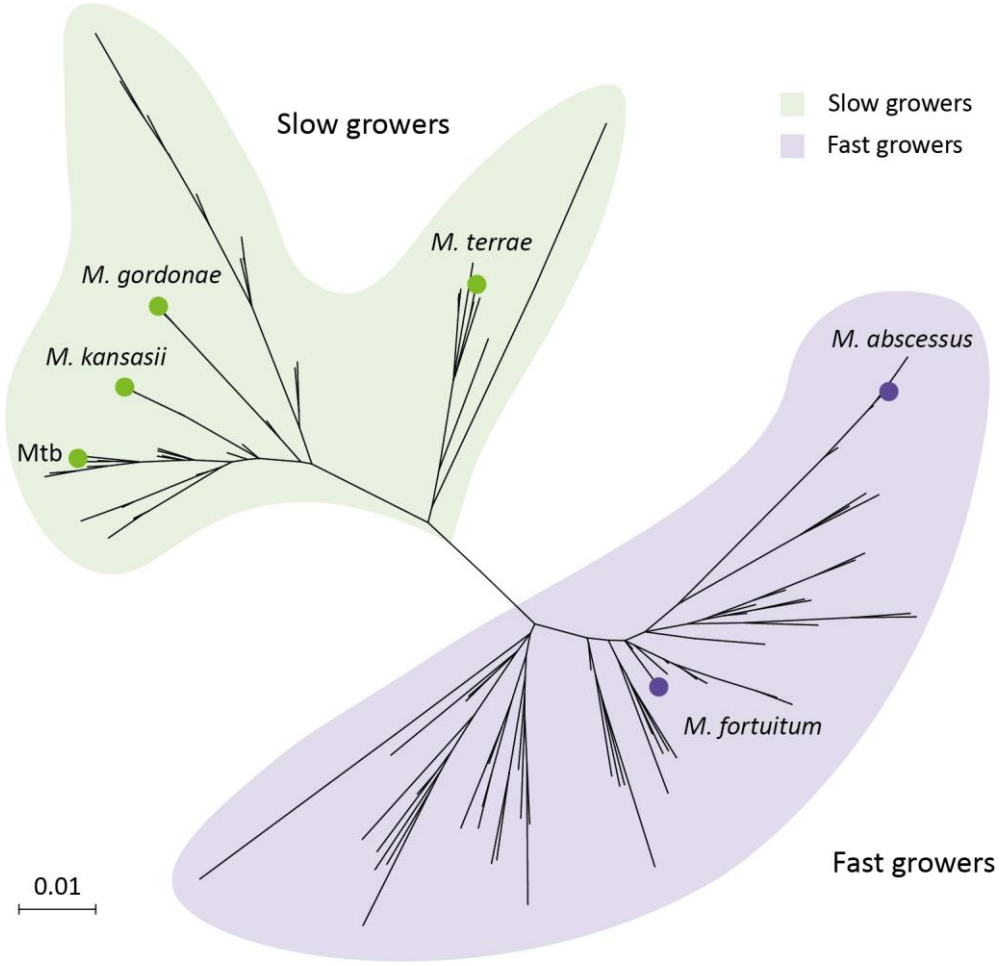


Figure 2: Mycobacterial phylogenetic tree constructed using 16S sequences.
Fast-growing (shaded purple) and slow-growing (shaded green) mycobacterial species appear as distinct groups in phylogenies constructed using 16S sequences. Mycobacterial phylogenetic tree derived from the September 2021 release of the All Species Living Tree Project, which includes 16S sequences for 203 mycobacterial species.[141,143] Selected species of interest are marked with circles for orientation purposes. The scale indicates the mean number of nucleotide substitutions per site.

1.2.3.2. Modern mycobacterial phylogenetics: concatenated sequence alignment

The inability of 16S sequencing to distinguish between some closely related species (**Section 1.2.3.1**) was apparent even in the earliest mycobacterial phylogenies generated using this method.[142,146] Consequently, the use of multiple taxonomic markers is required to construct robust phylogenetic trees for the genus. Concatenated

sequence alignments, for example of multiple housekeeping genes, provide substantially more reliable trees, and their use has become commonplace in modern mycobacterial phylogenetics.[145,147–149] Trees constructed using concatenated sequence alignments (such as that shown in **Figure 3**) show the same clear separation between the slow and fast growers as 16S phylogenies, and indicate that the last common ancestor of mycobacteria would have been a fast grower closely related to the extant *M. abscessus* complex.[150–153]

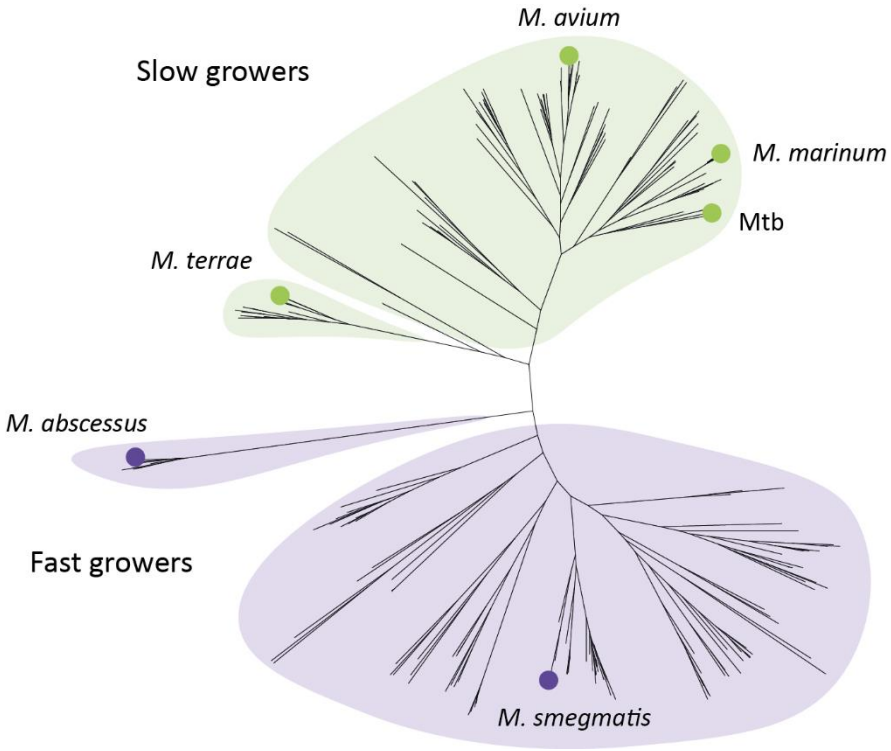


Figure 3: Mycobacterial phylogenetic tree constructed using concatenated sequence alignment.
Trees constructed using concatenated sequence alignment show the same clear separation between the slow and fast growers as trees constructed using 16S sequences. Mycobacterial phylogenetic tree constructed by Acely Garza-Garcia.

1.2.3.3. Division and reconstitution of the genus: an ongoing controversy

Five distinct clades, shown in **Figure 4**, appear repeatedly in phylogenetic and phylogenomic analyses of mycobacteria.[150,153] In 2018, Gupta *et al.* proposed that *Mycobacterium* Lehmann and Neumann 1896 should accordingly be split into five

genera: an emended genus *Mycobacterium* (*Tuberculosis-Simiae* clade), *Mycobacteroides* gen. nov. (*Abscessus-Chelonae* clade), *Mycolicibacterium* gen. nov. (*Fortuitum-Vaccae* clade), *Mycolicibacillus* gen. nov. (*Triviale* clade), and *Mycolicibacter* gen. nov. (*Terrae* clade).[150] These genera were validly published, and have consequently been adopted by public databases such as GenBank.[154] However, the division of the genus has proved contentious among the mycobacteriology community, with criticism falling along two main lines.

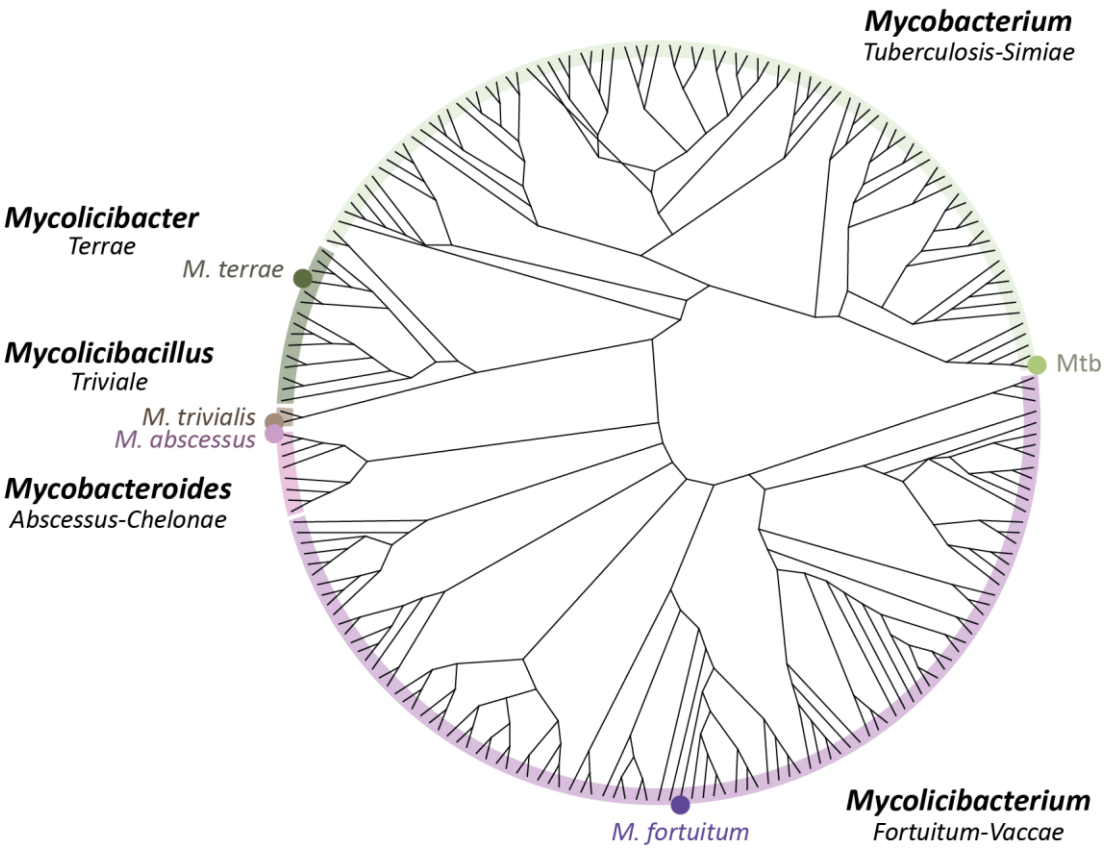


Figure 4: Proposed division of the genus *Mycobacterium*.
Gupta *et al*. proposed that the five main mycobacterial clades should be considered separate genera. The type species of the new genera are marked with circles (*Mycobacterium*, Mtb; *Mycolicibacterium*, *M. fortuitum*, *Mycobacteroides*, *M. abscessus*; *Mycolicibacillus, M. trivialis*; *Mycolicibacter, M. terrae*).[150] Mycobacterial phylogenetic tree constructed by Acely Garza-Garcia.

Firstly, there are concerns that the renaming of a large number of mycobacterial species could create confusion in clinical settings. Gupta *et al.* argue that the majority of

clinically relevant species remain within the emended genus *Mycobacterium*, citing the separation of pathogenic and non-pathogenic species as a benefit of the division.[150] However, the important emerging pathogens of the *Abscessus-Chelonae* clade (most notably *M. abscessus*) are situated outside the emended genus, belonging instead to the new genus *Mycobacteroides*. Further, the rarer mycobacterial infections caused by other species assigned to new genera (for example members of the *Fortuitum-Vaccae* clade) are already challenging to identify and treat correctly. Since considering non-tuberculous mycobacterial infections collectively rather than as distinct entities has led to more effective care and better outcomes for patients, creating distance between these species by assigning them to separate genera could hinder future clinical progress in this area.[154]

Secondly, the question arises of whether or not the differences between clades noted by Gupta *et al*. are sufficient to justify the creation of separate genera. Indeed, an earlier proposal to split *Mycobacterium* Lehmann and Neumann 1896 into two subgenera, *Mycobacterium* (slow growers) and *Mycomycobacterium* (fast growers), was rejected on the grounds that there were insufficient differences between the two groups.[118,155] Recently, Meehan *et al*. considered the division of the genus in the context of current genome-based approaches to the demarcation of bacterial genera. Their analysis of mycobacterial genomes using two indices of relatedness, the Percentage of Conserved Proteins (POCP) and Genome Alignment Fraction/Average Nucleotide Identity (AF/ANI) methods, found the original genus *Mycobacterium* to be more appropriate for the classification of mycobacterial species than the new genera proposed by Gupta *et al*. As a result, they proposed the reconstitution of *Mycobacterium* Lehmann and Neumann

1896.[156] At present, the two sets of names exist as heterotypic synonyms, meaning that either may be correctly used.[154] In line with the arguments outlined above, our preference is to consider all mycobacterial species as a single genus. For the rest of this thesis, therefore, *Mycobacterium* should be taken to mean *Mycobacterium* Lehmann and Neumann 1896, unless explicitly stated otherwise.

1.2.4. Clinical relevance remains of interest in mycobacterial classification

Since the inception of the genus *Mycobacterium* in 1896, numerous phenotypic, phylogenetic and phylogenomic techniques have been used to describe and classify its members.[116] Considering the history of mycobacterial classification, clinical relevance emerges as a repeated theme, appearing as both underlying motivation and end goal. Predicting the clinical relevance of mycobacterial species was the explicit aim of the Runyon classification (**Section 1.2.2.1**), while the development of other phenotypic methods for the classification of mycobacteria was driven in part by the need to reliably identify pathogens at the species level (**Section 1.2.2.2**).[118] Seventy years after Runyon first sought to categorise mycobacterial species, clinical relevance remains an important consideration in mycobacterial classification, appearing in arguments for both the division and the reconstitution of the genus (**Section 1.2.3.3**).[150,154]

Although methods for the study and classification of mycobacteria have advanced hugely since the 1950s, the fundamental question posed by Runyon remains incompletely answered and of interest to mycobacteriologists today. Why have some mycobacterial species become clinically significant pathogens while others remain innocuous? Clinical relevance indicates that a mycobacterial species is able to engage in certain interactions with a human and/or animal host. Understanding the molecular

mechanisms underlying these interactions will enable the development of new strategies to combat existing mycobacterial pathogens and the identification of mycobacterial species at risk of becoming clinically relevant in the future, activities which will be essential for addressing the burden of current and emerging mycobacterial diseases.[157–160] Since the host is a highly specialised ecological niche, a useful approach is to consider mycobacteria in terms of their ecology.[161]

1.3. Mycobacterial ecology and host adaptation

1.3.1. Mycobacterial niches

While Mtb is exclusively associated with the human host, the genus *Mycobacterium* is ecologically diverse. Mycobacterial species have been isolated from a wide range of natural, built and host niches, from cockroaches and vacuum cleaners to distilled water and the space station Mir.[159,162–164] The majority of mycobacterial niches fall into three categories: soil, water and water management systems, and hosts.

1.3.1.1. Soil

Mycobacterium is consistently among the more abundant genera identified in analyses of soil bacterial communities, with some samples containing one million mycobacterial cells per gram.[165–167] Mycobacteria have been isolated from a wide variety of soil types, including peat bogs, swamps, mountain soil, forest soil, arable soil, and domestic soil.[163,168–170] Several mycobacterial species are able to degrade polyaromatic hydrocarbons, and can be found in oil wells and contaminated soils.[136,159,171]

1.3.1.2. Natural water and water management systems

Mycobacteria are able to replicate in fresh water, and are present in both surface water sources (for example rivers, streams, lakes and reservoirs) and underground water

sources (such as underground lakes, springs and wells).[116,163] They are also found in brackish waters (1—2% NaCl) such as estuaries, but generally do not tolerate the higher salinities characteristic of marine environments (>3% NaCl).[167,172,173] As previously discussed (**Section 1.2.2.2**), mycobacteria are hydrophobic owing to the high lipid content of their cell walls.[133] In consequence, mycobacteria in water tend to aggregate at the air-water interface or associate with suspended matter such as soil particles.[133,167] Additionally, mycobacterial biofilms are readily formed on surfaces, most notably the pipes supplying water to domestic, medical and industrial settings.[173–175]

Systems for the distribution and consumption of water therefore constitute a major source of mycobacteria in the built environment. Mycobacteria tend to be enriched in municipal water systems because they resist chemical treatment strategies such as chlorination.[174,176,177] As a result, mycobacteria are among the most common organisms isolated from showers, and can also be found in drinking water, hot tubs and swimming pools.[133,174,176–181] Mycobacteria have also been identified in hospital ice machines, endoscopes, dental water lines, and industrial process fluids (such as metal working water).[119,159,175,181–183] Finally, mycobacteria are unsurprisingly found in facilities that handle waste water, such as sewage treatment plants.[184]

1.3.1.3. Host niches

Mycobacteria may be isolated from eukaryotic hosts in either the presence or absence of disease.[175] The simplest hosts are free-living amoebae and other protozoa, unicellular eukaryotes which, like mycobacteria, are prevalent in soil and water.[185–189] Beyond humans, the animal hosts of mycobacteria are too numerous to list comprehensively

here. They include nematodes, turtles, fish, goats, monkeys, antelopes, deer, cows, pigs, birds, rabbits, badgers, horses, snakes, and elephants.[116,159]

1.3.2. Ecology of infection as a basis for the classification of mycobacteria

Some mycobacterial species are restricted to particular ecological niches, while others are more flexible about their living arrangements. *M. leprae*, for example, is so host-dependent as to be unculturable in artificial media, meaning that laboratory studies of leprosy must be conducted using nine-banded armadillos (for biomass production) or mouse foot-pads (for experimental infection studies).[190] In contrast, *M. avium* is culturable and isolable from a range of niches, including soil, water, birds and humans.[191,192] Mycobacteria can be divided into three groups according to their ecological preferences, with the key differentiator between the groups being their behaviour with respect to host niches. "Environmental saprophytic" mycobacterial species are free-living organisms that occupy non-host niches such as soil and surface water. Although they are sometimes isolated from host niches (in particular, from protozoa), these species do not cause disease.[116]

In contrast, mycobacterial species from the other two groups interact with host niches in ways that may provoke disease. The "obligate pathogenic" mycobacteria – Mtb being the most notable example – are strongly to exclusively host-associated. They cause disease in their preferred host (which may be humans or another animal) and are highly virulent.[116] Obligate pathogenic mycobacteria live intracellularly within the host. For example, Mtb resides primarily in lung macrophages, but can also survive inside other cell types, such as neurons, lymphatic endothelial cells, and adipocytes.[129,193–197] Obligate pathogenic mycobacterial infections are transmissible between members of

the host species, and in some cases between different host species.[116] The spreading of bovine TB (caused by *M. bovis*) by badgers is perhaps the best-known example of interspecies transmission of a mycobacterial infection.[198,199]

"Potentially pathogenic" mycobacteria comprise the third group, which spans the gap between the environmental saprophytes and the obligate pathogens. Thought to live primarily in non-host niches, potentially pathogenic mycobacteria are able to colonise susceptible hosts, causing disease.[116] Since around one-third of mycobacterial species are considered potentially pathogenic and infections with these species are typically acquired from the environment, non-host niches are significant reservoirs of mycobacterial infections.[116,133,180] An important species falling into this category is *M. ulcerans*, which causes Buruli ulcer, the third most common mycobacterial disease in humans.[200] Another example is the emerging pathogen *M. abscessus*, which is responsible for an increasing number of lung infections in people with cystic fibrosis.[201]

Using ecology of infection to classify mycobacterial species highlights the potential clinical relevance of each species to humans and/or animals. Like other pathogenicity-focused classifications, however, it ignores the possibility of commensal or symbiotic host-mycobacteria interactions.[116,202] While mycobacteria have not traditionally been considered commensals or symbiotes, some plant-mycobacteria relationships likely fall into these categories, as mycobacteria are thought to form part of the rhizosphere.[116,203] Additionally, it has been suggested that the principles of commensal ecology should be applied to the study of long-term host-mycobacteria relationships under immune control, such as those seen in latent TB.[204] Furthermore, some host-mycobacteria interactions may benefit the host.[205,206] For instance, prior inoculation with *M. vaccae*

increases resilience-associated behaviours in mice subjected to psychosocial stress.[205,207] These examples highlight that, although pathogenicity is a useful and accessible classification metric, we must be careful not to conflate pathogenicity and host adaptation – not all host-mycobacteria interactions result in disease. Indeed, transient colonisations are thought to make up the majority of host-mycobacteria interactions.[133,159] Nonetheless, if we consider the three ecological categories as a spectrum, the pattern of niche tolerances seen in the transition from environmental saprophyte to obligate pathogen brings to mind the general stages of bacterial host adaptation.[116,158,208]

1.3.3. Ecological profiles as stages of mycobacterial host adaptation

1.3.3.1. Stages of bacterial adaptation to intracellular lifestyles

Using whole genome sequencing data from 350 phylogenetically diverse bacterial species, Toft and Andersson proposed viewing various types of host-bacteria interaction as evolutionary stages on the pathway of bacterial adaptation to an intracellular lifestyle. In this depiction of host adaptation, bacteria begin as exclusively free-living organisms, unable to survive inside other cells. On gaining the ability to survive inside a eukaryotic cell, bacteria initially adopt a facultative intracellular lifestyle, meaning that they can grow both inside and outside of host cells. The ability to live extracellularly (including on artificial media) is then lost during the transition to an obligate intracellular lifestyle. Host specialisation is also seen at the obligate intracellular stage. Once host dependency is established, host-bacteria coevolution may become selective for benefit to the host, with the bacteria becoming obligate intracellular mutualists. At the highest level of host adaptation, bacteria are incorporated into the host cell as organelles, losing their taxonomic status as bacteria.[208]

1.3.3.2. Are ecological profiles indicative of evolutionary status?

To what extent do the ecological categories of mycobacteria – environmental saprophytes, potential pathogens and obligate pathogens – correlate with stages in the bacterial host adaptation pathway proposed by Toft and Andersson? The earliest mycobacteria, like many extant environmental saprophytic mycobacteria, are thought to have been exclusively free-living.[116,185–187,209] Environmental saprophytic mycobacteria able to survive inside amoebae could be considered to exhibit early-stage facultative intracellular lifestyles.[188] Indeed, amoebae are thought to serve as a 'nursery' in which free-living bacteria can take their first steps towards survival inside eukaryotic cells.[133,157,158,191,208,210] Free-living bacteria are liable to phagocytosis by the amoebae with which they share ecological niches, and so may evolve mechanisms to evade destruction.[188,189] These mechanisms can later be adapted to survive and replicate inside other cells, such as macrophages.[159,187,208,210] Consistent with this theory, the ability of mycobacteria to replicate in macrophages correlates with their ability to replicate in amoebae.[210] Similarly, potentially pathogenic mycobacteria are able to live freely in non-host niches, and many species can replicate inside amoebae.[133,186,187] Any potentially pathogenic mycobacterial species able to live both inside and outside of cells could be considered to lead a facultative intracellular lifestyle; however, isolation of that species from a complex eukaryote does not guarantee that it has adopted an intracellular lifestyle within the host. For example, *M. avium* has been shown to replicate on mucosal surfaces, while *M. abscessus* forms biofilms during infection of human lung tissue.[211,212]

As obligate pathogenic mycobacteria are intracellular pathogens and are not found in non-host niches, they have arguably progressed beyond facultative intracellularity.[213] However, with the exception of *M. leprae*, they can still be grown on artificial media and so do not quite meet the criteria for obligate intracellularity laid out by Toft and Andersson.[208] Finally, while it is well known that *M. leprae* has undergone substantial genome degradation, it is presently unclear whether the genomic features of Toft and Andersson's host adaptation model apply to the genus *Mycobacterium* as a whole.[158,190,208,214–220]

Clearly, the ecological categories of environmental saprophytic, potentially pathogenic and obligate pathogenic mycobacteria do not correspond directly to the evolutionary categories of extracellular, facultative intracellular and obligate intracellular bacteria (**Figure 5**). However, the niche occupancy pattern exhibited by the ecological categories – broadening with the development of host tolerance and then narrowing with host specialisation – is consistent with that expected for progressive host adaptation.[208] Consequently, classifying mycobacteria based on their ecological profiles provides a useful, if somewhat imprecise, indication of their host adaptation status.[c] Comparing mycobacterial species with different ecological profiles may therefore be of use in

[c] Ecological classification presents different challenges to other methods for the classification of mycobacteria (such as the Runyon, chemotaxonomic and genomic methods discussed earlier) because it is based on literature reports of isolation and/or infection rather than on experimental data. Since literature coverage of the genus *Mycobacterium* is biased towards Mtb and human disease, one may not always have enough information to make a confident classification, particularly when considering recently identified species or species that have historically been conflated. As a result of the literature focus on pathogenicity, "environmental" is often used to refer to all mycobacterial species found in non-host niches, not only those classified as environmental saprophytic mycobacteria. After extensive discussion (with Luiz Carvalho, Acely Garza-Garcia, Catherine Hubert and Fernanda Subtil) we concluded that the literature can be used to infer the ecological status of most mycobacterial species.

understanding the evolution of mycobacterial host adaptation and the cellular processes involved therein.

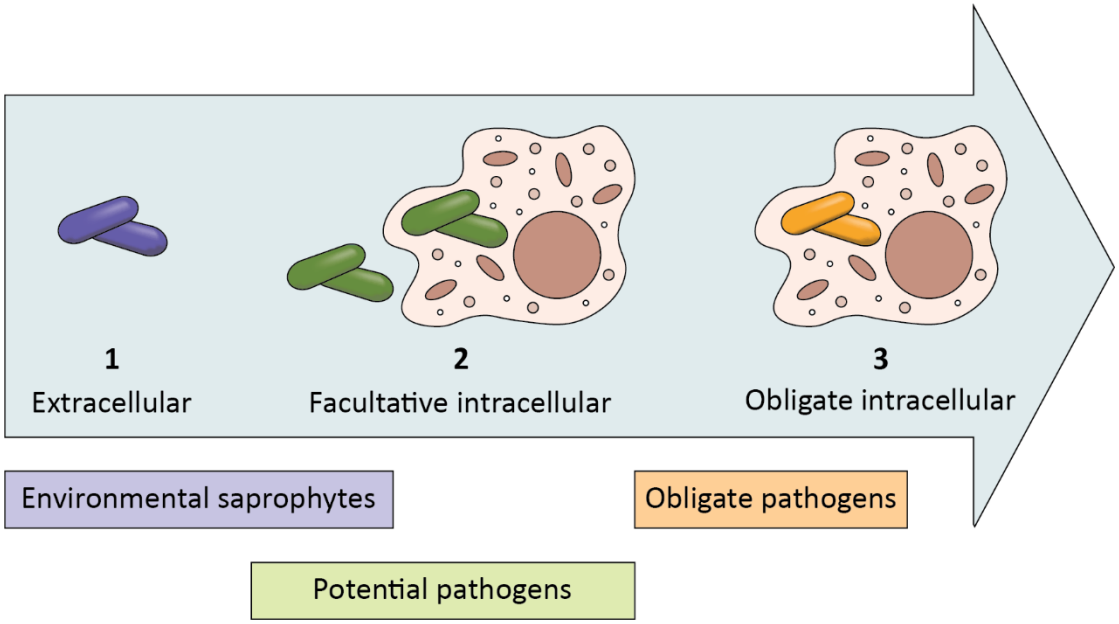


Figure 5: Ecology of infection as an indicator of host adaptation status.
The blue arrow (top) shows the first three stages in the bacterial host adaptation pathway proposed by Toft and Andersson and discussed in **Section 1.3.3.1**. Extracellular bacteria (**1**) acquire the ability to survive inside cells and become facultative intracellular bacteria (**2**). Host dependence then increases and the bacteria eventually lose the ability to live outside the host (**3**).[208] Using niche occupancy patterns to map the ecological categories of mycobacteria – environmental saprophytes, potential pathogens and obligate pathogens – onto the bacterial host adaptation pathway shows that ecology of infection is broadly reflective of host adaptation status. (Mycobacteria have not progressed beyond stage **3** of this pathway.)

1.4. Genus-level comparison as a tool to understand mycobacterial phenotypes

1.4.1. Comparison as a tool to understand bacterial phenotypes

Comparing sets of related bacteria that display different phenotypes can help us to understand the mechanisms underlying ecological adaptations such as host-microbe interactions and nutrient utilisation profiles.[221,222] For example, a comparison of the lipidomes of pathogenic and non-pathogenic serovars of *Salmonella enterica* revealed a previously unknown family of trehalose phospholipids that are enriched in the pathogenic serovar *S.* Typhi. One of these lipids was identified as potentially

immunogenic and is likely to be involved in the interaction between *S.* Typhi and the human host, together with other differentially produced cell wall lipids.[223,224] As another example, a comparison of proteomic remodelling in three species of marine bacteria entering the stationary phase of growth found that the species employed different metabolic strategies for dealing with nutrient depletion.[225] Such observations can help to explain the changes in microbial community composition that occur in response to ecological events, such as phytoplankton blooms, which alter the nutrient content of the marine environment.[225,226]

1.4.2. Mtb exhibits low phenotypic and genetic variability between strains

The details we can uncover using a comparative strategy naturally depend on the level of diversity present in the sample set. Compared to other bacterial species such as *S. enterica* and *E. coli*, Mtb exhibits low genetic diversity, as measured by its pan-genome composition.[227,228] The pan-genome of a bacterial species is all the genes that are associated with that species and is divided into the core genome (genes that are present in all strains) and the accessory genome (genes that are only present in some strains).[229,230] A recent pan-genomic analysis of Mtb identified 3679 core genes. These core genes comprise approximately 87% of the genome of any individual Mtb strain, and around 64% of the Mtb pan-genome.[231] In contrast, core genes account for two-thirds of the genome of any individual *E. coli* strain, and less than 10% of the *E. coli* pan-genome.[232]

The high genetic diversity of *E. coli* is reflected in the wide phenotypic variability between strains. For example, *E. coli* K12 is a mammalian intestinal commensal, while *E. coli* O157:H7 causes food poisoning with bloody diarrhoea.[233,234] Thus, *E. coli* strains,

like the *S. enterica* serovars discussed above, exhibit sufficient diversity that comparative experiments can readily identify new molecules or pathways involved in interactions with the host niche. In contrast, all Mtb strains are human pathogens, and the Mtb accessory genome is responsible for more nuanced differences between strains.[231]

The high similarity between Mtb strains limits the ability of a species-level comparative approach to inform on the mechanisms underlying host adaptation and pathogenicity. To gain comparative power, we must increase the genetic and phenotypic diversity of our sample set by stepping beyond Mtb. However, the genetic diversity must not become so great that we lose the ability to pinpoint the origins of phenotypes of interest. In seeking to strike a balance between genetic similarity and phenotypic diversity, perhaps the most natural starting point is to compare Mtb to its closest relatives: other members of the genus *Mycobacterium*.[235] While they are phenotypically diverse, mycobacteria are genetically more closely related than the members of other bacterial genera, suggesting that inter-species comparison is likely to informatively balance genetic similarity and phenotypic diversity.[236]

1.4.3. Individual mycobacterial species are suboptimal models for Mtb

As access to Biosafety Level 3 laboratories is required for Mtb work, there has long been an interest in using mycobacterial species which can be handled at lower containment levels as model organisms for Mtb, particularly in the context of drug discovery.[237–239] The most commonly selected species are *M. smegmatis* (which has only a limited ability to survive inside macrophages), *M. marinum* (which is used in the zebrafish model of tuberculosis) and *M. bovis* BCG (the attenuated vaccine strain discussed in **Section**

1.1.4).[240–244] These species are suboptimal models for Mtb because they engage differently with the human host, meaning that the mechanisms underlying many Mtb disease processes are absent or altered.[245–249] When seeking to identify these mechanisms and understand their evolutionary origins, however, the very features that limit other mycobacterial species as model organisms for Mtb make them valuable comparators.

How, then, should we choose which mycobacterial species to compare? Individual species will have evolved unique features to suit their own particular lifestyles.[240] Indeed, around 40% of mycobacterial accessory genes are species-specific.[250] Thus, broadly relevant molecular mechanisms might be more effectively identified and understood by considering the whole genus, rather than through the study of individual species or pairs of species. We therefore propose to take a genus-level comparative approach to the elucidation of the molecular mechanisms underlying mycobacterial phenotypes, focusing particularly on the processes involved in antibiotic resistance and metabolic adaptation to the host niche.

1.4.4. A genus-level comparative approach to the elucidation of the molecular mechanisms underlying mycobacterial phenotypes

To recapitulate, comparison is a powerful tool for the elucidation of the molecular mechanisms underlying bacterial phenotypes (**Section 1.4.1**). Effective comparison requires careful balancing of genetic similarity and phenotypic diversity, such that features of interest are both identifiable and explainable (**Section 1.4.2**). In the case of mycobacteria, we believe that this balance falls at the genus level, and that multispecies studies will deliver broader insights than pairwise comparisons (**Section 1.4.3**). Further,

the integration of multiple techniques (such as bioinformatics, enzymology, microbiology and metabolomics) at the genus level will facilitate *in vitro* validation of the results of comparative genomics studies and, conversely, provide genetic explanations for the patterns observed in biochemical and microbiological data.[251,252]

Accordingly, members of our laboratory have assembled a library of mycobacterial species considered to be representative of the genus as a whole. In 2018, when work on this thesis began, 32 species were available for study in our laboratory. Our collection has since expanded, and currently comprises 74 mycobacterial species. By using members of the library in combinations of comparative studies, we hope to uncover the biochemical processes involved in mycobacterial metabolism and antibiotic resistance and situate these processes in their evolutionary and ecological context. Ultimately, discovery of the molecular mechanisms by which mycobacteria tolerate antibiotics and the host niche will expand our understanding of basic mycobacterial biology, identify novel targets for the development of antimycobacterial agents, and highlight mycobacterial species at risk of becoming clinically relevant in the future. Herein, we take a genus-level comparative approach to the study of isocitrate metabolism, a process of critical importance during Mtb infection.

1.5. The importance of isocitrate metabolism in Mtb

1.5.1. Metabolic plasticity is critical during Mtb infection

Mtb undergoes extensive metabolic remodelling in response to the lung environment, a process which begins immediately after macrophage infection.[253] The transcriptome stabilises between 24 and 48 hours after infection, then continues to evolve more slowly over the following two weeks. Microarray analysis (validated by qRT-PCR) 24 hours after

infection reveals a 'core intramacrophage transcriptome': 280 Mtb genes for which changes in expression are conserved across 15 Mtb strains and in both resting and interferon-γ-activated murine macrophages. Of these genes, 168 are induced. They are involved in processes such as cell wall remodelling, fatty acid metabolism, and stress responses (hypoxia, oxidative stress and nitrosative stress), indicating that these activities are important for Mtb survival inside host macrophages.[102,254,255]

A major metabolic change that accompanies the transition from standard laboratory culture media to the macrophage is a shift in carbon utilisation from carbohydrates (glucose and glycerol) to lipids (fatty acids and cholesterol). By 24 hours after infection, all 18 genes required for the activation and β-oxidation of fatty acids have been upregulated, indicating that Mtb perceives the macrophage as fatty acid rich and carbohydrate poor (discussed further in **Section 1.9.1**).[253,256] Given that modulation of carbon metabolism plays an essential role in adjustment to the intramacrophage environment, it is perhaps unsurprising that deletion or inactivation of enzymes dispensable for growth on standard *in vitro* mycobacterial culture media results in some of the most profound attenuations in Mtb pathogenicity and persistence ever seen in mouse models.[257,258] A well-known example is isocitrate lyase 1, one of four Mtb enzymes for the metabolism of isocitrate.[259]

1.5.2. The metabolic position of isocitrate

Isocitrate is the product of citrate isomerisation by aconitase.[260] In Mtb, as in many other bacterial species, isocitrate sits at a branch point in central carbon metabolism (**Figure 6**).[261] The tricarboxylic acid (TCA) cycle can be continued through the decarboxylation of isocitrate to α-ketoglutarate, a reaction catalysed by isocitrate

dehydrogenases (IDHs). Alternatively, carbon flux may be diverted through the glyoxylate shunt, a two-step pathway.[261–263] The first reaction in the glyoxylate shunt is the conversion of isocitrate to glyoxylate and succinate by isocitrate lyases (ICLs). This step is followed by the synthesis of malate from glyoxylate and acetyl-CoA (Ac-CoA), which is catalysed by malate synthase.[264–266]

While the TCA cycle and the glyoxylate shunt are typically active concurrently, the balance of flux between the two changes in response to the nutrient environment.[258,262,264,267] The glyoxylate shunt bypasses the two decarboxylation steps of the TCA cycle, enabling the synthesis of succinate from Ac-CoA without loss of carbon.[264,268] Conservation of carbon is essential for anaplerosis (the replenishment of TCA cycle intermediates) and gluconeogenesis during growth on 'two-carbon' substrates such as fatty acids, ethanol and alkanes, which enter central carbon metabolism as Ac-CoA.[261,264–266,269,270] When the glyoxylate shunt is active, some level of flux must be maintained through an $NADP^+$-dependent IDH to generate NADPH for biosynthesis (discussed further in **Section 1.7.1**).[261] For example, during growth of *M. smegmatis* on acetate, 32—36% of succinate is synthesised by the glyoxylate shunt, with the remainder being produced by the TCA cycle.[267] [d]

[d] Flux partitioning at the isocitrate branch point contributes to the process of balancing the cellular demand for energy with the need to produce biosynthetic precursors. Microbial carbon use efficiency is the proportion of incoming carbon atoms that are incorporated into biomass (rather than being respired to CO_2 to meet the cellular energy demand).[553,554] *In silico* models of bacterial metabolism indicate that carbon use efficiency varies between species and carbon sources. In Mtb, the mean carbon utilisation efficiency during growth on single carbon sources is about 0.3, meaning that around 30% of the carbon that enters the cell ends up as biomass.[553] Models typically assume that oxygen is not limiting and that intermediates (such as β-oxidation products) are not accumulated; in reality either/both of these assumptions may not hold.[554]

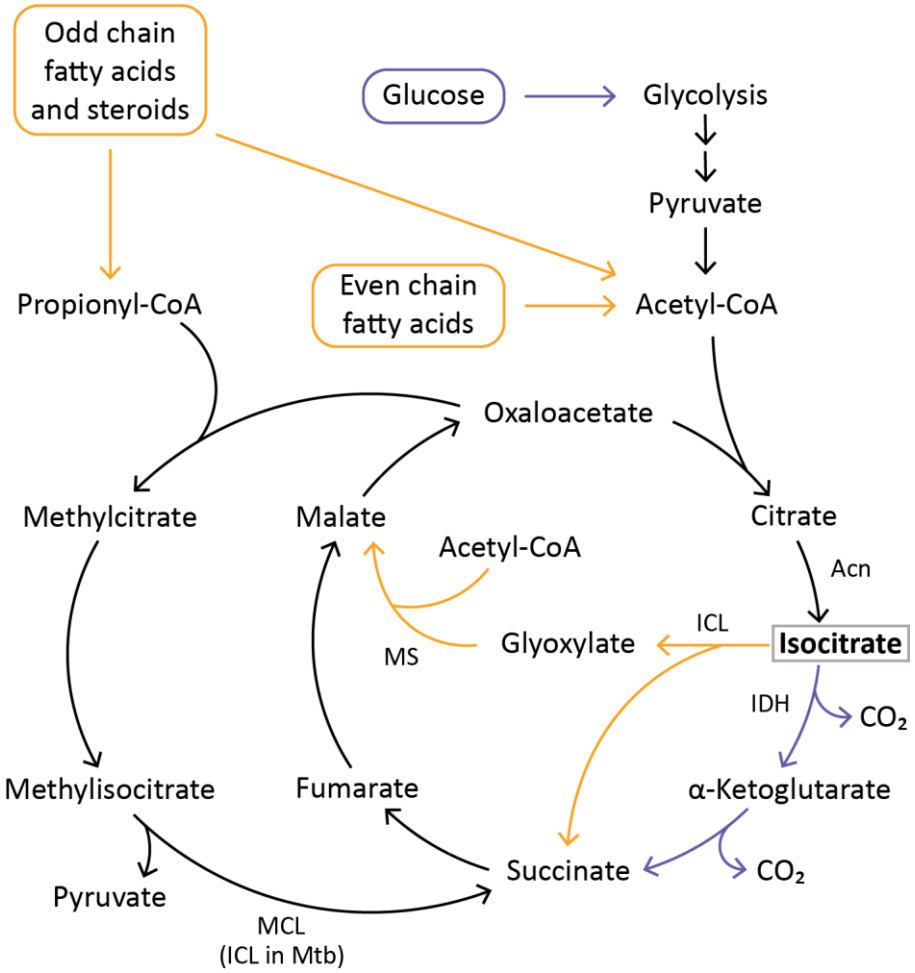


Figure 6: The position of isocitrate in mycobacterial central carbon metabolism.
Produced from citrate by aconitase (Acn), isocitrate is situated at the branch point between the TCA cycle and the glyoxylate shunt. Continuation of the TCA cycle (purple arrows) involves the decarboxylation of isocitrate to form α-ketoglutarate, catalysed by isocitrate dehydrogenase (IDH). The glyoxylate shunt (yellow arrows) allows the two decarboxylation steps in the TCA cycle to be bypassed, enabling the accumulation of carbon during growth on fatty acids. Isocitrate lyase (ICL) converts isocitrate to glyoxylate and succinate. Malate synthase (MS) then catalyses the production of malate from glyoxylate. In Mtb, ICLs also function as methylisocitrate lyases (MCLs), converting methylisocitrate to succinate and pyruvate. This reaction is part of the methylcitrate cycle, which is used to metabolise the propionyl-CoA produced during the breakdown of odd chain fatty acids and cholesterol. Metabolism of steroids produces both Ac-CoA and propionyl-CoA.[261,271–273]

Many species of bacteria and fungi use the glyoxylate shunt to facilitate growth on two-carbon substrates acquired from the environment.[265] In higher plants, meanwhile, this pathway enables the synthesis of carbohydrates from stored lipids during seed germination.[265,266,274,275] The glyoxylate shunt is absent in animals, with the exception of some nematodes during embryogenesis. Coupled with the presence and importance of

the pathway in bacterial and fungal pathogens, this absence has resulted in substantial interest in glyoxylate shunt enzymes as drug targets, albeit with limited success.[261,265,274,276]

1.5.3. Structure of the isocitrate branch point in Mtb

Consisting of four distinct enzymes, the isocitrate branch point in Mtb is notable for its complexity (**Figure 7**). Two $NADP^+$-dependent IDHs contribute to flux through the TCA cycle: IDH1 (encoded by *icd1*, *rv3339c*) and IDH2 (encoded by *icd2*, *rv0066c*). While most organisms that employ the glyoxylate shunt possess only one ICL, Mtb has two: ICL1 (encoded by *icl*, *rv0467*) and ICL2 (encoded by *aceA*, *rv1915—1916*).[261,277]

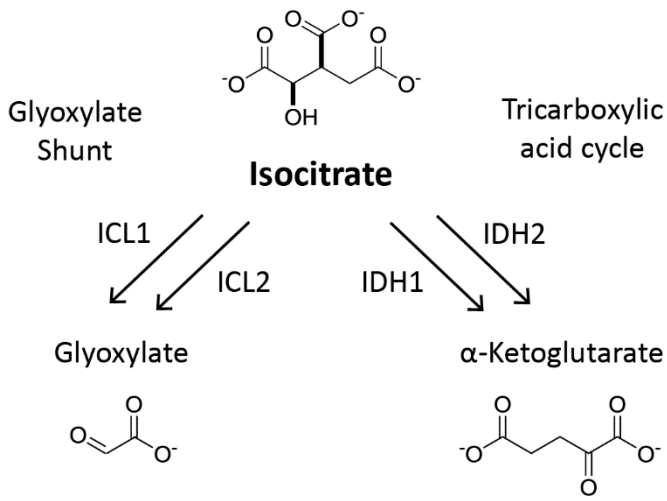


Figure 7: Mtb encodes four distinct enzymes for the metabolism of isocitrate.
Continuation of the TCA cycle from isocitrate begins with the production of α-ketoglutarate and CO_2. In Mtb, this reaction is catalysed by two $NADP^+$-dependent IDHs, IDH1 (encoded by *icd1*, *rv3339*c) and IDH2 (encoded by *icd2*, *rv0066*c). Diversion of carbon flux through the glyoxylate shunt begins with isocitrate lysis, which produces glyoxylate and succinate. This reaction is catalysed by ICLs, of which Mtb has two: ICL1 (encoded by *icl*, *rv0467*) and ICL2 (encoded by *aceA*, *rv1915—1916*).

Before reviewing the biochemistry and evolution of these enzymes in more detail in **Sections 1.6** (ICLs) and **1.7** (IDHs), we explore the importance of the glyoxylate shunt during growth on fatty acids (**Section 1.5.4**) and during infection (**Section 1.5.5**).

1.5.4. Mtb isocitrate metabolism during growth on fatty acids

The glyoxylate shunt is expected to be most active during growth on carbon sources which enter central carbon metabolism as Ac-CoA, such as even-chain fatty acids (**Section 1.5.2**). Indeed, compared to growth on glucose, total isocitrate lyase activity in the Mtb strains H37Rv and CDC 1551 is doubled when acetate is the limiting carbon source and tripled when palmitate is the limiting carbon source.[278] Deletion of a single isocitrate lyase (Δ*icl* or Δ*aceA*) does not affect the ability of the Erdman strain of Mtb to grow on either carbohydrates or fatty acids (**Table 2**). The double deletion mutant Δ*icl*Δ*aceA* is unable to grow on fatty acids. Growth on these substrates can be restored by complementation with either *icl* or *aceA*, suggesting that ICL1 and ICL2 may be functionally redundant.[279]

Exposure to small-molecule ICL inhibitors such as 3-nitropropionate (3-NP) produces a growth phenotype similar to that of the Δ*icl*Δ*aceA* mutant. Mtb Erdman cannot grow on fatty acids in the presence of 0.1 mM 3-NP, but growth on glucose is only slightly reduced and growth on glycerol is unaffected.[279] Similarly, treatment with 0.2 mM 3-NP reduces the growth of Mtb H37Rv by 97% when acetate is the sole carbon source, compared to 34% when glucose is the sole carbon source.[280] Another ICL inhibitor, itaconate, also inhibits Mtb H37Rv growth on acetate, but does not affect growth on glucose or glycerol.[281,282] These results further demonstrate the conditional essentiality of the glyoxylate shunt for Mtb growth on even-chain fatty acids such as acetate.

Table 2: Growth of Mtb Erdman isocitrate lyase mutants on various carbon sources.
Deletion of a single isocitrate lyase gene does not affect growth. Mutants where both *icl* and *aceA* have been deleted cannot grow on fatty acids, but retain the ability to grow on carbohydrates. Complementation with either *icl* or *aceA* restores growth on fatty acids. [a] Polysorbate 20 is a complex ester of lauric acid (C_{12}); mycobacteria access its fatty acid chains using Tween 20 hydrolase.[283] [b] Growth is less than the wild type; [c] exhibits a lag phase compared to the wild type, indicating that there is some flux through the glyoxylate shunt during wild type growth on glucose. **Abbreviations:** Polysorbate 20, polyoxyethylene sorbitan monolaurate; WT, wild type; pICL1, complemented with a plasmid containing *icl*; pICL2, complemented with a plasmid containing *aceA*; ✓, grows; x, does not grow.[279]

Carbon source	**Type**	**WT**	**Δ*icl***	**Δ*aceA***	**Δ*icl*Δ*aceA***	**pICL1**	**pICL2**
Glycerol	Carbohydrate	✓	✓	✓	✓	✓	✓
Glucose	Carbohydrate	✓	✓	✓	✓[c]	✓	✓
Acetate	Short chain fatty acid, C_2	✓	✓	✓	x	✓	✓
Propionate	Short chain fatty acid, C_3	✓	✓[b]	✓	x	✓	✓
Butyrate	Short chain fatty acid, C_4	✓	✓	✓	x	✓	✓
Polysorbate 20[a]	Long chain fatty acid, C_{12}	✓	✓	✓	x	✓	✓
Palmitate	Long chain fatty acid, C_{16}	✓	✓	✓	x	✓	✓

The observations that Δ*icl*Δ*aceA* Mtb Erdman is unable to grow on the three-carbon fatty acid propionate (**Table 2**) and that itaconate inhibits the growth of wild-type Mtb Erdman on propionate highlight the second important function of ICLs in Mtb metabolism, as methylisocitrate lyases (MCLs).[279,282] MCLs catalyse the conversion of 2-methylisocitrate to pyruvate and succinate. This reaction is part of the methylcitrate cycle (**Figure 6**), which metabolises the propionyl-CoA produced by β-oxidation of odd-chain fatty acids, branched-chain fatty acids and branched-chain amino acids.[282,284]

Mtb does not encode a separate MCL, instead using its ICLs to perform this catalytic function.[285] Incubation of Δ*icl* Mtb Erdman with propionate results in depletion of TCA cycle intermediates and accumulation of the propionate metabolites 2-methylcitrate, propionyl-CoA and methylmalonyl-CoA. Gluconeogenesis becomes impaired by (i) a lack of oxaloacetate and malate, and (ii) the inhibition of fructose 1,6-bisphosphatase by 2-methylcitrate. The cellular NAD^+/NADH balance, membrane potential, and pH homeostasis are disrupted, resulting in acidification of the cytosol. These effects cumulatively result in cell death.[282] The *in vitro* growth experiments discussed above showed that ICLs are essential for the metabolism of both even-chain and odd-chain fatty acids in Mtb. Given that fatty acids are major carbon sources for Mtb in lung macrophages (**Section 1.5.1**), one might expect control of flux through isocitrate metabolising enzymes to be critical for Mtb during infection.

1.5.5. Mtb isocitrate metabolism during infection

Of the isocitrate metabolism genes in Mtb (*icl*, *aceA*, *icd1* and *icd2*), only *icl* is significantly upregulated following macrophage infection.[256,286] Upregulation of *icl* can be observed within two hours of murine macrophage infection with Mtb CDC 1551. Levels of the *icl* transcript continue to increase during the first 24 hours after infection, and other components of fatty acid metabolism are also upregulated during this time.[253] Elevated levels of *icl* mRNA have also been observed in Mtb H37Rv following infection of the macrophage-mimicking human cell line THP-1, and in lung samples from mice and humans infected with Mtb.[256,287,288] In mouse lung samples, *icl* remains upregulated 45 days after infection, suggesting a continuing requirement for ICL1 during chronic Mtb infection.[289]

Transcriptomics experiments suggest that *icl* upregulation is independent of the activation state of the macrophage, with *icl* being strongly induced (12—20-fold) in both resting and interferon-γ-activated mouse bone marrow macrophages.[254,256] However, fluorescence-based protein quantification experiments found that the expression of ICL1 fused to green fluorescent protein (GFP) in Mtb CDC 1551 was influenced by the macrophage activation state. In activated macrophages, ICL1-GFP levels rose shortly after infection and remained elevated throughout the experiment. While the same initial rise in ICL1-GFP levels was seen in resting macrophages, the increase was not sustained, with ICL1-GFP returning to background levels by 24 hours after infection.[259]

Macrophage experiments with Δ*icl* Mtb Erdman also indicate that the requirement for ICL1 depends on the host cell activation status. The phenotypes of the wild type and Δ*icl* strains diverge by 24 hours after infection, with Δ*icl* bacilli becoming less viable than the wild type. In resting macrophages, 63% of Δ*icl* bacilli survive 120 hours after infection, compared to 87% of wild-type bacilli. The divergence in viability is far more pronounced in activated macrophages, where 1.2% of Δ*icl* bacilli and 41% of wild-type bacilli survive for 120 hours.[259] Survival of the double deletion mutant Δ*icl*Δ*aceA*, meanwhile, is impaired irrespective of the macrophage activation state. Complementation with a plasmid containing either *icl* or *aceA* restores survival in activated macrophages and replication in resting macrophages.[279] These observations suggest that, as we might expect from the *in vitro* data for carbon sources thought to be relevant during infection (**Table 2**), some degree of (methyl)isocitrate lyase activity is required for Mtb survival inside macrophages.

ICL1 is essential for the virulence and persistence of Mtb in immunocompetent mouse models. When immunocompetent mice are exposed to Δ*icl* Mtb Erdman, wild-type bacterial growth and lung pathology are seen during the acute phase of infection (0—2 weeks). However, the mutant is unable to persist and is cleared from the lungs and extrapulmonary organs by week 16.[259,279,290] In contrast, Δ*icl* Mtb Erdman is lethal to immunodeficient (interferon-γ knockout) mice, again supporting the conclusion that *icl* essentiality is dependent on the immune status of the host.[259] The double deletion mutant Δ*icl*Δ*aceA* is unable to grow in either immunocompetent or immunodeficient mice, and is rapidly eliminated. Immunocompetent mice infected with this mutant were still culture negative after two years (when the experiment was terminated), confirming that (methyl)isocitrate lyase functionality is required for Mtb survival inside the host.[279]

Deletion of *aceA* alone does not affect Mtb persistence in immunocompetent mice, suggesting that *aceA* is less important than *icl* during infection.[279] Although ICL1 and ICL2 appear to be functionally redundant *in vitro* (**Section 1.5.4**), the macrophage and mouse experiments discussed above indicate that this is not the case in the host niche. The more aggressive the host-bacteria interaction, the more critical *icl* becomes and the more inadequate *aceA* becomes as a substitute. *Icl* may therefore have additional roles during infection, beyond the metabolism of fatty acids. This hypothesis is supported by the observations that (i) upregulation of *icl* is involved in the response to iron starvation, hypoxia, and antibiotic-induced oxidative stress; and (ii) *glcB* (encoding malate synthase, the second enzyme in the glyoxylate shunt) is not upregulated during infection.[270,286,291–294]

1.6. Isocitrate Lyases

1.6.1. Introduction to isocitrate lyases

Isocitrate lyases (ICLs) catalyse the first reaction in the glyoxylate shunt: the retro-aldol cleavage of (2*R*,3*S*)-isocitrate to glyoxylate and succinate (**Figure 8**).[261,295] They belong to the phosphoenolpyruvate mutase/isocitrate lyase-like family of enzymes and are assigned the Enzyme Commission (EC) number 4.1.3.1.[295–297] ICLs are found in many species of bacteria, plants and fungi (**Section 1.5.2**), and are homotetramers in solution.[275,276,295] The ICL-MCL bifunctionality exhibited by the Mtb ICLs (**Section 1.5.4**) is unusual – most ICLs have no or very weak MCL activity and cannot support growth on propionate.[272,284,298] All ICLs contain a distorted α/β barrel fold.[297] Eight parallel β-strands form a central barrel, which is surrounded by seven α-helices.[296,297,299–301] As in most α/β barrel enzymes, the active site is situated at the *C*-terminal ends of the β-strands.[276,300–302] A divalent cation (Mg^{2+} or Mn^{2+}) and a thiol are required for ICL activity.[295]

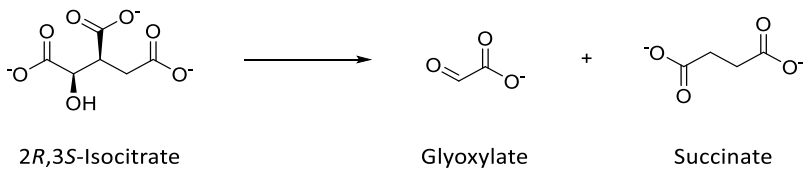


Figure 8: ICLs catalyse the retro-aldol cleavage of (2*R*,3*S*)-isocitrate to glyoxylate and succinate.

While the α/β barrel is always present, the ICLs from prokaryotes and eukaryotes differ in their overall length and domain organisation. On the basis of these differences, which are immediately visible on sequence alignment, the known ICLs can be divided into three groups (**Figure 9**).[279,303] Group I ICLs are the smallest, containing around 430 amino acids. Prokaryotic ICLs, for example *E. coli* ICL and Mtb ICL1, belong to this

group.[279,300,301] The eukaryotic ICLs, represented in **Figure 9** by *Aspergillus nidulans* ICL, contain an additional domain and comprise Group II. X-ray crystallography reveals that the additional amino acids (Domain II, coloured yellow in **Figure 9**) do not disrupt the α/β barrel fold (Domains I and III, coloured dark and light teal respectively in **Figure 9**), instead forming a series of helices and loops on the outside of the protein.[279,299,302]

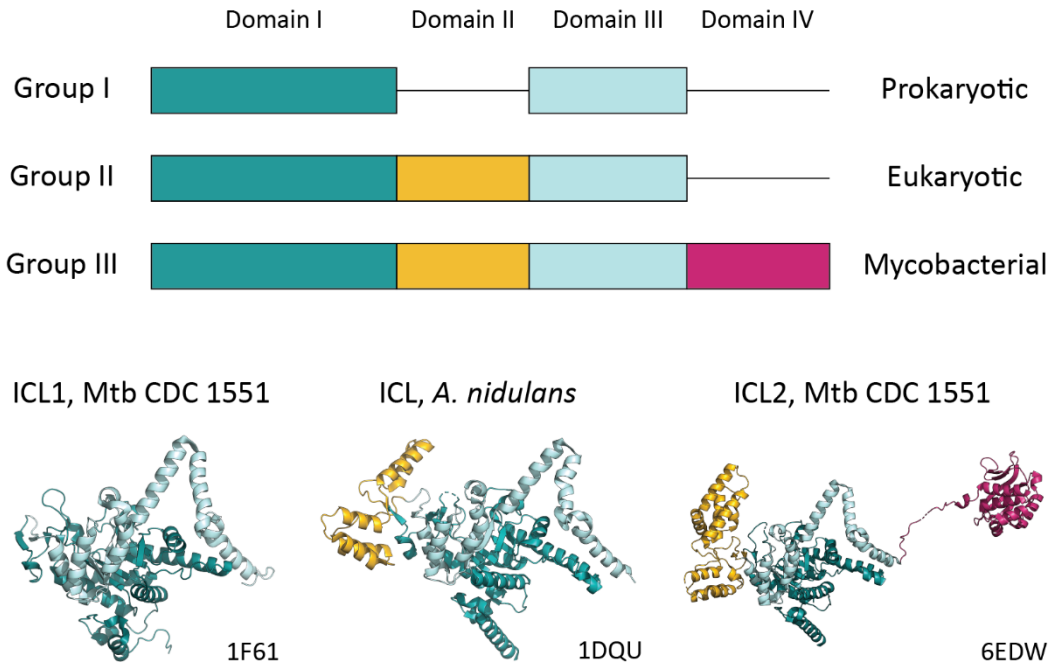


Figure 9: Structure-based division of ICLs into three groups.
ICLs can be classified into three groups based on differences in their length and domain organisation. Mycobacterial ICL1s (represented by Mtb CDC 1551 ICL1) and other prokaryotic ICLs have the simplest structure, and belong to Group I. Two domains (I and III, coloured dark and light teal respectively) form a distorted α/β barrel fold. The active site is situated at the *C*-terminal ends of the parallel β-strands that make up the centre of this barrel. The key catalytic residues (Cys191 and Asp108 in Mtb ICL1) are located in Domain I. Eukaryotic ICLs (represented by *Aspergillus nidulans* ICL) possess an additional domain (Domain II, coloured yellow) in the middle of the protein, and are classified as Group II ICLs. Mycobacterial ICL2s (represented by Mtb CDC 1551 ICL2) contain the three Group II domains and a fourth domain (Domain IV, coloured pink) located at the *C*-terminus. Until recently, they were the only known members of Group III.[279] Domain lengths vary between ICLs, but typical lengths are: Domain I, 250—270 aa; Domain II, 100—175 aa; Domain III, 160—180 aa; Domain IV, ~175 aa.

Until recently, mycobacterial ICL2s were the only reported members of Group III.[276,302,304] They are the longest ICLs, possessing a *C*-terminal domain (Domain IV, coloured pink in **Figure 9**) in addition to the three domains seen in the Group II

ICLs.[279] The *C*-terminal domain displays no sequence homology to any known protein, but structurally resembles a Gcn5-related *N*-acetyltransferase (GNAT).[276,302,303,305] Group III ICLs are commonly described as "eukaryotic-like" because they share Domain II with the Group II enzymes.[271,304,306,307] However, it cannot be assumed that the structural similarity between the Group II and Group III ICLs is representative of their evolutionary relationship. Given the ubiquitous nature of the α/β barrel fold, the similarities in domain organisation could well have arisen through convergence rather than divergence.

1.6.2. Mtb isocitrate lyase 1 (ICL1)

Encoded by *icl* (*rv0467*), Mtb isocitrate lyase 1 (ICL1) is the most studied mycobacterial isocitrate metabolising enzyme due to its essentiality in mouse models of TB (**Section 1.5**).[277] The first crystal structure of ICL1 was published in 2000 (PDB: 1F61) and was the second ICL structure to be solved, after ICL from *A. nidulans* (PDB: 1DQU).[299,301] Each ICL1 monomer comprises 428 amino acids and includes an α/β barrel with $(\beta\alpha)_2\alpha(\beta\alpha)_5\beta$ topology.[301,302,308,309] ICL1 oligomerisation involves helix swapping, where the *C*-terminal helices of one monomer are exchanged for those of a neighbouring monomer to form a stable dimer. These dimers then associate to form the ICL1 homotetramers observed in solution and in crystal structures.[301,310–313] The exchanged *C*-terminal helices are also involved in the conformational change that occurs on substrate binding. Isocitrate binding results in the movement of bound Mg^{2+}, causing ICL1 to shift from the open conformation to the closed conformation. The catalytic loop closes over the active site and the *C*-terminal helices of the neighbouring subunit form a lid on top, excluding solvent from the active site.[301,302,311,314,315]

Catalysis follows an ordered uni-bi mechanism. Isocitrate binds, retro-aldol cleavage takes place, and the products are released sequentially: succinate followed by glyoxylate.[316] The chemical mechanism of isocitrate lysis has been determined using structural, kinetic and computational data, and is shown in **Figure 10**.[301,317–320] Isocitrate and water coordinate Mg^{2+} in the active site.[317] Asp108 deprotonates a water molecule, which in turn deprotonates isocitrate, leading to retro-aldol cleavage.[320] The resulting enolate is then protonated by Cys191 to form succinate.[317] Kinetic parameters are given in **Table 3**.

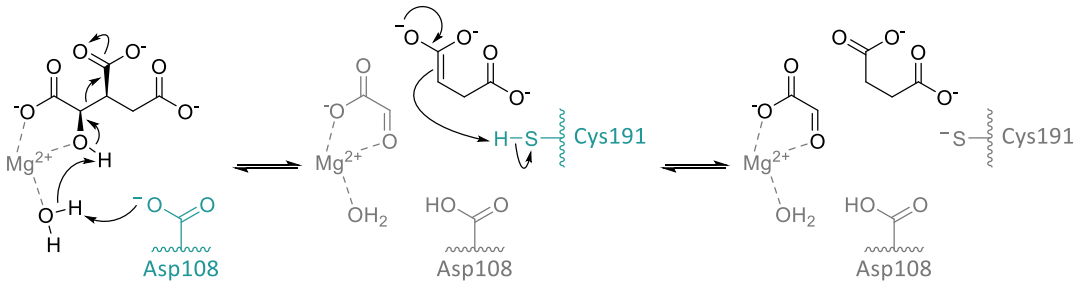


Figure 10: Chemical mechanism of isocitrate lysis by Mtb ICL1.
Mg^{2+} in the active site is coordinated by isocitrate and water. Deprotonation of a water molecule by Asp108 leads to the deprotonation of isocitrate, which triggers retro-aldol cleavage. Protonation of the resulting enolate by Cys191 gives succinate. Succinate is released from the active site, followed by glyoxylate. As noted in **Figure 9**, both Cys191 and Asp108 are located in Domain I of the enzyme.[301,317–320]

ICL1 is a validated Mtb drug target, and numerous inhibitor design studies have been carried out over the last fifteen years.[307,314,316–318,321–334] In addition, several natural product libraries have been docked to the ICL1 active site, including the Chinese Medicine Library and the Malaysian Local Natural Compound Database.[335–338] The majority of biochemically characterised ICL1 inhibitors are mechanism-based covalent inhibitors – analogues of isocitrate or succinate that react with the active site cysteine of ICL1 (**Figure 11**).[314,316,317,331–333,339] To date, the utility of these compounds has been limited by a lack of selectivity and poor mycobacterial cell permeability.[261,271,307,340]

Table 3: Kinetic parameters for Mtb ICL1.
For the first four entries (data from [311,318,319,329]), values for k_{cat}/K_M were calculated from the literature k_{cat} and K_M values. Experimental conditions (such as assay method, buffer, pH, and temperature) varied between studies. T, temperature; ns, not specified; Ref, reference.

Substrate	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (µM)	T (°C)	Ref
D-isocitrate	$2.7\times10^5 \pm 8.5\times10^4$	12.2 ± 0.6	45 ± 14	37	[318]
DL-isocitrate	$1.2\times10^5 \pm 1.9\times10^4$	5.2 ± 0.2	45 ± 7	ns	[319]
DL-isocitrate	$8.6\times10^3 \pm 5.0\times10^2$	9.78 ± 0.15	1142 ± 65	30	[311]
DL-isocitrate	$1.5\times10^4 \pm 6.2\times10^2$	4.3 ± 0.1	290 ± 10	27	[329]
DL-isocitrate	2.79×10^4	5.24	188	ns	[284]
DL-2-methylisocitrate	1.74×10^3	1.25	718	ns	[284]

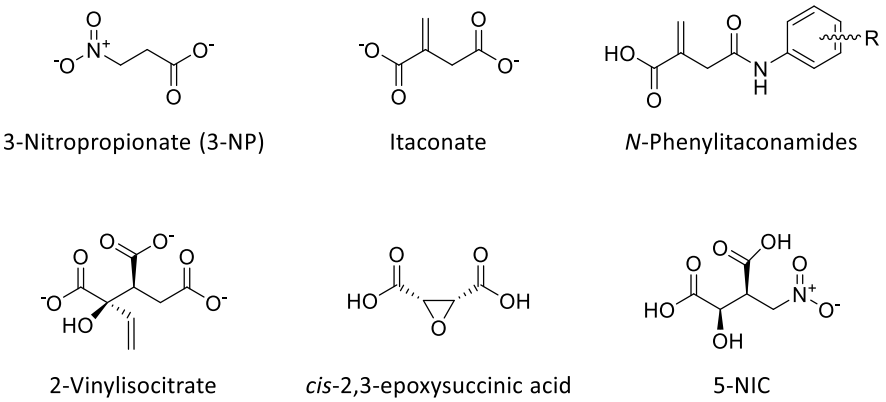


Figure 11: Examples of mechanism-based ICL1 inhibitors.
3-Nitropropionate (3-NP); itaconate; *N*-phenylitaconamides; 2-vinylisocitrate; *cis*-2,3-epoxysuccinic acid; 5-NIC ((2*R*,3*S*)-2-hydroxy-3-(nitromethyl)succinic acid). These compounds inhibit ICL1 by forming a covalent adduct with the active site cysteine. 2-Vinylisocitrate and 5-NIC initially undergo retro-aldol cleavage in the ICL1 active site, unmasking an electrophile (2-vinylglyoxylate and 3-NP respectively) which then reacts with the active site cysteine.[314,316,317,331–333]

1.6.3. Mtb isocitrate lyase 2 (ICL2)

Among the organisms that encode the glyoxylate shunt, mycobacteria are unusual in having two structurally distinct ICLs (**Section 1.5.3**). The second isocitrate lyase encoded by Mtb is the Group III enzyme ICL2 (*aceA*, *rv1915—1916*).[276,279,302] To date, ICL2 has been less studied than ICL1, and its role *in bacterio* remains unclear. This is in part due

to the presence of a frameshift mutation in *aceA* in the laboratory reference strain Mtb H37Rv, which splits the gene into two open reading frames, *rv1915* and *rv1916*.[276,302,341] Full length ICL2 has not been observed in Mtb H37Rv, as a result of which *rv1915—1916* is sometimes assumed to be a pseudogene.[276,302,303,342] Recombinant Rv1915 and Rv1916 each display ICL and MCL activities, and can associate *in vitro* to form a protein-protein complex with higher ICL and MCL activities than the constituent enzymes.[303,304,343] It is not yet known whether this complex forms *in bacterio*. Indeed, there is presently no evidence for the expression of Rv1916 *in bacterio* – *rv1915* mRNA and Rv1915 peptides have been observed, but *rv1916* mRNA and Rv1916 peptides have not.[304,344,345]

Some other strains of Mtb, such as CDC 1551 and Erdman, have a single frame *aceA* and can express a full length, functional ICL2.[276,302] Kinetic parameters have been determined for recombinant ICL2 from Mtb CDC 1551; these are shown in **Table 4**.[278,305,317] The enzyme has a lower affinity for isocitrate than Mtb ICL1 and turns over the substrate more slowly, observations which contributed to the view that ICL2 is subordinate to ICL1 in Mtb isocitrate metabolism.[276,278,279,302] This perspective has recently been revised, following the discovery that binding of Ac-CoA to the *C*-terminal domain of Mtb CDC 1551 ICL2 results in activation of the enzyme.[305]

In the presence of 25 μM Ac-CoA, the catalytic efficiency (k_{cat}/K_M) of ICL2 with respect to DL-isocitrate is 50 times higher than in the absence of Ac-CoA. Binding of Ac-CoA also activates the enzyme with respect to 2-methylisocitrate. For both DL-isocitrate and 2-methylisocitrate, the catalytic efficiency of ICL2 becomes comparable to that of ICL1 in the presence of Ac-CoA. Propionyl-CoA also activates both the isocitrate lyase and

methylisocitrate lyase activities of ICL2. These observations led Bhusal *et al*. to propose that the activation of ICL2 by Ac-CoA regulates flux through the glyoxylate shunt by increasing the total ICL activity when Ac-CoA levels are elevated.[305] However, the concentration of Ac-CoA in bacterial cells is typically in the range 20—600 μM.[346] When compared to the reported K_M of 2.9 μM for Ac-CoA binding to ICL2, this suggests that the Ac-CoA-bound, activated state may in fact be the usual form of ICL2 *in bacterio*.

Table 4: Kinetic parameters for Mtb CDC 1551 ICL2.
As noted above, ICL2 has a lower affinity for isocitrate (higher K_M) than ICL1 and turns over isocitrate more slowly (lower k_{cat}). However, detailed numerical comparisons are of limited value due to differences in the experimental conditions between literature reports. aa 1—605, ICL2 containing amino acids 1—605 (missing the *C*-terminal domain); *, parameters determined in the presence of 25 μM Ac-CoA; n.d., not determined; Ref, reference.

Length	Substrate	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (μM)	T (°C)	Ref
Full length	D-isocitrate	n.d.	n.d.	1300	37	278
aa 1—605	D-isocitrate	$1.7\times10^4 \pm 2.0\times10^3$	1.7 ± 0.05	100 ± 12	37	317
Full length	DL-isocitrate	755 ± 68	0.45 ± 0.1	606 ± 96	27	305
Full length	DL-isocitrate	300 ± 35	0.17 ± 0.1	570 ± 30	~21	305
Full length	DL-isocitrate*	$3.7\times10^4 \pm 6.2\times10^3$	4.6 ± 0.3	125 ± 27	27	305
Full length	DL-isocitrate*	$6.9\times10^3 \pm 1.3\times10^3$	1.7 ± 0.7	250 ± 80	~21	305
Full length	2-methylisocitrate	No reaction			27	305
Full length	2-methylisocitrate*	$1.2\times10^3 \pm 1.8\times10^2$	0.86 ± 0.1	717 ± 215	27	305

The ligand-free (PDB: 6EDW) and Ac-CoA-bound (PDB: 6EE1) crystal structures of ICL2 from Mtb CDC 1551 were published in 2019 and are the first structures solved for a Group III ICL.[305] Mtb CDC 1551 ICL2 crystallises as a homotetramer, with each chain consisting of 766 amino acids. As mentioned in **Section 1.6.1** above, the ligand-free crystal structure reveals that the *C*-terminal domain of ICL2 is structurally similar to a

GNAT acetyltransferase, but without the helices that comprise the acetyl acceptor binding site in these enzymes. The homotetramer adopts an elongated structure, with weak associations forming between the *C*-terminal domains of monomer units (**Figure 12**). In the presence of Ac-CoA, each *C*-terminal domain binds a molecule of Ac-CoA through hydrogen bonding. This triggers a significant conformational change in the ICL2 tetramer, as illustrated in **Figure 12**. New, more stable dimer interfaces are formed between the *C*-terminal domains of monomers that were previously on opposite ends of the structure. The rearrangement of the *C*-terminal domains causes the active site loop to shift into a more catalytically relevant position, resulting in enzyme activation.[303,305]

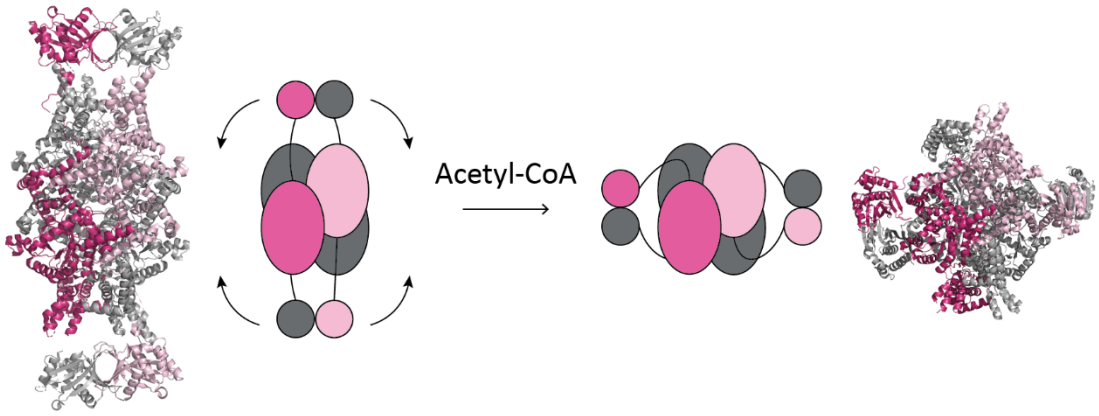


Figure 12: Conformational rearrangement of ICL2 on Ac-CoA binding.
The *C*-terminal domains of ICL2 monomers associate weakly in the absence of Ac-CoA. Binding of one molecule of Ac-CoA to each monomer triggers a conformational rearrangement of the tetramer. New, more stable dimer interfaces are formed between the *C*-terminal domains of monomers that were previously on opposite ends of the structure, and the active site loop moves into a more catalytically relevant position.[305]

1.6.4. ICLs from other mycobacterial species

Literature coverage of ICLs from mycobacteria other than Mtb is limited. Studies in the 1980s and 1990s identified glyoxylate shunt activity in *M. leprae*, *M. smegmatis*, *M. flavescens*, *M. microti*, *M. avium* and *M. vaccae*.[278,347–349] Regulation of the isocitrate

branch point has since been studied in *M. smegmatis*.[267,350–352] Additionally, various ICL inhibitors have been screened for binding to *M. ulcerans* ICL1 *in silico* and for antimycobacterial activity against *M. smegmatis*, *M. avium*, and *M. kansasii*.[336,353,354]

1.7. Isocitrate Dehydrogenases

1.7.1. Introduction to isocitrate dehydrogenases

Isocitrate dehydrogenases (IDHs) catalyse the oxidative decarboxylation of (2*R*,3*S*)-isocitrate to α-ketoglutarate (**Figure 13**), an ancient and ubiquitous metabolic reaction which evolved as part of the glutamate synthesis pathway before being incorporated into the TCA cycle.[355–357] They are dependent on either NAD^+ (EC 1.1.1.41) or $NADP^+$ (EC 1.1.1.42) and require a divalent cation (Mg^{2+} or Mn^{2+}) as a cofactor.[355,358–360] IDHs belong to the isocitrate/isopropylmalate dehydrogenase-like superfamily and can be divided into three phylogenetic groups, designated Types I, II and III.[355,361] The members of each group can be categorised based on the domain of life they are found in, together with their coenzyme specificity and multimeric state.[355] Examples of IDHs from each group are presented in **Table 5**. The general chemical mechanism of isocitrate decarboxylation by IDHs is shown in **Figure 14**.

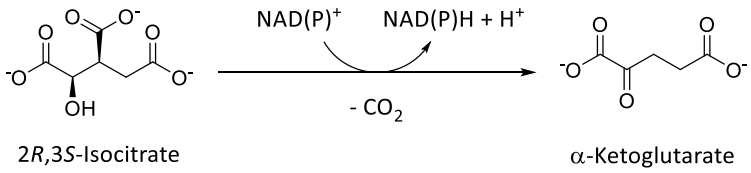


Figure 13: IDHs catalyse the decarboxylation of 2*R*,3*S*-isocitrate to α-ketoglutarate.

Table 5: Examples of enzymes from the three IDH subfamilies.
IDHs can be classified into three phylogenetic groups: Type I, Type II and Type III. Each group has members from different domains of life and with different oligomeric states and coenzyme specificities.[355]

Type of IDH		**Coenzyme**	**Example species**
Type I	Bacterial homodimer	$NADP^+$	*Escherichia coli*
			Salmonella enterica
		NAD^+	*Streptococcus mutans*
	Bacterial homotetramer	NAD^+	*Shewanella piezotolerans*
	Archaeal homodimer	$NADP^+$	*Thermoplasma acidophilum*
			Haloferax volcanii
	Eukaryotic (mitochondrial) heterooligomer	NAD^+	*Glycine max*
			Homo sapiens
			Saccharomyces cerevisiae
Type II	Bacterial homodimer	$NADP^+$	*Bifidobacterium longum*
			Mtb IDH1
	Bacterial homodimer or homotetramer	$NADP^+$	*Dictyoglomus turgidum*
			Thermotoga maritima
	Eukaryotic homodimer	$NADP^+$	*Homo sapiens*
			Zea mays
	Eukaryotic homodimer	NAD^+	*Ostreococcus tauri*
Type III	Bacterial monomer	$NADP^+$	*Azotobacter vinelandii*
			Corynebacterium glutamicum
			Mtb IDH2
	Bacterial monomer	NAD^+	*Campylobacter* sp. FOBRC14
	Archaeal monomer	$NADP^+$	*Methanococcoides methylutens*

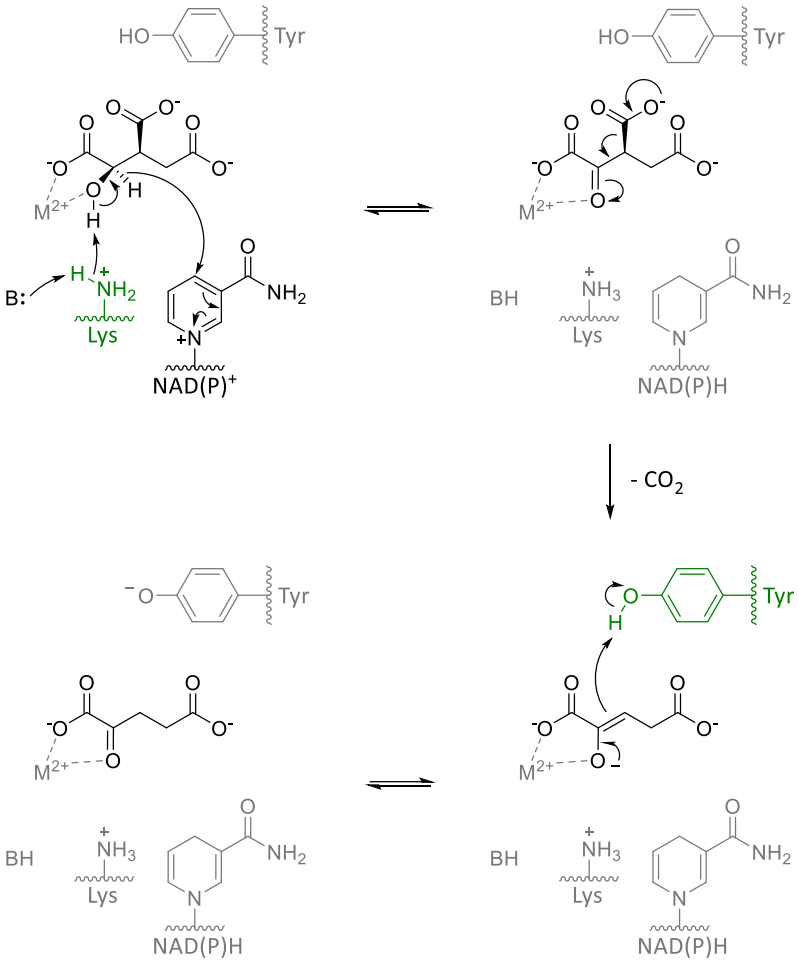


Figure 14: Suggested general chemical mechanism of isocitrate decarboxylation by IDHs.
The metal ion-dependent pyridine dinucleotide-linked β-hydroxyacid oxidative decarboxylases are a group of enzymes which includes IDHs, malate dehydrogenases, isopropylmalate dehydrogenases, homoisocitrate dehydrogenases, tartrate dehydrogenases, and 6-phosphogluconate dehydrogenases. It has been proposed that these enzymes share a general acid-base mechanism involving a lysine-tyrosine pair. A divalent metal ion (Mg^{2+} or Mn^{2+}) is coordinated by isocitrate (and other residues) in the active site. Deprotonation of lysine allows it to act as a general base and deprotonate isocitrate, prompting hydride transfer to $NAD(P)^+$. The residue responsible for lysine deprotonation varies between enzymes, but is often aspartate. Isocitrate decarboxylation follows, producing an enolate which is then protonated by tyrosine to generate α-ketoglutarate. The proposed mechanism is consistent with that suggested for Mtb IDH1; Mtb IDH2 has not been mechanistically characterised.[362–364]

Type I and Type II IDHs belong to the isocitrate/isopropylmalate dehydrogenase family of enzymes and are multimeric, with each chain containing around 400 amino acids.[355,365] The Type I subfamily includes eukaryotic mitochondrial heterooligomeric IDHs, bacterial homotetrameric NAD^+-dependent IDHs, archaeal homodimeric IDHs, and the majority of bacterial homodimeric IDHs.[355,360,365] Most Type II IDHs are

$NADP^+$-dependent homodimers from bacteria and eukaryotes. A limited number of NAD^+-dependent IDHs also belong to the Type II subfamily; these are all homodimeric IDHs from eukaryotic algae.[355,359,366] The best-known Type II IDHs are the human enzymes IDH1 and IDH2, mutations in which are associated with a variety of cancers, including glioblastoma and acute myeloid leukaemia.[367–369] These mutations usually affect a catalytic arginine residue and result in the production of the oncometabolite 2-hydroxyglutarate, which disrupts tissue homeostasis and gene regulation by competitively inhibiting α-ketoglutarate-dependent enzymes.[367,370,371]

Type III IDHs belong to the monomeric isocitrate dehydrogenase family. Containing around 740 amino acids, they are larger than the Type I and Type II IDHs and usually exist as monomers, although dimer formation has been observed in some bacterial species.[355,357,358,365,372,373] The family includes both NAD^+-dependent and $NADP^+$-dependent enzymes.[355] All reported Type III IDHs were from bacteria until 2020, when the first archaeal family member was identified in the marine species *Methanococcoides methylutens*. Putative Type III IDHs have since been annotated in several methylotrophic marine archaea, with sequence homology suggesting that these enzymes were acquired through horizontal gene transfer from marine bacteria.[365]

While members of different IDH subfamilies display less than 20% sequence similarity, their crystal structures show that the amino acid residues involved in isocitrate and metal ion binding are conserved, indicating that the three types of IDH evolved from a single ancestor.[374] In contrast, the residues comprising the NAD^+/$NADP^+$ binding site vary substantially between IDH subfamilies, indicating that coenzyme specificity evolved on at least three separate occasions.[361,374] $NADP^+$-dependent IDHs evolved from

ancestral NAD^+-dependent IDHs approximately 2 to 3.5 billion years ago, around the time that mitochondria developed.[361,375–377] In prokaryotes, it is thought that the adoption of $NADP^+$ as the coenzyme for IDH was to facilitate growth on acetate.[360,378] Carbon entering metabolism as acetate cannot be directed through the oxidative pentose phosphate pathway, which is the main source of NADPH during growth on glucose.[361] Bacterial species that encode only NAD^+-dependent IDHs are therefore unable to grow on acetate. The adaption of IDH to use $NADP^+$ provides an alternative source of NADPH, allowing biosynthesis to be sustained during growth on acetate.[375] For example, the $NADP^+$-dependent IDH from *E. coli* supplies approximately 90% of the cellular NADPH when acetate is the sole carbon source present.[361,375] Consequently, all prokaryotes that use the glyoxylate shunt to metabolise acetate possess at least one $NADP^+$-dependent IDH.[361] Mtb has two: IDH1 and IDH2.[277]

1.7.2. Mtb isocitrate dehydrogenase 1 (IDH1)

Mtb IDH1 (encoded by *icd1*, *rv3339c*) is a Type II homodimeric $NADP^+$-dependent IDH, in which each chain of 409 amino acids forms a Rossman fold and two β-hairpins.[277,355,379] A molecule of NADP(H) binds to the Rossman fold of each monomer, while Mn^{2+} binds at the dimer interface and is coordinated by residues from both monomers. Inhibition studies suggest that IDH1 catalysis follows an ordered sequential mechanism in which $NADP^+$ binds first ($K_M(NADP^+) = 15 \pm 3$ μM), followed by isocitrate (K_M(isocitrate) = 50 ± 7 μM). After the reaction, CO_2 is released first, α-ketoglutarate second, and NADPH last.[363]

1.7.3. Mtb isocitrate dehydrogenase 2 (IDH2)

Mtb IDH2 (encoded by *icd2*, Rv0066c) is a Type III monomeric $NADP^+$-dependent IDH, with each monomer comprising 745 amino acids.[277,380] Comparing IDH2 to IDH1 reveals low amino acid sequence similarity (~14%), as expected for members of different IDH subfamilies.[379] Indeed, IDH2 is more similar to other monomeric IDHs, such as those from *Corynebacterium glutamicum* (56%) and *Azotobacter vinelandii* (60%).[381] IDH2 crystallises as a monomer (PDB: 5KVU), but is among the monomeric IDHs observed to oligomerise in solution (**Section 1.7.1**). Dimers, trimers, and tetramers of IDH2 have been observed; the dominant state varies with the experimental conditions.[379,381] The catalytic mechanism has yet to be determined.

1.7.4. IDHs from other mycobacterial species

Studies in the 1980s and 1990s demonstrated IDH activity in a number of other mycobacterial species,[e] but did not explicitly distinguish between the two isoenzymes.[382–385] While genome annotation indicates that not all mycobacterial species encode both IDHs (for example, *M. smegmatis* has only IDH2), the distribution of the two isoenzymes among mycobacterial species has yet to be analysed. Of the non-tuberculous mycobacterial IDHs, only three IDH2s (*M. bovis* BCG, *M. smegmatis* and *M. phlei*[f]) have been expressed and purified, and only *M. smegmatis* IDH2 has been crystallised.[267,386–388] Kinetic and structural data for mycobacterial IDHs are consequently extremely limited.

[e] *M. leprae*, *M. lepraemurium*, *M. avium*, *M. bovis* BCG, *M. fortuitum*, *M. gastri*, *M. intracellulare*, *M. kansasii*, *M. phlei*, *M. senegalense*, *M. simiae*, *M. smegmatis*, and *M. vaccae*.

[f] Although not explicitly identified as such, this enzyme is an IDH2; *M. phlei* does not possess IDH1.

1.8. Control of flux through the isocitrate branch point

To recap, changes in carbon metabolism are an important part of the mycobacterial response to intracellular environments.[257,389] Control of flux at the isocitrate branch point is particularly critical for Mtb during infection, when diversion of flux through the glyoxylate shunt is required to accommodate a shift in carbon availability from carbohydrates to fatty acids (**Section 1.5**). In Mtb, four enzymes contribute to flux through the isocitrate branch point: ICL1 (**Section 1.6.2**), ICL2 (**Section 1.6.3**), IDH1 (**Section 1.7.2**) and IDH2 (**Section 1.7.3**). Altering the flux balance between the TCA cycle and the glyoxylate shunt requires changing the ability of one or more of these enzymes to compete for isocitrate.[261,278] Potential regulatory mechanisms therefore fall into two broad categories: control of enzyme concentrations (by modulating transcription, translation or degradation) and control of enzyme kinetic parameters (through post-translational modification (PTM) and activation or inhibition by metabolites).

1.8.1. Control of the ICL1 concentration through transcriptional regulation of *icl*, *rv0467*

As previously noted (**Section 1.5.5**), *icl* is the only gene from Mtb isocitrate metabolism to be significantly upregulated in response to macrophage infection.[256] *Icl* is also upregulated under various *in vitro* conditions relevant to infection, including hypoxia, oxidative stress, and growth on acetate or fatty acids.[256,390,391] Control of the ICL1 concentration through transcriptional regulation of *icl* is therefore a regulatory mechanism for the isocitrate branch point in Mtb. The local regulator of *icl* transcription in Mtb is Rv0465c, which is also known as LrpI (local regulatory protein of *icl*) or RamB (after the homologous protein in *Corynebacterium glutamicum*).[261,392] RamB does not regulate *rv1915—1916* (*aceA*) or any other genes involved in acetate metabolism, but

does exhibit negative autoregulation.[390] The reported mechanism for RamB-mediated control of *icl* transcription varies according to the stimulus or stressor applied; it appears that RamB functions as both a transcriptional activator and a transcriptional repressor in Mtb.[390,391]

1.8.1.1. RamB represses *icl* transcription during growth on glucose

Icl mRNA levels in wild-type Mtb H37Rv are significantly higher when acetate is the sole carbon source than when glucose is the sole carbon source (5.9-fold after 24 hours). Transcriptomics and enzymology experiments using a *ramB* deletion mutant indicate that this phenotype results from RamB-mediated repression of *icl* transcription in the presence of glucose. Specifically, deletion of *ramB* does not affect levels of *icl* mRNA or ICL activity in H37Rv grown on acetate. During growth on glucose, however, the Δ*ramB* mutant has significantly higher *icl* mRNA levels and ICL activity than wild-type H37Rv – both *icl* mRNA and ICL activity reach levels comparable to those seen during growth on acetate. This pattern suggests that the increase in *icl* transcription seen when the carbon source is changed from glucose to acetate is the result of de-repression rather than activation.[390] In contrast, RamB functions as a transcriptional activator during hypoxia, in conjunction with the global regulator σ^B.[391]

1.8.1.2. RamB and σ^B jointly activate *icl* transcription in hypoxic culture

Bacterial RNA polymerases have six subunits: a five-subunit core enzyme and a sigma factor. The sigma factor determines the promoter specificity of the polymerase; changing the sigma factor changes the set of genes which are transcribed.[393,394] One of 13 sigma factors encoded by Mtb, σ^B (Rv2710, SigB) is expressed in response to stressors such as hypoxia, oxidative stress and cell envelope stress – all conditions under which

the glyoxylate shunt is employed.[277,394,395] Expression of σ^B is regulated by another sigma factor, σ^E (Rv1221, SigE), and both σ^B and σ^E are conserved in mycobacteria.[394,396] Under hypoxic conditions, transcriptional regulation of *icl* involves a nested feed-forward loop: σ^B controls transcription of *ramB*, then σ^B and RamB jointly positively regulate the transcription of *icl*.[391]

However, this coherent nested feed-forward loop does not fully explain the experimentally observed transcription dynamics during hypoxia.[397] Specifically, *icl* and *ramB* mRNA levels exhibit non-monotonic dynamics, rising to a peak on day 3 and returning to pre-induction levels by day 4, while *sigB* mRNA levels are monotonic (continually increasing) during this time period.[391,397] To produce this pattern, some form of negative regulation must be taking place in addition to the positive regulatory interactions that have previously been described. Combining transcriptomics and knockout studies with mathematical models suggests that this negative regulation is post-translational, with the primary target being σ^B. Proteolytic degradation of σ^B, for example by Clp protease, has been proposed as the mechanism.[397]

1.8.1.3. PrpR activates *icl* transcription during growth on propionate

In the context of the methylcitrate cycle, *icl* transcription is controlled by PrpR (propionate regulator, Rv1129c), a paralog of RamB. PrpR is required for Mtb growth on propionate (a methylcitrate cycle substrate), but is not required for growth on glucose or acetate. Expression of *prpR* is carbon source dependent, and is highest when propionate is the sole carbon source (on propionate, *prpR* expression is 7 times higher than on glucose and 30 times higher than on acetate). During growth on propionate, PrpR activates transcription of *icl* and two other methylcitrate cycle genes: *prpD*

(methylcitrate dehydratase) and *prpC* (methylcitrate synthase). PrpR also represses transcription of *ramB*; it has been suggested that this serves to prevent any repression of *icl* transcription by RamB.[398]

1.8.1.4. Additional regulatory mechanisms may be involved in the presence of other stressors

Studies in *M. smegmatis* have identified additional transcription factors that contribute to regulation of *icl* transcription. In *M. smegmatis*, *icl* transcription is repressed by RegX3 (part of the phosphate-sensing two-component system SenX3-RegX3) during phosphate starvation, by the global nitrogen metabolism regulator GlnR under nitrogen-limiting conditions, and by cAMP receptor protein (Crp) during growth on glucose.[350–352,399] In all three cases, it is thought that repression of *icl* transcription results from obstruction of the RamB binding sites (in contrast to Mtb, RamB is consistently reported to be a transcriptional activator in *M. smegmatis*).[350] However, the contributions of these mechanisms to the regulation of *icl* transcription in Mtb have yet to be examined.

1.8.2. Control of enzyme activities using post-translational modifications

1.8.2.1. Phosphorylation of IDHs

Post-translational modification (PTM) of IDH can be used to control carbon flux through the isocitrate branch point. In *E. coli*, for example, the flux distribution between the TCA cycle and the glyoxylate shunt is determined by the extent of phosphorylation of IDH on a serine residue (S113) in the active site. The negative charge introduced by this PTM creates electrostatic repulsion that prevents isocitrate binding, rendering the enzyme catalytically inactive.[400] As the proportion of phosphorylated IDH increases, the total IDH activity decreases. *E. coli* ICL (which has a lower affinity for isocitrate than *E. coli* IDH) then becomes able to compete with IDH for isocitrate, resulting in increased flux

through the glyoxylate shunt.[401] Phosphorylation of *E. coli* IDH is reversible, with both reactions being catalysed by the kinase/phosphatase AceK.[402] The direction of AceK catalysis is allosterically regulated by various metabolites, most importantly isocitrate, NADPH, phosphoenolpyruvate and adenine nucleotides (AMP, ADP and ATP).[403]

As Mtb does not possess AceK, IDH activity must be regulated differently than in *E. coli*. Nonetheless, serine phosphorylation has been hypothesised as a mechanism for the regulation of Mtb IDH activity. Of the 11 serine/threonine phosphatase/kinases (STPKs) encoded by Mtb, protein kinase G (PknG) was determined to be the most likely alternative to AceK.[277,404] It has been shown that PknG can phosphorylate recombinant IDH1 *in vitro*, resulting in inactivation of the latter.[404] Molecular dynamics simulations identified several IDH1 serine residues as possible phosphorylation sites, but these have not been confirmed experimentally.[405] IDH2, like other monomeric IDHs, is thought to be phosphorylation insensitive.[267,358,401] Structural studies of IDH2 from *M. smegmatis* predict that the residues corresponding to the *E. coli* AceK recognition sequence would be sterically inaccessible to a kinase.[267]

Phosphoproteomics experiments have yet to identify either IDH as a substrate of PknG *in bacterio*.[406–409] Additionally, deletion experiments showed that (i) *pknG* is dispensable for the growth of Mtb and *M. smegmatis* on acetate, and (ii) deletion of *pknG* does not alter the total IDH activity in cell-free extracts of Mtb and *M. smegmatis* grown on glucose or acetate. These observations are contrary to what one would expect if PknG mediated isocitrate branch point flux in an analogous manner to AceK.[267] Finally, irrespective of *pknG* status (wild type or Δ*pknG*), the total IDH activity in cell-free extracts of Mtb, *M. bovis* BCG and *M. smegmatis* was higher when acetate was used as

the carbon source than when glucose was used as the carbon source. This finding indicates that flux through the glyoxylate shunt in mycobacteria is not facilitated by IDH inactivation, excluding PknG-mediated phosphorylation as the primary mechanism of flux control at the isocitrate branch point.[267]

1.8.2.2. Acetylation of IDHs

Phosphorylation is not the only PTM known to influence IDH activity. Depending on which lysine residue is modified, acetylation of *E. coli* IDH can significantly enhance or completely eliminate its activity.[410] Lysine acetylation affects IDH activity in plants and in human mitochondria, and acetylation of both Mtb IDHs has been observed in proteomics experiments.[411–414] The conserved mycobacterial protein Rv2170, which contains a GNAT domain, is able to acetylate recombinant Mtb IDH1, but not recombinant Mtb IDH2. IDH1 acetylation occurs at two positions, K30 and K129, and reduces the activity of the enzyme by 70%. However, deletion of *rv2170* does not affect the growth of Mtb on either TCA cycle substrates (glucose and glycerol) or glyoxylate shunt substrates (acetate and stearic acid), suggesting that it is not essential for flux control at the isocitrate branch point.[415] As noted in **Section 1.8.2.1** above, IDH inactivation by PTM does not appear to be a dominant regulatory mechanism for flux through the mycobacterial isocitrate branch point.

1.8.2.3. Acetylation of ICL1

ICL1 has been identified in the acetylomes of Mtb H37Rv and H37Ra. Acetylation of K322 decreases both the ICL activity and the MCL activity of ICL1, while acetylation of K392 increases the ICL activity and does not affect the MCL activity.[414,416] ICL1 acetylated on K392 is thought to be less susceptible to degradation than unacetylated or

K322-acetylated ICL1 – the level of K392-acetylated ICL1 does not decrease in extended stationary phase culture, when total ICL1 and K322-acetylated ICL1 levels decrease significantly.[414] Acetylation may therefore be involved in the long-term control of ICL1 activity, following on from the initial transcriptional response discussed in **Section 1.8.1**. Other PTMs may also influence ICL1 activity – succinylation of ICL1 has been observed in Mtb, while mycothiolation of ICL1 has been observed in *M. smegmatis*.[417–419]

1.8.3. Enzyme activation by metabolites

1.8.3.1. Activation of IDH2 by glyoxylate

Glyoxylate, one of the products of ICL, binds close to the active site of *M. smegmatis* IDH2, resulting in activation of the enzyme. IDH2 activation is dose-dependent for glyoxylate concentrations in the range 0—200 μM, concentrations which are thought to be in the physiological range. Allosteric cross-activation of IDH2 explains the (initially counterintuitive) observation that the total IDH activity in several mycobacterial species increases during growth on acetate, despite the fact that the IDH protein abundance is unchanged.[267] However, the effect of glyoxylate on Mtb IDH2 has yet to be confirmed.

1.8.3.2. Activation of ICL2 by Ac-CoA

As discussed in **Section 1.6.3**, binding of Ac-CoA or propionyl-CoA to the *C*-terminal domain of ICL2 results in enzyme activation, leading to the suggestion that allosteric regulation of ICL2 by Ac-CoA contributes to the control of flux through the glyoxylate shunt.[305] However, as typical concentrations of Ac-CoA in bacteria are 10 to 100 times higher than K_M(Ac-CoA) for ICL2, the activated state may in fact be the dominant form of the enzyme *in bacterio*.[305,346,420] The activation of ICL2 by Ac-CoA highlights an important consideration: where enzyme-metabolite interactions have been identified

as potential regulatory mechanisms *in vitro*, critical evaluation of their physiological relevance is required. Such evaluation requires an understanding of metabolite pool sizes and how these change in response to conditions of interest. To contribute to flux control, the metabolite pool size must be variable in a kinetically relevant concentration range. While the pool sizes of around 100 core metabolites have been quantified in *E. coli* grown with glucose, glycerol, and acetate as the sole carbon source, equivalent data are not yet available for mycobacteria.[420]

1.8.4. Regulation of the isocitrate branch point in Mtb

Having reviewed various mechanisms for the control of flux through the four isocitrate metabolising enzymes in Mtb (summarised in **Figure 15**), we can conclude that regulation of the isocitrate branch point is complex and incompletely understood. Although the mechanisms regulating flux through the isocitrate branch point in Mtb and the ways in which these mechanisms interact have yet to be fully elucidated, the environmental conditions that prompt changes in isocitrate flux are well established. Diversion of flux through the glyoxylate shunt is an essential part of the mycobacterial response to the challenges of the intracellular environment, most notably changes in carbon source availability (**Section 1.5**). We will therefore finish by comparing the major carbon sources in infected macrophages and in soil, to explore the demands exerted by host and non-host niches on mycobacterial isocitrate metabolism.

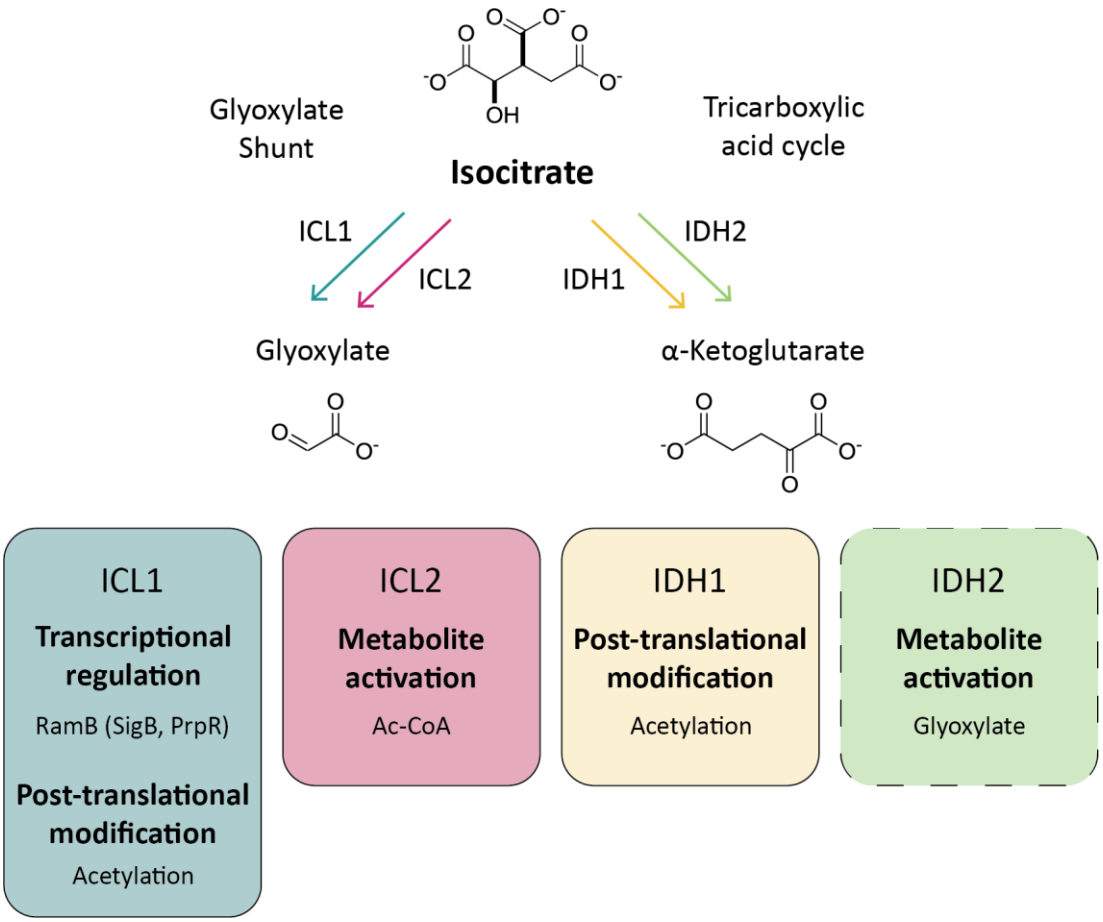


Figure 15: Mechanisms contributing to the regulation of the isocitrate branch point in Mtb.
ICL1: A key mechanism for increasing flux through the glyoxylate shunt is increasing the concentration of ICL1 by increasing the transcription of *icl*. In the context of the glyoxylate shunt, *icl* transcription is primarily controlled by the local regulator RamB, which is in turn influenced by other transcription factors such as SigB (**Section 1.8.1**). In the context of the methylcitrate cycle, *icl* transcription is regulated by PrpR (**Section 1.8.1.3**). Post-translational control of ICL1 activity can be achieved through acetylation (**Section 1.8.2.3**). **ICL2:** To date, the only potential regulatory mechanism identified for ICL2 is activation by Ac-CoA, the physiological relevance of which is not known (**Section 1.8.3.2**). **IDH1:** Post-translational modification of IDH1 homologues forms the basis of isocitrate flux control in many organisms, but this is not the case in mycobacteria (**Section 1.8.2.1**). Although IDH1 activity can be modulated through acetylation, it is unclear to what extent this PTM is involved in regulation of flux through the isocitrate branch point (**Section 1.8.2.2**). **IDH2**: In *M. smegmatis*, the total flux through the isocitrate branch point increases when the glyoxylate shunt is active, a consequence of allosteric cross-activation of IDH2 by glyoxylate. The effect of glyoxylate on Mtb IDH2 is presently unknown (**Section 1.8.3.1**).

1.9. Host and non-host niches are nutritionally distinct

1.9.1. Lipids are the dominant carbon sources in infected macrophages

As discussed earlier, studies of Mtb metabolism during infection indicate that the macrophage is a fatty acid rich, carbohydrate poor environment (**Section 1.5.1**).[253,256]

This perspective is supported by studies of the macrophage response to Mtb infection.

Once inside an alveolar macrophage, Mtb cell wall components trigger various signalling pathways, prompting a two-stage remodelling of macrophage carbon metabolism.[421,422] The first stage lasts for around 8 hours after infection and is similar to the Warburg effect seen in cancer cells. Glucose uptake and aerobic glycolysis are increased, leading to the production of lactate, while oxidative phosphorylation is impaired.[422–426] Interruption of the mitochondrial TCA cycle at IDH leads to the accumulation of citrate, which is used to produce Ac-CoA for the synthesis of long-chain fatty acyl-CoAs and prostaglandins.[421,422]

By 24 hours after infection, the second phase of remodelling is established. Glucose uptake and glycolysis are shut down, and the macrophage accumulates triacylglycerols, phospholipids, cholesterol and cholesteryl esters, becoming increasingly lipid-rich.[421,422] Continued lipid accumulation leads to the development of a "foamy" phenotype.[102,427] As infection progresses, foamy macrophages become part of granulomas (organised collections of macrophages and other immune cells).[428–430] On macrophage death, Mtb and lipids are released into the centre of the granuloma, from where they enter sputum when the granuloma eventually ruptures.[427,431] Mtb is therefore exposed to a lipid-rich environment throughout its time in the host niche.[421,427]

NMR-based metabolomic profiling has been used to explore how chronic Mtb infection affects the concentrations of other potential carbon sources in infected mouse and guinea pig tissues, most importantly acetate and lactate.[432–434] Acetate progressively accumulates in guinea pig granulomas during chronic infection (15—60 days post-infection) with Mtb H37Rv or W-Beijing, and the acetate concentration is significantly elevated compared to controls at all time points.[432,433] The concentration of lactate

depends on the infecting Mtb strain, increasing during chronic H37Rv infection but decreasing during chronic W-Beijing infection.[432–434] These results highlight that carbon sources other than lipids are likely accessible to Mtb *in vivo*. Nonetheless, ^{13}C labelling studies indicate that the carbon sources metabolised by intracellular Mtb enter central carbon metabolism primarily as acetate or Ac-CoA, and so will ultimately be metabolised through the glyoxylate shunt.[435]

1.9.2. Sugars are the dominant carbon sources in soil

1.9.2.1. Sources and availability of organic carbon in soil

Sources of organic compounds in soil include microorganisms and their metabolites; plant root exudates; and decaying plant and animal tissues.[436] The soil organic carbon content varies geographically and is affected by factors such as the soil type, vegetation, land use, and climate (temperature, precipitation and humidity).[437,438] For example, arable land has a lower average organic carbon content (2.0%) than forests (7.5%), moorlands (7.2%), or wetlands (13.5%).[437] Importantly, not all of the organic carbon present in soil is assimilable by bacteria. Incompletely degraded biopolymers are too large for bacterial transporters, which usually accept substrates smaller than 600 kDa, while molecules small enough for transport may be rendered inaccessible by aggregation or adsorption to inorganic surfaces.[439–441] Potential bacterial carbon sources are therefore limited to the dissolved organic matter, a collection of soluble low molecular weight organic compounds consisting mainly of monosaccharides, disaccharides, organic acids, and amino acids.[442] Sugars are 2—3 times more abundant in dissolved organic matter than organic acids and amino acids, and are consequently the major assimilable carbon source for soil bacteria.[442]

1.9.2.2. Simple sugars: glucose is the most abundant assimilable carbon source in soil

Around 87% of European soils have a low to moderate organic carbon content (<20%) and are classified as mineral soils.[438,442,443] In mineral soils, sugars account for 10 ± 5% of the total organic carbon content and 30% of the dissolved organic matter.[442,444] While the total sugar content is a relatively consistent proportion of both total and dissolved organic carbon, the sugar profile depends on the soil type and land use.[442,445] Glucose is always present and is the most abundant sugar, comprising 30% of the sugar content of dissolved organic matter.[442,446] There are two reasons for the dominance of glucose. Firstly, glucose is the decomposition product of cellulose, the most abundant biopolymer on Earth, and is also prevalent in other plant polysaccharides.[442,447] Secondly, plant root exudates can be up to 50% glucose, depending on the species.[442] Other sugars and amino-sugars frequently identified in soil include galactose, arabinose, mannose, xylose, fructose, fucose, rhamnose, ribose, glucosamine, galactosamine and mannosamine.[445] Collectively, sugars contribute 5—130 mg of carbon per kilogram of soil. Measured concentrations of individual sugars depend on the ecosystem and the extraction method, but values in the range 100—4000 μg/g soil are commonly reported.[445]

1.9.2.3. The organic acids present in soil include glyoxylate shunt substrates

Comprising 0.9—10% of dissolved organic carbon, organic acids are less abundant in soil than sugars.[445,448,449] As the organic acids in soil come from plant root exudates and microorganisms, the organic acid content of the soil is ecosystem-dependent.[445] The organic acids present may include acetic, adipic, benzoic, butyric, caffeic, citric, ferulic, formic, gallic, gentisic, glycolic, isocitric, lactic, malic, malonic, *m*-hydroxybenzoic, oxalic, *p*-coumaric, phenylacetic, phthalic, *p*-hydroxybenzoic, propionic, protocatechuic,

shikimic, sinapic, succinic, syringic, tartaric, tosmarinic, *trans*-aconitic, *trans*-cinnamic, and vanillic acid.[445] Although individual acid concentrations may be up to 6 mM, concentrations of monovalent acids (such as acetic, formic and lactic acid) are more commonly in the range 1—1000 μM, while concentrations of divalent and trivalent acids (such as citric, malic and oxalic acid) are typically in the range 0.1—50 μM.[445,450,451] Some organic acids, such as *trans*-cinnamic acid and caffeic acid, are bacteriostatic.[452] Others, like acetic acid and lactic acid, can be used as carbon sources by a range of microorganisms.[453–456]

1.9.2.4. Amino acids are usually considered as nitrogen sources in soil

Amino acids in soil are produced by the decomposition of plant and animal proteins and collectively account for 30% of soil nitrogen.[457] Individual amino acids are present at concentrations between 0.1 and 5 μM.[450] Amino acids in soil are usually considered in their role as nitrogen sources, as soil respiration studies indicate that the proportion of amino acid carbon metabolised to CO_2 is low compared to other carbon sources.[450,457,458]

1.9.2.5. Some bacterial species are able to metabolise hydrocarbons

Some soils are contaminated with hydrocarbons as a result of human activities such as oil drilling.[459,460] Various soil microorganisms are able to degrade these hydrocarbons and have gained attention due to their potential use as bioremediation agents.[460–462] As noted in **Section 1.3.1.1**, several mycobacterial species were originally isolated from oil-contaminated soils and can use straight-chain aliphatic and/or polyaromatic hydrocarbons as their sole carbon source.[463–470]

1.9.3. Changes in isocitrate metabolism may be associated with mycobacterial niche adaptation

Sugars, which are metabolised through the TCA cycle, are the dominant carbon sources available to mycobacteria in soil (**Section 1.9.2.2**). However, glyoxylate shunt substrates – organic acids and hydrocarbons – are also typically present (**Sections 1.9.2.3 and 1.9.2.5**). Unlike *E. coli*, which uses catabolite repression to sequentially consume its preferred carbon sources, mycobacteria can co-catabolise carbon substrates.[291,435,471,472] It follows that soil mycobacteria encoding the glyoxylate shunt may use sugars, organic acids, and hydrocarbons contemporaneously. If this is the case, mycobacterial glyoxylate shunt activity in soil would be higher than that observed in laboratory media with glucose (or another sugar) as the sole carbon source. Nonetheless, the abundance of glucose in soil means that environmental saprophytic mycobacteria are unlikely to encounter glyoxylate shunt substrates as their primary carbon source.

In contrast, the ability to subsist primarily on glyoxylate shunt substrates is critical for the survival of Mtb in the host. As discussed earlier, studies of Mtb and macrophage metabolism have established fatty acids and cholesterol as the canonical carbon sources for Mtb *in vivo*.[253,256,389,473–477] More recent work indicates that other carbon sources, notably acetate and lactate, are also accessible to Mtb residing in the host niche.[432–434,453] Since carbon entering the central metabolism of intracellular Mtb does so primarily as acetate or Ac-CoA, metabolism of the carbon sources abundant *in vivo* ultimately requires ICL.[435] Deletion studies confirm that ICL activity is essential during infection (**Section 1.5.5**).[259,279] Since the glyoxylate shunt is essential for mycobacterial growth in host niches but is not required for the metabolism of the sugars abundant in

soil, changes in isocitrate metabolism may occur as part of the process of host adaptation. Specifically, host-adapted mycobacterial species (such as Mtb) might have evolved to be more efficient at metabolising glyoxylate shunt substrates than environmental saprophytic mycobacteria. We therefore hypothesise that mycobacterial species adapted to nutritionally distinct ecological niches will exhibit differences in isocitrate metabolism.

1.10. A genus-level comparative study of mycobacterial isocitrate metabolising enzymes

1.10.1. Do mycobacterial species adapted to nutritionally distinct ecological niches exhibit differences in isocitrate metabolism?

TB and other mycobacterial diseases have significantly impacted on human health and society throughout recorded history, and continue to do so today (**Section 1.1**). Since the causative agents of TB and leprosy were identified and classified as mycobacteria, mycobacteriologists have sought to determine why some mycobacterial species cause disease while others are innocuous, despite being frequently encountered due to their ubiquity in natural and built environments. This question is becoming increasingly pressing due to the rise in antibiotic resistant and non-tuberculous mycobacterial infections (**Section 1.2**). Pathogenic mycobacteria have evolved molecular mechanisms to engage, exploit and elude their hosts – that is, they have become host-adapted. In contrast, host-pathogen interaction mechanisms may be absent or less effective in mycobacterial species adapted to non-host niches such as soil (**Section 1.3**). Consequently, we consider that comparison of mycobacterial species adapted to different ecological niches will be of use in elucidating the molecular mechanisms underlying host adaptation. We propose that studying mycobacteria at the genus level

will maximise the power of comparison by effectively balancing genetic similarity with phenotypic diversity. To this end, we have assembled the Mycobacterial Species Library, a collection of 74 mycobacterial species which we believe to be representative of the genus as a whole (**Section 1.4**).

In Mtb, changes in metabolic fluxes form an essential part of the response to host niches. Control of the fate of isocitrate, which sits at the branch point between the TCA cycle and the glyoxylate shunt, is particularly crucial during infection, when flux must be diverted through the glyoxylate shunt to process the Ac-CoA produced from the metabolism of fatty acids and cholesterol (**Section 1.5**). Four enzymes contribute to flux through the isocitrate branch point in Mtb: ICL1, ICL2, IDH1 and IDH2 (**Sections 1.6 and 1.7**). These enzymes have been kinetically characterised, but their regulation is incompletely understood (**Section 1.8**).

Host and non-host niches exert different demands on the isocitrate branch point as a consequence of their distinct organic carbon content (**Section 1.9**). The glyoxylate shunt is essential for Mtb survival inside macrophages, but dispensable for growth on TCA cycle substrates such as glucose (**Section 1.5**), which are the dominant carbon sources in soil (**Section 1.9.2**). It follows that changes in isocitrate metabolism may be associated with mycobacterial niche adaptation. In particular, host-adapted mycobacterial species may have evolved a more efficient glyoxylate shunt as a result of the selection pressure exerted by the host niche (**Section 1.9.3**). Herein, we will evaluate whether mycobacterial species adapted to nutritionally distinct ecological niches exhibit differences in isocitrate metabolism by conducting a genus-level comparative study of mycobacterial isocitrate metabolising enzymes.

1.10.2. Project aims

In the following chapters, we will combine bioinformatics and genus-level comparative enzymology to explore how the structure and kinetic function of the isocitrate branch point varies between mycobacterial species. First, we will **establish how the composition of the isocitrate branch point varies across the genus *Mycobacterium*** (**Chapter 2**). We know that Mtb and *M. smegmatis* have different isocitrate branch point structures – specifically, *M. smegmatis* lacks IDH1. This observation raises the possibility that the presence or absence of individual isocitrate metabolism genes could contribute to differences in branch point function between species. An understanding of branch point composition will therefore be essential for the design and interpretation of comparative experiments using other mycobacterial species, in which isocitrate metabolism is presently uncharacterised.

Second, we will **determine whether the frameshift mutation in Mtb *aceA* that creates *rv1915—1916* is restricted to the laboratory-adapted strains H37Rv and H37Ra** (**Chapter 3**). In the frequently studied Mtb strains H37Rv and H37Ra, *aceA* is often assumed to be a pseudogene because it is split into two open reading frames, *rv1915* and *rv1916*.[276,302] However, translation of the single-frame *aceA* encoded by Mtb CDC 1551 produces a functional ICL2, which has been hypothesised to be important in regulating isocitrate flux.[305] Determining whether the frameshift mutation that creates *rv1915—1916* is widespread among Mtb strains or simply a consequence of laboratory adaptation will help us to more clearly define the role of ICL2 *in bacterio*. The status of ICL2 will also be an important consideration when selecting an Mtb strain for comparison to other mycobacterial species.

Third, we will draw on the literature to **establish methods for the preparation and kinetic characterisation of mycobacterial isocitrate metabolising enzymes in our laboratory** (**Chapter 4**). Working with the four branch point enzymes from Mtb CDC 1551, we will focus on reproducibility and comparability to ensure that variation across the subsequent genus-level dataset results from differences between the enzymes and not the experimental conditions.

Finally, we will **clone, express, purify and kinetically characterise the isocitrate metabolising enzymes from a representative subset of the Mycobacterial Species Library** (**Chapter 5**). Comparison of these enzymes to each other and to the Mtb enzymes characterised in **Chapter 4** will reveal whether isocitrate metabolism kinetics vary across the genus *Mycobacterium*. Collectively, these experiments will constitute the first detailed examination of isocitrate metabolism in non-tuberculous mycobacteria. They will allow us to evaluate whether changes in isocitrate metabolism are associated with mycobacterial niche adaptation, or whether highly host-adapted species such as Mtb exploit enzymatic activity that is available across the genus.

Chapter 2. Distribution of isocitrate metabolising enzymes in the genus *Mycobacterium*

2.1. Chapter introduction

Four distinct enzymes contribute to flux through the isocitrate branch point in Mtb: ICL1 (encoded by *icl*, *rv0467*), ICL2 (encoded by *aceA*, *rv1915—1916*), IDH1 (encoded by *icd1*, *rv3339c*), and IDH2 (encoded by *icd2*, *rv0066c*).[277] We reviewed these enzymes in **Chapter 1**. In *M. smegmatis*, meanwhile, the isocitrate branch point consists of three enzymes: ICL1 (MSMEG_0911), ICL2 (MSMEG_3706), and IDH2 (MSMEG_1654). As we noted in our review of mycobacterial isocitrate metabolism, little is known about the ICLs and IDHs from other mycobacterial species (**Chapter 1, Sections 1.6.4 and 1.7.4**). To study mycobacterial isocitrate metabolism at the genus level, we first need to understand how the composition of the isocitrate branch point varies between mycobacterial species. Accordingly, this chapter will examine how ICL1, ICL2, IDH1, and IDH2 are distributed in the genus *Mycobacterium*.

Further details of the work described in this chapter are provided in **Chapter 7**. Throughout this chapter, the terms 'fast growers' and 'slow growers' will be used to refer to the major phylogenetic divisions in the genus *Mycobacterium*, rather than the absolute growth rate of any given species *in vitro* (as discussed in **Chapter 1, Section 1.2**).[478] Mycobacterial clades (such as *Abscessus-Chelonae*) are named as described in **Chapter 1, Section 1.2**.

2.2. Distribution of ICL1 and ICL2 in the genus *Mycobacterium*

Mycobacterial ICLs were identified by BLAST searching, as described in **Chapter 7**. The distribution of ICL1 and ICL2 in the genus *Mycobacterium* is illustrated in **Figure 16**. All 129 mycobacterial species analysed possess ICL1, and most mycobacterial species also have ICL2. The species without ICL2 form two phylogenetic clusters, shaded blue in **Figure 16**: the *Abscessus-Chelonae* clade and a group of species closely related to *M. fallax*.

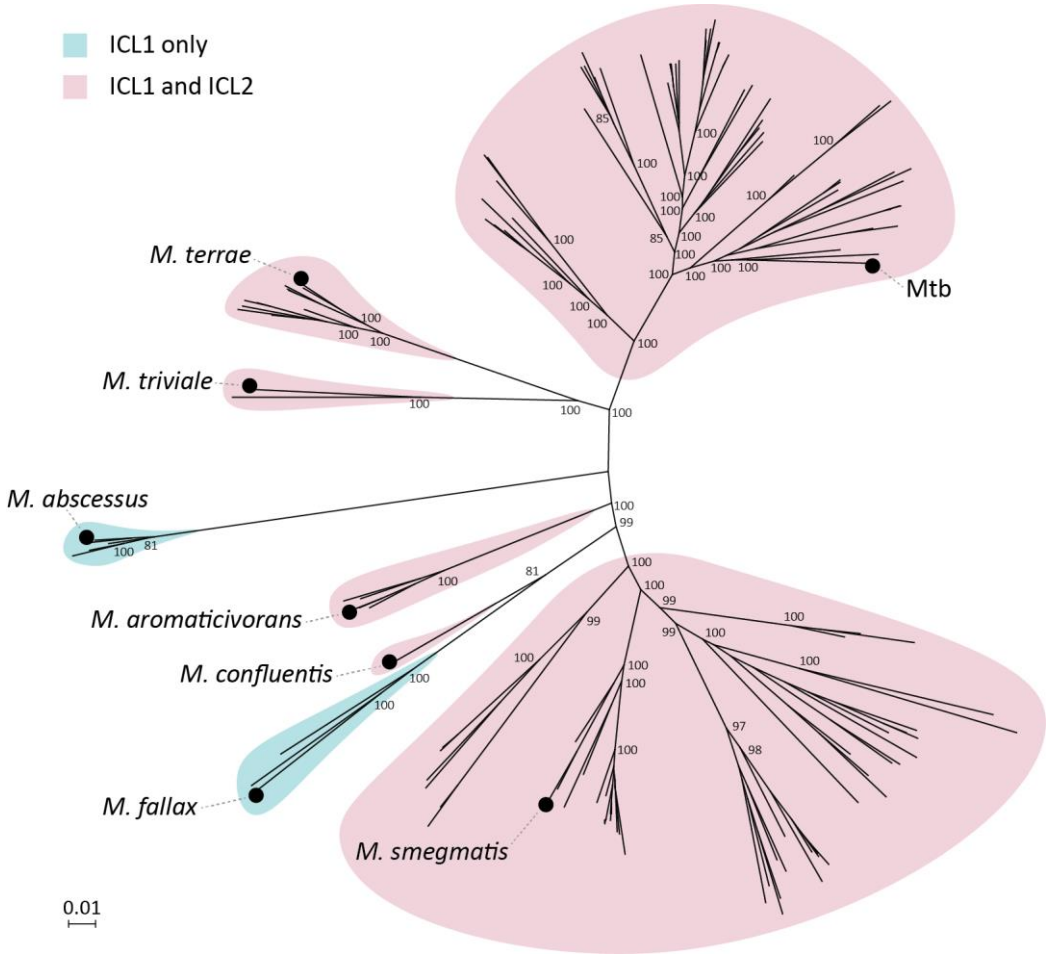


Figure 16: Distribution of ICL1 and ICL2 in the genus *Mycobacterium*.
Species in clades shaded pink have both ICL1 and ICL2; species in clades shaded blue have ICL1 only. For orientation purposes, key species are marked with black circles. The mycobacterial phylogenetic tree used in this figure was constructed by Acely Garza-Garcia.

2.3. Distribution of IDH1 and IDH2 in the genus *Mycobacterium*

Mycobacterial IDHs were identified by BLAST searching, as described in **Chapter 7**. The distribution of IDH1 and IDH2 in the genus *Mycobacterium* is illustrated in **Figure 17**. All 129 mycobacterial species analysed have at least one IDH. IDH2 is present in nearly all mycobacterial species. We identified four species (marked with pink circles in **Figure 17**) that have lost IDH2; these species all possess IDH1.

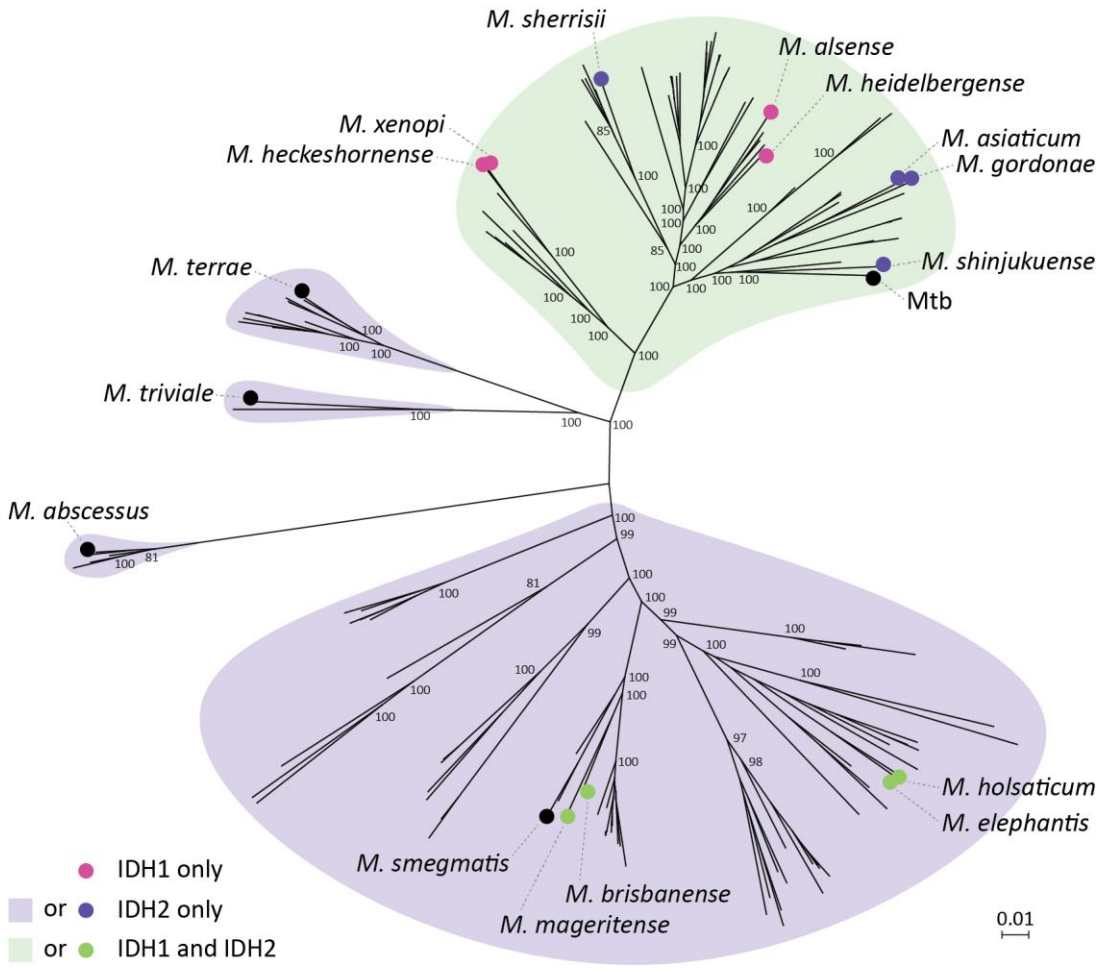


Figure 17: Distribution of IDH1 and IDH2 in the genus *Mycobacterium*.
Species in clades shaded purple typically have IDH2 only. The four species marked with green circles (*M. holsaticum*, *M. elephantis*, *M. brisbanense*, and *M. mageritense*) are the exceptions; they possess both IDH1 and IDH2. Species in the *Tuberculosis-Simiae* clade, shaded green, typically have both IDH1 and IDH2. The exceptions are those species marked with purple and pink circles. The four species marked with purple circles (*M. shinjukuense*, *M. asiaticum*, *M. gordonae*, and *M. sherrisii*) have only IDH2. The four species marked with pink circles (*M. alsense*, *M. heidelbergense*, *M. xenopi*, and *M. heckeshornense*) have only IDH1. For orientation purposes, a key species from each clade is marked with a black circle. The mycobacterial phylogenetic tree used in this figure was constructed by Acely Garza-Garcia.

Overall, IDH1 is phylogenetically distributed, being present in the *Tuberculosis-Simiae* clade of slow growers (shaded green in **Figure 17**), but absent in the other clades (shaded purple in **Figure 17**). Of note, we identified four fast growing species that possess IDH1 (marked with green circles in **Figure 17**), presumably having acquired it through horizontal gene transfer, and four slow growing species that have lost this enzyme (marked with purple circles in **Figure 17**).

2.4. Isocitrate metabolising enzymes are phylogenetically distributed in the genus *Mycobacterium*

Combining the ICL and IDH distributions obtained above, we find that each mycobacterial species has between two and four enzymes for the metabolism of isocitrate, including at least one ICL and at least one IDH (**Figure 18** and **Table 6**). There are three common isocitrate branch point structures: four enzymes (ICL1, ICL2, IDH1 and IDH2); three enzymes (ICL1, ICL2 and IDH2); and two enzymes (ICL1 and IDH2). A fourth combination (ICL1, ICL2 and IDH1) was observed in four species. The species possessing each branch point structure are summarised in **Table 6**. Mapping the isocitrate branch point structures onto the mycobacterial phylogenetic tree (**Figure 18**) shows that variation in branch point composition is phylogenetically associated. The presence or absence of individual isocitrate metabolising enzymes does not appear to be associated with mycobacterial ecology or pathogenicity.

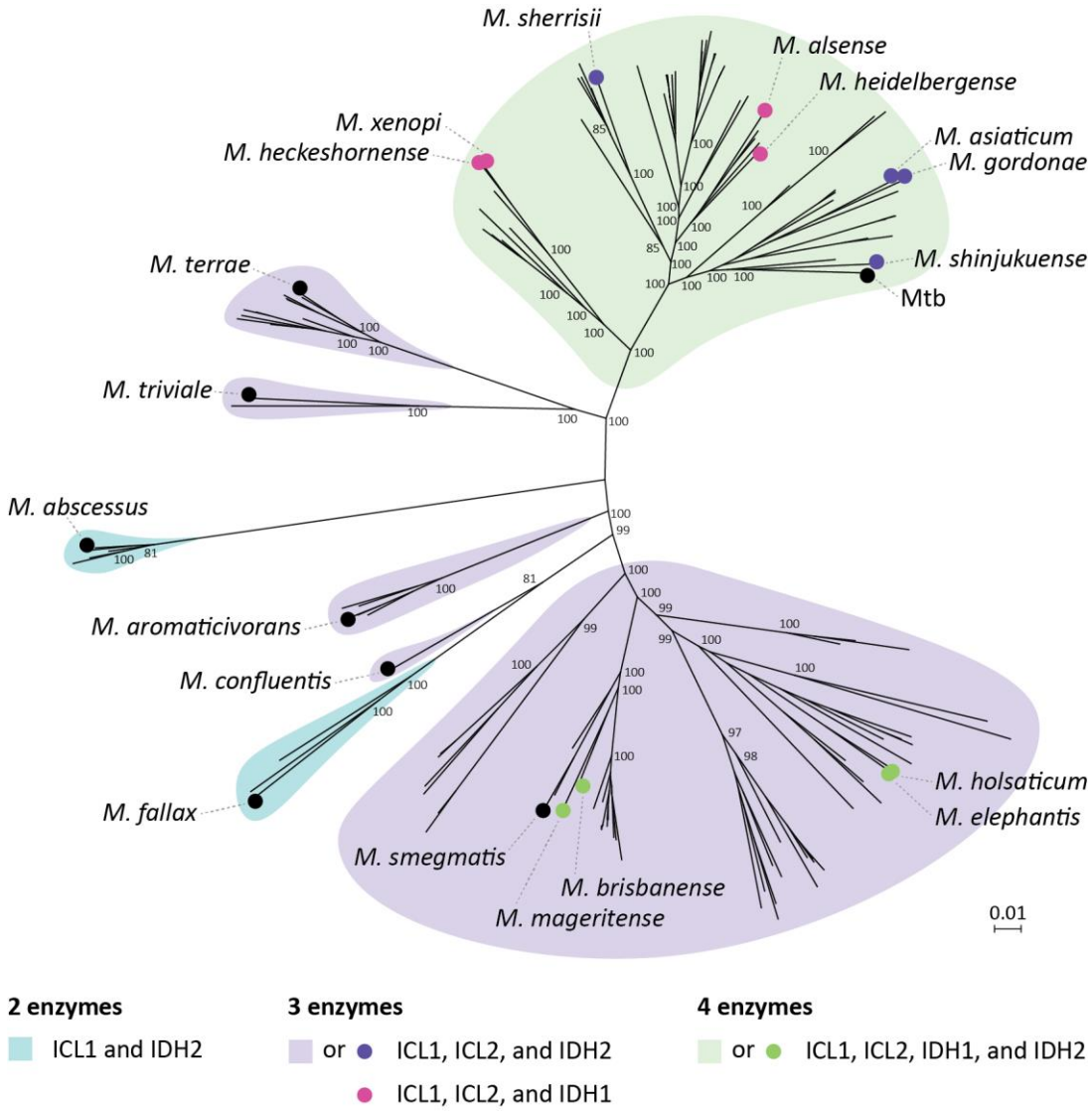


Figure 18: Distribution of the four isocitrate metabolising enzymes in the genus *Mycobacterium*. Species in clades shaded blue have ICL1 and IDH2 only. Species in clades shaded purple typically have ICL1, ICL2, and IDH2. The four species marked with green circles (*M. holsaticum*, *M. elephantis*, *M. brisbanense*, and *M. mageritense*) are the exceptions; they also possess IDH1. Species in the *Tuberculosis-Simiae* clade, shaded green, typically have all four enzymes. The exceptions are those species marked with purple and pink circles, which have three enzymes. The four species marked with purple circles (*M. shinjukuense*, *M. asiaticum*, *M. gordonae*, and *M. sherrisii*) have ICL1, ICL2, and IDH2. The four species marked with pink circles (*M. alsense*, *M. heidelbergense*, *M. xenopi*, and *M. heckeshornense*) have ICL1, ICL2, and IDH1. For orientation purposes, a key species from each clade is marked with a black circle. The mycobacterial phylogenetic tree used in this figure was constructed by Acely Garza-Garcia.

Table 6: Mycobacterial isocitrate branch point structures.
2 enzymes: Species with a two-enzyme isocitrate branch point have ICL1 and IDH2. They are all fast growers and form two phylogenetic clusters, the *Abscessus-Chelonae* clade and a group of species closely related to *M. fallax*. **3 enzymes:** Species with a three-enzyme isocitrate branch point have ICL1, ICL2, and either IDH2 (usual) or IDH1 (rarely). **4 enzymes:** Most species with a four-enzyme isocitrate branch point are in the *Tuberculosis-Simiae* clade of slow growers. A few fast-growing species, for which a three-enzyme branch point with IDH2 would be predicted based on their phylogenetic position, also have a four-enzyme branch point.

Enzymes		**Phylogenetic growth rate**	**Species**	**Presentation in Figure 18**
2	ICL1, IDH2	Fast	*Abscessus-Chelonae* clade *M. brumae*, *M. fallax*, *M. insubricum*	Shaded blue
3	ICL1, ICL2, IDH2	Fast	Most *Fortuitum-Vaccae* clade	Shaded purple
		Slow	*Triviale* clade *Terrae* clade	Shaded purple
			M. asiaticum, *M. gordonae*, *M. sherrisii*, *M. shinjukuense*	Purple circles
3	ICL1, ICL2, IDH1	Slow	*M. alsense*, *M. heckeshornense*, *M. heidelbergense*, *M. xenopi*	Pink circles
4	ICL1, ICL2, IDH1, IDH2	Slow	Most *Tuberculosis-Simiae* clade	Shaded green
		Fast	*M. brisbanense*, *M. elephantis*, *M. holsaticum*, *M. mageritense*	Green circles

2.5. Chapter conclusion

In this chapter, we analysed the distribution of ICL1, ICL2, IDH1, and IDH2 in the genus *Mycobacterium*. Mycobacterial species have between two and four enzymes for the metabolism of isocitrate, and variation in isocitrate branch point composition is phylogenetically rather than ecologically associated. We identified four mycobacterial isocitrate branch point structures, each of which includes ICL1 and at least one IDH. These branch point structures will guide the selection of representative mycobacterial species for kinetic characterisation (**Chapter 5**). Before proceeding to genus-level enzymology, however, we will study the isocitrate metabolising enzymes in Mtb. We will begin with a feature of some Mtb strains: a frameshift mutation in *aceA* which may result in a partial or absent ICL2.

Chapter 3. Exploring the origin of the frameshift mutation that creates *rv1915—1916* from *aceA*

3.1. Chapter introduction

3.1.1. The role of ICL2 *in bacterio* is unclear

While ICL1 has been convincingly established as conditionally essential for Mtb growth on substrates that feed into the glyoxylate shunt and during infection (**Chapter 1, Section 1.5**), the role of ICL2 is much less clear (**Chapter 1, Section 1.6.3**). One reason for this lack of clarity is that in the laboratory reference strain Mtb H37Rv, a frameshift mutation splits *aceA* into two open reading frames, *rv1915* and *rv1916*.[276,277,302,341] The transcriptional and translational consequences of this mutation, which is also present in the avirulent laboratory strain H37Ra, are incompletely understood. Historically, it has been assumed that the frameshift mutation renders *rv1915—1916* a pseudogene, and that H37Rv and H37Ra do not express a functioning ICL2.[276,302,303,342] More recently, recombinant Rv1915 and Rv1916 have both been found to exhibit ICL activity, and the formation of an Rv1915—Rv1916 protein-protein complex has been seen *in vitro*.[303,304,343] However, it is not known whether this complex can assemble *in bacterio*, where Rv1916 expression has yet to be observed.[304,344,345] The status of ICL2 in H37Rv and H37Ra therefore remains unclear.

In Mtb CDC 1551 and Erdman, *aceA* consists of a single reading frame.[276,302] These strains can express a functional ICL2, which is homologous to the ICL2 present in the majority of other mycobacterial species (**Chapter 2**). Although allosteric mediation of ICL2 activity by Ac-CoA was recently proposed to regulate flux through the glyoxylate shunt in Mtb, deletion of single frame *aceA* does not affect Mtb growth on glyoxylate shunt substrates or persistence in mice unless *icl* is also deleted.[279,305] These

observations highlight that the role of ICL2 in Mtb isocitrate metabolism has yet to be fully elucidated.

Given the ongoing uncertainty surrounding the function and importance of ICL2, it would be helpful to clarify whether the frameshift mutation that splits *aceA* into *rv1915—1916* is present in Mtb strains other than H37Rv and H37Ra. It has been suggested that the mutation is a laboratory adaptation of H37Rv and H37Ra, as these strains have been cultured for extended periods without lipids as their major carbon source.[341] If this hypothesis is correct, and frameshifted *aceA* is restricted to H37Rv and H37Ra, these strains would not exhibit representative Mtb isocitrate metabolism, and so would not be appropriate choices for comparison to other mycobacterial species, which have a single frame *aceA*. Understanding how frameshifted and single frame *aceA* are distributed among Mtb strains will therefore aid in the development and interpretation of our proposed comparative enzymology study.

3.1.2. The evolutionary relationships between Mtb strains: lineages of the *Mycobacterium tuberculosis* complex (MTBC) and their assignment

3.1.2.1. Structure and evolution of the MTBC

As discussed in **Chapter 1**, humans and animals are frequently exposed to mycobacteria through contact with soil and water. Such exposure typically results in neither disease nor the establishment of long-term host-mycobacteria relationships.[133,159] Between 35,000 and 70,000 years ago, however, a mycobacterial species living in the Horn of Africa became adapted to the human host.[108,109,479] Having made the ecological transition from an alternative niche to the human body, this mycobacterial species accompanied human populations migrating out of Africa.[108,480] Its continued evolution produced a set of closely related species and subspecies known as the *Mycobacterium*

tuberculosis complex (MTBC). The members of the MTBC are host-adapted and cause TB in either humans (Mtb *sensu stricto* and *M. africanum*) or animals (*M. bovis*, *M. caprae*, *M. microti*, *M. mungi*, *M. orygis*, *M. pinnipedi*, and *M. suricattae*).[10,235,481] The human-adapted members of the MTBC can be divided into seven phylogeographic lineages, as shown in **Figure 19**.[10,482,483] Mtb *sensu stricto* comprises five lineages (1, 2, 3, 4, and 7), while *M. africanum* consists of two lineages (5 and 6).[10,481–483] These lineages vary in their virulence, prevalence, transmissibility, and geographic distribution.[481,484,485]

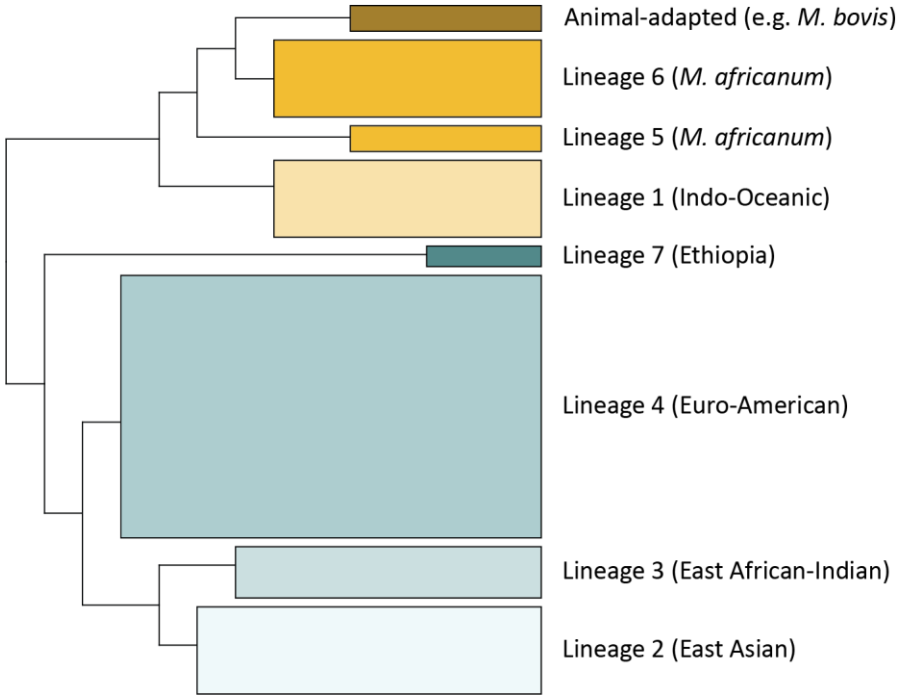


Figure 19: Phylogenetic tree of the *Mycobacterium tuberculosis* complex.
The human-adapted members of the MTBC can be divided into seven lineages, five lineages of Mtb *sensu stricto* and two lineages of *M. africanum*.[482,483] The phylogenetic position of the animal-adapted MTBC members indicates that they derived from a human-adapted ancestor.[486] Lineage names and numbers are those described by Gagneux *et al*.[482] A phylogenetic tree of the MTBC constructed by Acely Garza-Garcia was used in the creation of this figure. The tree is presented in a non-scaled format. Box sizes are proportional to the number of sequences from each lineage that were used to construct the tree. Colour is used for aesthetic purposes and to facilitate visual separation of the lineages.

3.1.2.2. Assignment of strains to MTBC lineages

By considering Mtb isolates in the context of their position in the MTBC, one can study the molecular epidemiology and evolution of TB.[487] Accordingly, various methods for MTBC lineage assignment have been developed.[488,489] One such method is spacer oligonucleotide typing (spoligotyping), which examines the direct repeat (DR) region of the Mtb genome to determine whether each of 43 spacer sequences is present or absent. The resulting presence/absence pattern is called a spoligotype.[490,491] While spoligotyping is traditionally a PCR-based experiment, it can also be performed *in silico* using whole genome sequencing data.[487,490,492,493] The spoligotypes of over 111,000 clinical isolates have been used to define spoligotype families (groups of related spoligotypes) and to construct international spoligotyping databases, allowing novel isolates to be classified.[487,491,494–497]

3.1.3. Towards the evolutionary origin of the frameshift mutation that creates *rv1915—1916* from *aceA*

In this chapter, we first identify the location of the frameshift mutation that creates *rv1915—1916* from *aceA* and the predicted consequences of this mutation at the protein level. We next demonstrate that frameshifted and single frame *aceA* can be easily distinguished using BLAST searching, and establish that the frameshift mutation is not a consequence of the laboratory adaptation of H37Rv and H37Ra. Having identified that frameshifted *aceA* is widespread among Mtb strains, we explore the evolutionary origins of the mutation by performing BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment for 3825 Mtb assemblies from the NCBI Microbial Genomes RefSeq database.

Further details of the work described in this chapter are provided in **Chapter 7**. Throughout this chapter, 'frameshifted *aceA*' refers to an *aceA* containing the two open reading frames *rv1915* and *rv1916*, as seen in H37Rv. 'Single frame *aceA*' refers to an *aceA* that can be translated in a single open reading frame, like *aceA* from CDC 1551 and non-tuberculous mycobacterial species.

3.2. Location and predicted consequences of the mutation

Figure 20A shows the position of the frameshift mutation p.(Tyr296ThrfsTer73), which results in the splitting of H37Rv *aceA* into *rv1915* and *rv1916*. The genomic locus of the mutation is given in **Table 7**. Deletion of a single thymine nucleotide alters the reading frame relative to CDC 1551 *aceA* at codon 296, changing tyrosine to threonine (**Figure 20B**). A further 71 residues (coloured brown in **Figure 20A**) can be translated in the new reading frame before encountering the stop codon (TGA) which is annotated as the end of *rv1915*. The start codon of *rv1916* overlaps the stop codon of *rv1915* by one base (**Figure 20C**). Translation of *rv1916* would therefore result in a correct partial ICL2 sequence, since the reading frame is the same as for CDC 1551 *aceA*. This prediction is consistent with the observation that recombinant Rv1916 displays ICL activity.[304]

Table 7: Mutations in H37Rv *aceA* identified by comparison to the CDC 1551 reference sequence.
The *aceA* sequences of CDC 1551 and H37Rv differ at two positions. Firstly, a single nucleotide polymorphism (SNP) is present which would result in the protein variant Asp179Gly. Secondly, deletion of a thymine residue results in the frameshift mutation p.(Tyr296ThrfsTer73) illustrated in **Figure 20**. Parentheses indicate that the protein products are hypothetical rather than experimentally observed. **NCBI Reference Sequences:** H37Rv, NC_000962.3 (*rv1915—1916* NC_000962.3:2160463-2162759); CDC 1551, NC_002755.2 (*aceA* NC_002755.2:2157754-2160051).

Mutation type	Change in DNA sequence	Predicted change in protein sequence
SNP	NC_002755.2:g2158289 A>G	WP_003409584.1:p.(Asp179Gly)
Frameshift	NC_002755.2:g.2158639delT	WP_003409584.1:p.(Tyr296ThrfsTer73)

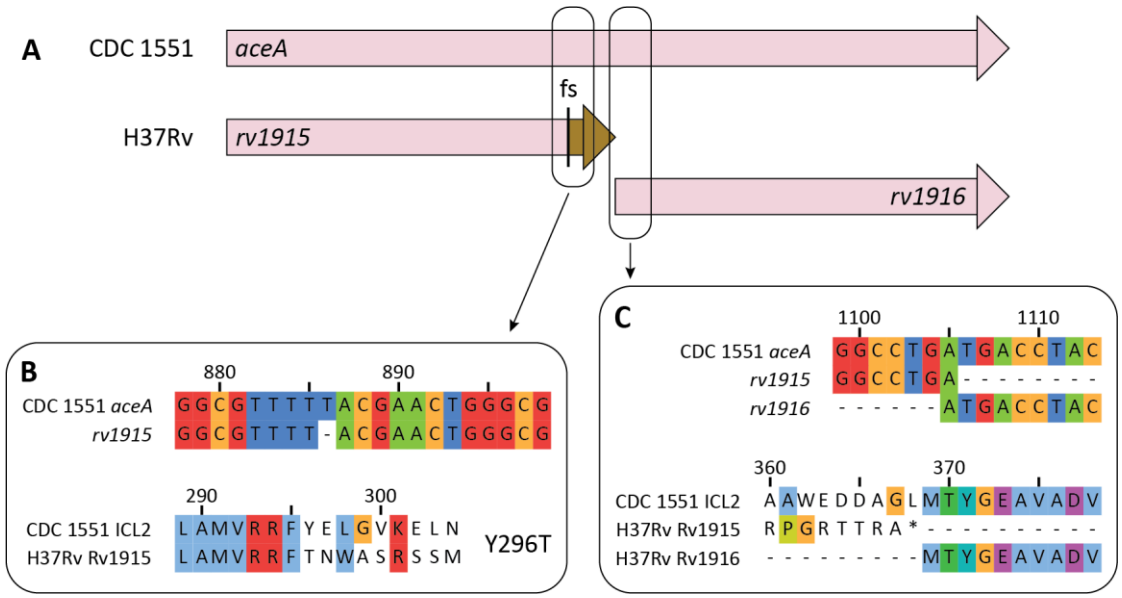


Figure 20: Alignment of the H37Rv and CDC 1551 *aceA* sequences to illustrate the position of the frameshift mutation p.(Tyr296ThrfsTer73) in H37Rv *aceA*.
A: CDC 1551 *aceA* consists of a single open reading frame, whereas H37Rv *aceA* is split into two parts, *rv1915* and *rv1916*. The split occurs due to a frameshift mutation in *rv1915* (labelled fs), which results in 72 altered codons (coloured brown) followed by a premature stop codon, which terminates the *rv1915* reading frame. **B: Comparison of the CDC 1551 and H37Rv DNA and protein sequences at the site of the frameshift mutation.** The reading frame in *rv1915* is shifted by deletion of a thymine nucleotide. The first altered amino acid is Tyr296, which becomes Thr. **C: Comparison of the CDC 1551 and H37Rv DNA and protein sequences at the end of the *rv1915* reading frame.** The start codon of *rv1916* overlaps the stop codon of *rv1915* by one nucleotide. Consequently, *rv1916* is read in the same frame as CDC 1551 *aceA* and, if translated, would result in a partial ICL2 with the correct sequence. **Genome sequences:** CDC 1551 *aceA*, NC_002755.2:2157754-2160051; H37Rv *aceA* (*rv1915—1916*), NC_000962.3:2160463-2162759.

Figure 21 shows how the areas of the sequence affected by the frameshift mutation map onto the protein structure of ICL2 (for an introduction to ICL structures, see **Chapter 1, Section 1.6**). The frameshift occurs in the part of the *aceA* sequence encoding Domain II, the series of helices and loops that join the two halves of the ICL α/β barrel. Thus, *rv1915* encodes the first five β-sheets of the α/β barrel (Domain I, coloured dark teal in **Figure 21**), followed by the start of Domain II (coloured yellow in **Figure 21**). The final 72 amino acids of Rv1915 – those altered by the frameshift mutation – also fall within Domain II, and are coloured brown in **Figure 21**. *Rv1916*, meanwhile, encodes the end of Domain II (coloured yellow in **Figure 21**), the final three β-sheets of the ICL α/β

barrel (Domain III, coloured light teal in **Figure 21**) and the *C*-terminal domain (Domain IV, coloured pink in **Figure 21**).

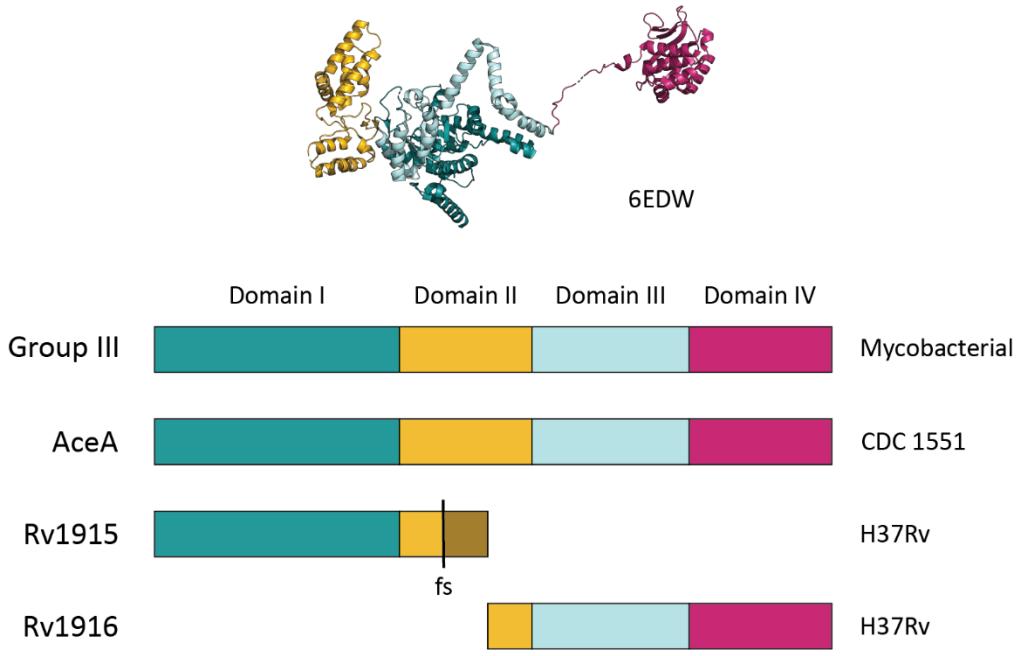


Figure 21: Mapping Rv1915 and Rv1916 onto the structure of ICL2.
ICL2s are classified as Group III ICLs, contain four domains, and are almost exclusively found in mycobacteria (**Chapter 1, Section 1.6**). *AceA* from Mtb CDC 1551 is translated in a single frame to produce a typical Group III ICL. In Mtb H37Rv, the frameshift mutation p.(Tyr296ThrfsTer73) splits *aceA* into two open reading frames, *rv1915* and *rv1916*. Rv1915 consists of Domain I (dark teal) and the first part of Domain II (yellow). The frameshift mutation (fs) occurs in Domain II, and 72 amino acids (brown) can be translated in the new reading frame before a stop codon is encountered. Rv1916 comprises the end of Domain II (yellow), Domain III (light teal), and Domain IV (pink). The typical ICL catalytic residues are located in Domain I; alternative catalytic residues have been suggested for Rv1916.[304]

3.3. The mutation is not a consequence of laboratory adaptation

BLAST searches were performed to identify whether the frameshift mutation was a consequence of the laboratory adaptation of H37Rv and H37Ra. The DNA sequences for H37Rv *aceA* (*rv1915—1916*, frameshifted) and CDC 1551 *aceA* (single frame) were translated in a single frame and the resulting protein sequences were used to query the NCBI Mtb RefSeq Genomes database. These queries returned a list of 1000 Mtb strains with frameshifted (H37Rv-like) *aceA* and 1000 Mtb strains with single frame (CDC 1551-like) *aceA*. The finding that frameshifted *aceA* is not restricted to H37Rv and

H37Ra indicates that the mutation did not occur during laboratory adaptation of these strains, but earlier in Mtb evolution. As H37Rv (frameshifted *aceA*) and CDC 1551 (single frame *aceA*) both belong to the Euro-American lineage (Lineage 4), the mutation may have occurred during the evolution of this lineage. To explore the evolutionary origin of frameshifted *aceA*, we evaluated the distribution of the two *aceA* genotypes (frameshifted and single frame) in the MTBC.

3.4. How are *aceA* genotypes distributed in the MTBC?

3.4.1. Methodology: *aceA* genotyping and *in silico* spoligotyping of Mtb strains from the NCBI Microbial Genomes RefSeq database

To establish the distribution of the two *aceA* genotypes in the MTBC, we subjected 3825 Mtb assemblies from the NCBI Microbial Genomes RefSeq database to BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based lineage assignment. *AceA* genotypes were determined by BLAST because our initial BLAST searches (**Section 3.3**) identified that single frame *aceA* and frameshifted *aceA* can easily be distinguished using this method. *In silico* spoligotyping (specifically, the SpoTyping method developed by Xia *et al*.) followed by spoligotype-based predictive lineage assignment was selected for the assignment of strains to MTBC lineages.[492,498] This method is fast and computationally accessible, allowing us to identify features of interest using a large number of strains.[492]

3.4.2. *AceA* genotyping

The *aceA* genotype of each strain (frameshifted or single frame) was determined by BLAST. Detailed parameters are given in **Chapter 7**. Briefly, the single frame translation of H37Rv *rv1915—1916* prepared in **Section 3.3** was used to determine whether frameshifted *aceA* was present. If a frameshifted *aceA* was not identified, CDC 1551

aceA was used to check for a single frame *aceA*. If neither query sequence returned a positive result, the genotype was left unassigned. Initially, 3636 strains (95% of the dataset) were successfully genotyped (**Table 8**). Failure to genotype *aceA* typically resulted from the presence of SNPs. Relaxation of the BLAST parameters to permit the presence of SNPs in *aceA* enabled the classification of a further 160 strains (**Table 8**). Ultimately, *aceA* genotypes were obtained for 99% of the analysed sequences, with 44% of strains possessing frameshifted *aceA*.

Table 8: Number of strains with each *aceA* genotype.
3825 Mtb assemblies from the NCBI Microbial Genomes RefSeq database were identified as suitable for analysis (as detailed in **Chapter 7**). *AceA* genotyping of these 3825 strains was performed as described above and in **Chapter 7**. Two different sets of BLAST output parameters were used. Initially, genotyping was performed using SNP-intolerant (strict) output parameters. To reduce the number of strains categorised as 'undefined', genotyping was repeated using SNP-tolerant (relaxed) output parameters.

***aceA* genotype**	**SNP-intolerant (strict) output parameters**		**SNP-tolerant (relaxed) output parameters**	
	Number of strains	**Percentage of dataset**	**Number of strains**	**Percentage of dataset**
Frameshifted	1608	42%	1669	44%
Single frame	2028	53%	2127	56%
Undefined	189	5%	29	1%
Total	3825	100%	3825	100%

3.4.3. *In silico* spoligotyping and major lineage assignment

In silico spoligotyping of each strain was performed using modified functions from SpoTyping, as described in **Chapter 7**.[492] Briefly, this method uses BLAST searches to identify whether each of the 43 spacer sequences used in spoligotyping is present or absent. The spoligotype is then determined from the pattern of spacer sequences. Next,

we used the online tool TB-Lineage to assign each strain to one of the major MTBC lineages. TB-Lineage offers two options for the prediction of major lineages from spoligotypes: (i) Rules, which performs rule-based major lineage predictions; and (ii) Conformal Bayesian Network (CBN), which performs probability-based major lineage predictions.[499,500] The lineages that can be predicted vary slightly between the two methods. Rules distinguishes between the two lineages of *M. africanum* (Lineages 5 and 6), while CBN does not. Unlike CBN, Rules does not include the animal-adapted strain *M. caprae*, instead classifying these strains as *M. bovis*. Finally, CBN always returns a lineage assignment, whereas Rules includes a final "unknown" category for those strains it is unable to assign.

We used both Rules and CBN to classify the 3825 strains in our dataset. The number of strains assigned to each lineage is shown in **Table 9**. Some lineages were more consistently assigned than others, likely due to the differences in the classification processes used by Rules and CBN (in particular, the order of precedence applied to lineages in Rules).[500] The two methods assigned the same 1001 strains to the East Asian lineage, while strains classified as Indo-Oceanic, *M. africanum* or *M. bovis* by CBN nearly always received the same assignment using Rules. In contrast, assignment of strains to the Euro-American and East African-Indian lineages was less consistent. Some strains assigned as Euro-American by CBN were designated Indo-Oceanic (74 strains) or *M. africanum* (11 strains) strains by Rules. Finally, Rules did not identify any strains as belonging to the East African-Indian lineage. The 69 strains assigned to this lineage by CBN were classified as Euro-American (12 strains), Indo-Oceanic (24 strains), *M. bovis* (9 strains), *M. africanum* (6 strains) or unknown (18 strains) using Rules.

Table 9: Assignment of 3825 Mtb strains to major MTBC lineages using Rules and CBN on TB-Lineage. **Rules – Number of strains assigned to each lineage using Rules.** The two *M. africanum* lineages (5 and 6) are counted together for easier comparison with CBN, which does not distinguish between them. Rules does not include the animal-adapted strain *M. caprae*. **CBN – Number of strains assigned to each lineage using CBN.** CBN always returns a lineage assignment and so does not include the 'Unknown' category. **Consensus – Number of strains confidently assigned to each lineage.** For Lineage 1, Lineage 2, Lineage 4, *M. africanum*, and *M. bovis*, the criteria for confident assignment were that Rules and CBN returned the same lineage, and that the CBN probability was >0.995. Strains assigned to Lineage 3 were those designated as unknown by Rules, but East African-Indian by CBN with a probability >0.995. Strains counted as *M. caprae* were those designated *M. bovis* by Rules, but *M. caprae* by CBN with a probability >0.995.

Lineage		**Assignment method**		
		Rules	**CBN**	**Consensus**
1	**Indo-Oceanic**	189	95	74
2	**East Asian**	1001	1001	1001
3	**East African-Indian**	0	69	17
4	**Euro-American**	2448	2521	2424
5 and 6	***M. africanum***	42	25	8
Animal	***M. bovis***	123	104	101
Animal	***M. caprae***	n/a	10	4
Unknown		22	n/a	n/a
Total		3825	3825	3629

Accordingly, to facilitate analysis of the distribution of *aceA* genotypes across the major lineages, we created a list of strains for which the lineage was confidently assigned (**Table 9, Consensus**). In most cases, the criteria for confident assignment were that Rules and CBN returned the same lineage, and that the CBN probability was >0.995. The exceptions were the East African-Indian lineage (strains designated as unknown by Rules, but East African-Indian by CBN with a probability >0.995) and *M. caprae* (strains designated *M. bovis* by Rules, but *M. caprae* by CBN with a probability >0.995). Using these criteria, a major lineage could be identified for 95% of the dataset (3629 strains) (**Table 9**). The majority of confidently assigned strains belong to the Euro-American (66.8%) and East Asian (27.6%) lineages, which are responsible for most cases of human

TB and are consequently highly represented in sequencing databases.[484] Indeed, a comparable bias towards these lineages was observed in a SNP-based lineage assignment study.[501]

3.4.4. Analysis of the distribution of *aceA* genotypes across the major MTBC lineages shows that frameshifted *aceA* is a Euro-American genotype

We next analysed the distribution of *aceA* genotypes across the major MTBC lineages, using the set of 3629 strains confidently assigned to lineages in **Section 3.4.3** (**Table 10** and **Figure 22**). This analysis found frameshifted *aceA* to be a predominantly Euro-American genotype, with 99.1% of strains encoding frameshifted *aceA* being assigned to the Euro-American lineage. The 15 strains with frameshifted *aceA* that were assigned to other lineages (0.4% of the total dataset) may well represent instances of lineage misassignment, as spoligotype-based major lineage assignment is around 95% accurate.[500,502,503] Assignment of these strains should be repeated using an alternative method prior to drawing a final conclusion. Within the Euro-American lineage, frameshifted *aceA* was the dominant genotype, accounting for two-thirds of sequenced strains. These results suggest that the frameshift mutation in *aceA* occurred after the establishment of the Euro-American lineage, as we might expect given that H37Rv (frameshifted *aceA*) and CDC 1551 (single frame *aceA*) both belong to this lineage (**Section 3.3**). Following this line of reasoning, we next examined whether frameshifted *aceA* is associated with a particular subset of Euro-American strains.

Table 10: *AceA* genotypes of the 3629 strains that were confidently assigned to major MTBC lineages. *AceA* genotypes were determined in **Section 3.4.2**; lineage assignments were made in **Section 3.4.3**.

Lineage		*aceA* genotype			Total
		Frameshifted	Single frame	Undefined	
1	**Indo-Oceanic**	2	68	4	74
2	**East Asian**	12	982	7	1001
3	**East African-Indian**	0	17	0	17
4	**Euro-American**	1611	798	15	2424
5 and 6	***M. africanum***	0	8	0	8
Animal	***M. bovis***	1	100	0	101
Animal	***M. caprae***	0	4	0	4
Total		1626	1977	26	**3629**

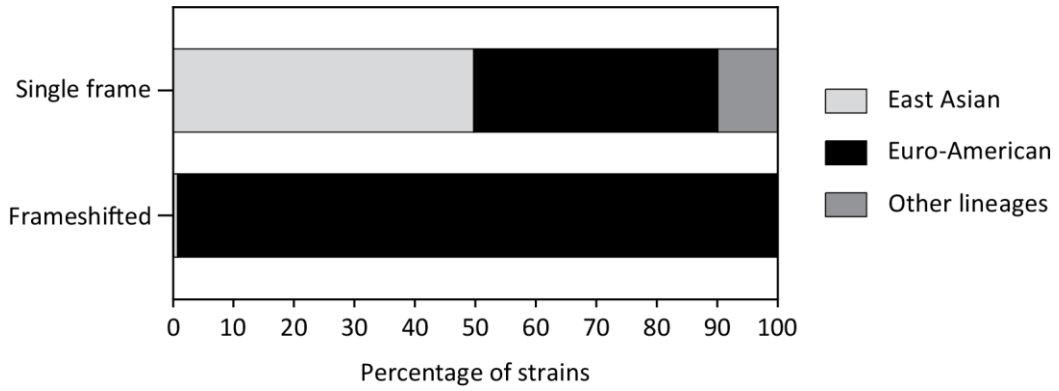


Figure 22: Breakdown of *aceA* genotypes by MTBC lineage.
This plot shows the percentage of strains with each genotype that belong to the different MTBC lineages (data from **Table 10**, collating Indo-Oceanic, East African-Indian, *M. africanum*, *M. bovis*, and *M. caprae* as 'Other lineages' for ease of visualisation). Single frame *aceA* is found in all lineages (**Table 10**), while frameshifted *aceA* is a predominantly Euro-American genotype.

3.4.5. Assignment of Euro-American strains to major spoligotype families

Above, we identified 2409 Euro-American strains with defined *aceA* genotypes (**Section 3.4.4**). Next, we searched for the spoligotypes of these strains in the international spoligotyping database SITVIT2.[496] 1820 strains had spoligotypes listed in SITVIT2. Of

these strains, 1817 were identified as belonging to Euro-American spoligotype families. The other three strains had a spoligotype classified as *M. africanum*. **Table 11** shows the distribution of *aceA* genotypes in the major Euro-American spoligotype families H, LAM, S, T, and X. **Figure 23** shows the percentage of strains in each spoligotype family that have each *aceA* genotype. All strains assigned to the S family and 98% of strains assigned to the LAM family have a frameshifted *aceA*, while 98% of strains with X family spoligotypes have a single frame *aceA*. The H and T spoligotype families contain both *aceA* genotypes.

Table 11: *AceA* genotypes of the 1817 strains that were assigned to Euro-American spoligotype families. *AceA* genotypes were determined in **Section 3.4.2**. The spoligotypes determined in **Section 3.4.3** were used to query SITVIT2; strains with recognised Euro-American spoligotypes were categorised by spoligotype family. The spoligotypes belonging to each family are given in **Chapter 7**.

Spoligotype family	Number of strains		
	Frameshifted *aceA*	Single frame *aceA*	Total
H	43	230	273
LAM	721	13	734
S	25	0	25
T	397	306	703
X	2	80	82
Total	**1188**	**629**	**1817**

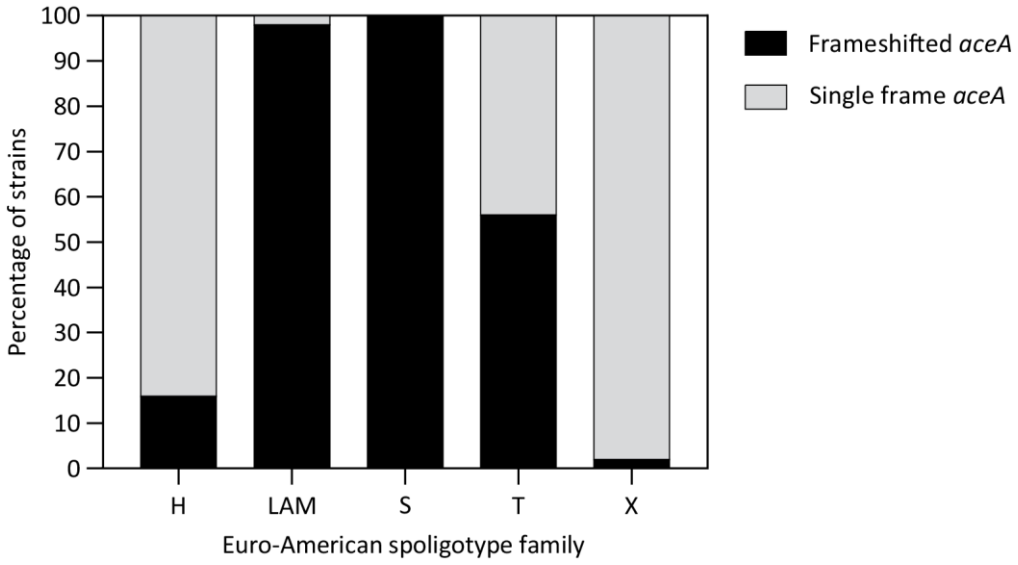


Figure 23: Percentage of strains in each Euro-American spoligotype family with each *aceA* genotype. Visualisation of the data from **Table 11** shows that frameshifted *aceA* is the dominant genotype in the S (100%) and LAM (98%) families, while single frame *aceA* is dominant in the X family (98%). The H and T families contain both *aceA* genotypes.

Figure 24 shows the phylogenetic distribution of the five spoligotype families within Lineage 4. The S and X families form distinct clades (Lineages 4.4.1.1 and 4.1.1 respectively), while the majority of LAM family strains are found in Lineage 4.3. In contrast, members of the H and T families are dispersed across many sublineages, meaning that spoligotypes from these families are less phylogenetically informative.[491,494] Indeed, lack of resolution (discriminatory power) is the main disadvantage of spoligotyping.[504]

Having reached the limit of phylogenetic resolution that can be obtained using a spoligotyping-based approach, we did not consider it appropriate to analyse the distribution of *aceA* genotypes on the level of individual spoligotypes.[503,505–507] Our suggestion that the frameshift mutation in *aceA* occurred during the evolution of Lineage 4 (**Section 3.3**) remains plausible, but cannot be proven with the data we have acquired thus far. To establish the evolutionary origin of frameshifted *aceA*, we would

need to more accurately determine the phylogenetic positions of our genotyped strains by using SNP-based lineage assignment.[485,489,508,509] Due to time constraints, we have yet to pursue this line of enquiry.

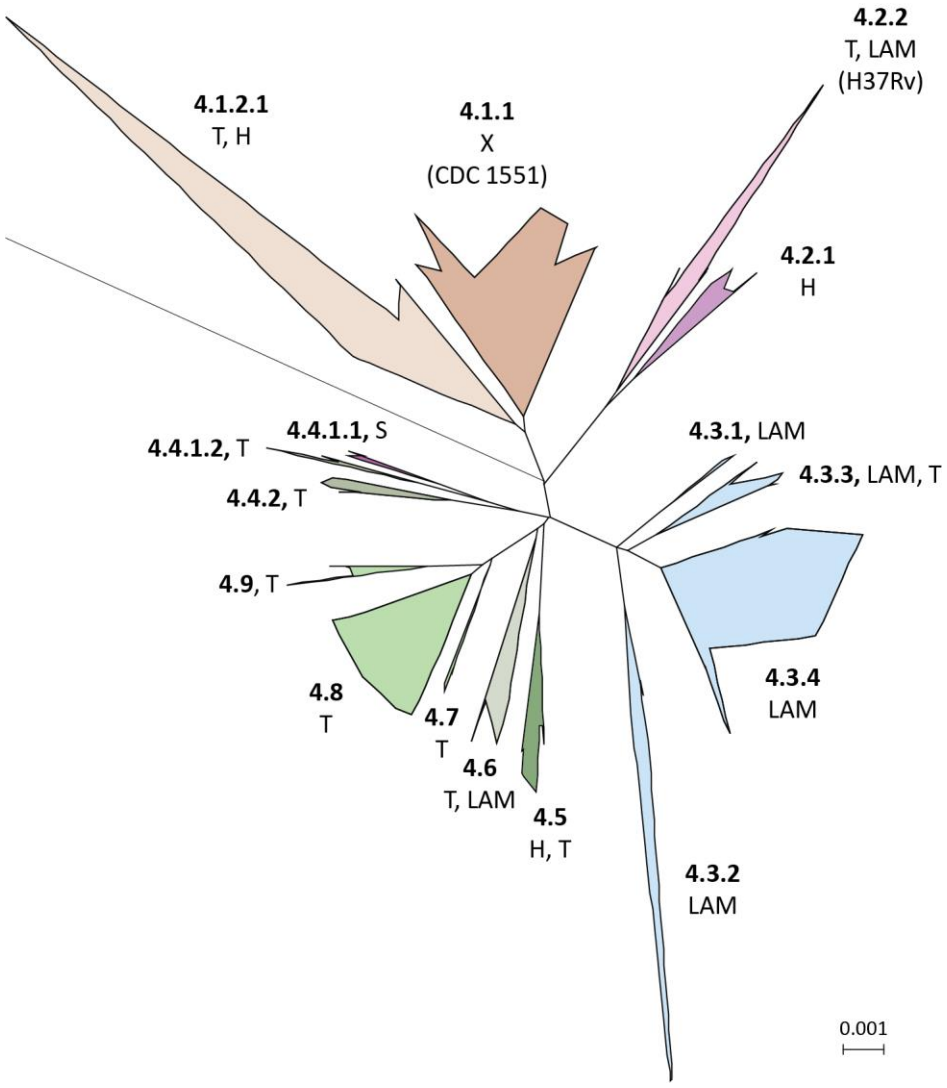


Figure 24: Phylogenetic distribution of the spoligotype families H, LAM, S, T, and X in Lineage 4.
While spoligotypes predict MTBC lineages with around 95% accuracy, they are not always phylogenetically informative on the sublineage level.[502,503] Mapping the Euro-American spoligotype families H, LAM, S, T, and X onto a SNP-based phylogenetic tree of Lineage 4 shows that the S (Lineage 4.4.1.1) and X (Lineage 4.1.1) families form distinct clades. LAM spoligotypes are found primarily in Lineage 4.3, while H and T spoligotypes are found in many sublineages. H37Rv belongs to Lineage 4.2.2, while CDC 1551 belongs to Lineage 4.1.1. This figure was constructed using the SNP-based phylogenetic tree of the MTBC and other data included in PhyTB.[508] The scale indicates the mean number of nucleotide substitutions per site. Annotations and colours were added in Adobe Illustrator to aid visual separation of sublineages.

3.5. Chapter conclusion

In this chapter, we identified the frameshift mutation that creates *rv1915—1916* from *aceA* as p.(Tyr296ThrfsTer73) (**Section 3.2**). BLAST searches established that at least 1000 Mtb strains encode frameshifted *aceA*, indicating that the mutation did not occur during the laboratory adaptation of H37Rv (**Section 3.3**). To explore the evolutionary origins of frameshifted *aceA*, we performed BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment for 3825 Mtb strains from the NCBI Microbial Genomes RefSeq database (**Section 3.4**). *AceA* genotypes were obtained for 99% of the analysed sequences, with 44% of strains possessing frameshifted *aceA* (**Section 3.4.2**). Predictive MTBC lineage assignment using spoligotypes generated *in silico* was possible for 3629 strains (95% of the dataset). As expected, the majority of these strains belonged to the Euro-American (66.8%) and East Asian (27.6%) lineages (**Section 3.4.3**).

Analysis of the distribution of *aceA* genotypes across the MTBC lineages showed that 99.1% of strains encoding frameshifted *aceA* were assigned to the Euro-American lineage, and that two-thirds of Euro-American strains possessed frameshifted *aceA* (**Section 3.4.4**). We were able to assign 1817 strains to the major Euro-American spoligotype families (H, LAM, S, T, and X) by searching for their spoligotypes in SITVIT2 (**Section 3.4.5**). Frameshifted *aceA* is the dominant genotype in the S (100%) and LAM (98%) families, while single frame *aceA* is dominant in the X family (98%). The H and T families contain both *aceA* genotypes.

We can conclude that the frameshift mutation in *aceA* occurred after the establishment of the Euro-American lineage, as we might expect given that this lineage includes both

H37Rv (frameshifted *aceA*) and CDC 1551 (single frame *aceA*). However, the phylogenetic resolution offered by our spoligotype-based method is insufficient to pinpoint the evolutionary origin of frameshifted *aceA*. To continue, we would need to take a higher resolution approach, coupling BLAST-based *aceA* genotyping to SNP-based lineage assignment. While we will not pursue this investigation further during this thesis, understanding the phylogenetic distribution and *in bacterio* consequences of the frameshift mutation that creates *rv1915—1916* from *aceA* will be important in resolving the ongoing uncertainty surrounding the function and importance of Mtb ICL2.

Having situated mycobacterial isocitrate metabolising enzymes in their evolutionary context (**Chapter 2** and this chapter), we move on to considering their kinetics, beginning with the four enzymes from Mtb CDC 1551 (**Chapter 4**).

Chapter 4. Preparation and kinetic characterisation of the isocitrate metabolising enzymes from *Mycobacterium tuberculosis* CDC 1551

4.1. Chapter introduction

Having analysed the distribution of the four isocitrate metabolising enzymes in the genus *Mycobacterium* (**Chapter 2**), we are poised to select a representative subset of the genus in which to characterise the kinetics of isocitrate metabolism. Before embarking on a large-scale comparative enzymology study, however, we must first establish methods for the preparation and kinetic characterisation of mycobacterial isocitrate metabolising enzymes in our laboratory. As reviewed in **Chapter 1**, literature data are available for all four of the Mtb enzymes, but isocitrate metabolism kinetics have been minimally characterised outside of Mtb. The Mtb enzymes thus represent a natural starting point for our own studies.

In order to begin method development, we must select the strain of Mtb that we will later compare to other mycobacterial species. This is important owing to the frameshift mutation in *aceA* that is present in many Mtb strains, including the laboratory reference strain H37Rv. In **Chapter 3**, we showed that the p.(Tyr296ThrfsTer73) mutation is not a laboratory adaptation of H37Rv, but a feature of certain Euro-American Mtb lineages. As previously discussed, the impact of this mutation on ICL2 status and function *in bacterio* remains unclear, as it is not known if the frameshift is overcome during translation or if it renders *aceA* a pseudogene. Given this uncertainty, Mtb strains possessing the mutation (such as H37Rv) would not be the best choice for comparison to other mycobacterial species, which encode a single-frame ICL2.

The frequently used virulent Mtb strain CDC 1551 (CSU93/Oshkosh) encodes a single-frame ICL2 which should be directly comparable to the ICL2s from other mycobacterial species. At the start of this work, a truncated form of the enzyme had been characterised, and partial data were available for the full-length form.[278,317] Kinetic data were also available for CDC 1551 ICL1.[317] CDC 1551 IDH2 and IDH1, meanwhile, are identical to the previously characterised H37Rv enzymes.[363,379] Consequently, CDC 1551 was identified as the most suitable Mtb strain for use in method development and future comparative work. An unexpected benefit of this choice was that the crystal structure and further characterisation data for Mtb CDC 1551 ICL2 were published during the project, supporting ICL2 assay development and data analysis.[305]

In this chapter, we draw on the literature to establish methods for the expression, purification and kinetic characterisation of Mtb CDC 1551 ICL1, ICL2, IDH2 and IDH1 in our laboratory. It is hoped that these methods will prove translatable to the orthologous enzymes in other mycobacterial species, facilitating the proposed comparative enzymology study. As this study will involve a large number of recombinant enzymes, our purification method development places speed and purity over yield, where such a choice needs to be made. Finally, detailed experimental procedures for the work described in this chapter are presented in **Chapter 8**.

4.2. Isocitrate lyase 1

4.2.1. Cloning, expression and purification of His_6-ICL1

Icl was amplified from Mtb CDC 1551 genomic DNA by PCR and assembled into the pNIC-NTH vector, which provides a tobacco etch virus (TEV) cleavable *N*-terminal hexahistidine (His_6) tag. The pNIC-NTH vector was selected because Mtb ICL1 had

previously been expressed bearing an *N*-terminal His_6 tag.[317,318] The resulting plasmid was used to transform *E. coli* BL21(DE3)Star and protein expression was performed using a protocol adapted from the literature.[317] As expected, recombinant His_6-ICL1 expressed well and was soluble (**Figure 25**). However, the protein proved unexpectedly challenging to purify by immobilised metal affinity chromatography (IMAC). While the literature reports that His_6-ICL1 can be purified to homogeneity in one step using IMAC, we found that the enzyme consistently eluted with a set of contaminant proteins.[318]

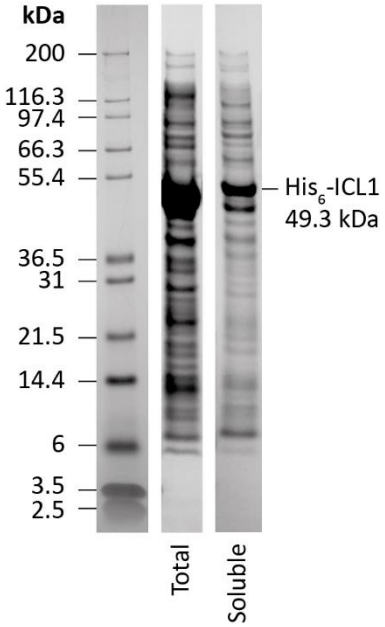


Figure 25: Expression of His_6-ICL1.
Recombinant His_6-ICL1 was expressed in *E. coli* BL21(DE3)Star as described in **Chapter 8**. LDS-PAGE analysis showed that a protein of the expected molecular weight of His_6-ICL1 (49.3 kDa) was present in both the complete cell lysate (**Total**) and in the soluble fraction (**Soluble**). The total and soluble protein lanes were taken from different parts of the same image; the ladder was taken from a separate image.

4.2.1.1. Initial attempts at purification using IMAC

The simple IMAC method outlined in **Figure 26** (**IMAC Method 1**) was used as the starting point for method development. While **IMAC Method 1** was successful in enriching His_6-ICL1, a substantial background of *E. coli* proteins was also present, indicating the need for method optimisation (**Figure 27A**).

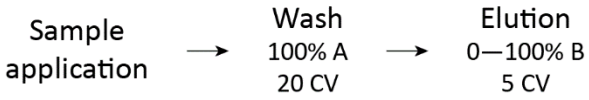


Figure 26: IMAC Method 1.
The sample was applied to a HisTrap FF 1 mL column equilibrated with buffer A. The column was washed with buffer A (20 CV) prior to gradient elution (0—100% B over 5 CV). Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

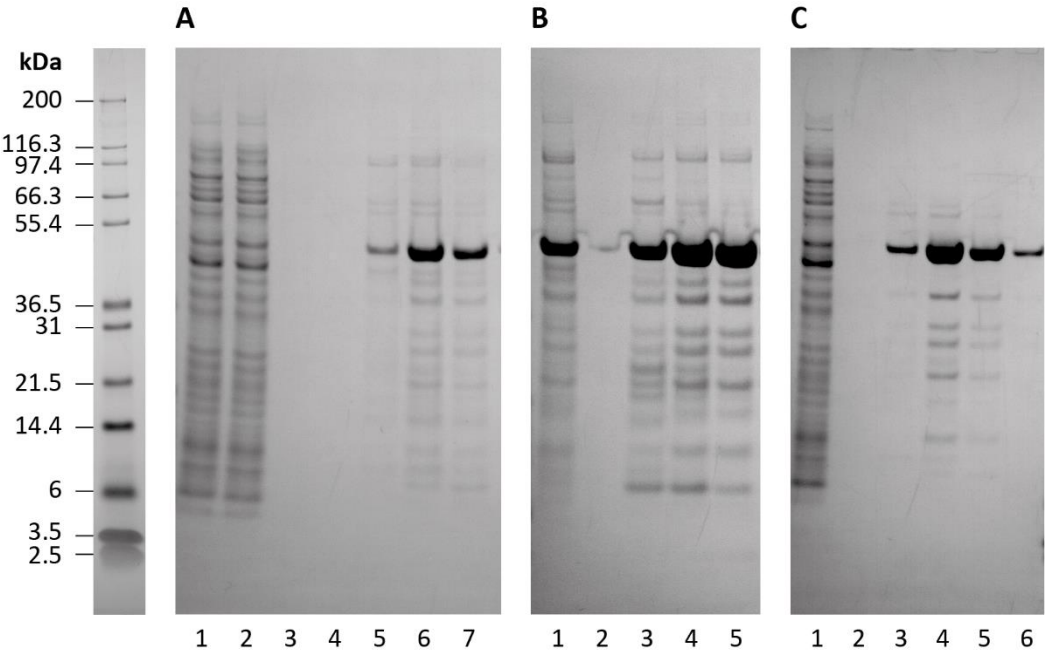


Figure 27: Initial attempts at purification of His_6-ICL1 by IMAC.
A – Purification using IMAC Method 1. His_6-ICL1 was purified using **IMAC Method 1**, as described in **Figure 26**. Lanes 1 and 2, fractions collected during sample application; lanes 3 and 4, fractions collected during the wash step; lanes 5—7, His_6-ICL1 elution peak. **B – Purification using IMAC Method 2.** His_6-ICL1 was purified using **IMAC Method 2**, as described in **Figure 28**. Lane 1, sample application; lane 2, modified wash step (500 mM NaCl, 25 mM imidazole); lanes 3—5, His_6-ICL1 elution peak. **C – Purification using IMAC Method 1, using protease inhibitors during sample preparation.** The inclusion of protease inhibitors in the lysis buffer reduced the number of proteins that co-eluted with His_6-ICL1. Lane 1, sample application; lane 2, wash step; lanes 3—6, His_6-ICL1 elution peak. Panels separated by white space were taken from separate images.

To begin, we altered **IMAC Method 1** to be more similar to the literature method.[318] Specifically, the imidazole and NaCl concentrations used in the column wash step were raised to 25 mM and 500 mM respectively (**Figure 28, IMAC Method 2**). However, these

modifications did not enable purification of His_6-ICL1 to homogeneity (**Figure 27B**) – the enzyme co-eluted with a number of other proteins.

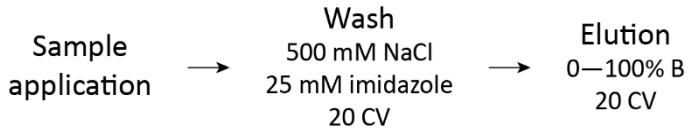


Figure 28: IMAC Method 2.
The sample was applied to a HisTrap FF 1 mL column equilibrated with buffer A. The column was washed with a high salt wash buffer (20 CV) prior to gradient elution (0—100% B over 20 CV). Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; wash buffer: 50 mM Tris, 500 mM NaCl, 25 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

As the majority of the co-eluting proteins were of lower molecular weight than His_6-ICL1 (**Figure 27B**), they could theoretically have been the result of *C*-terminal proteolysis of the enzyme by *E. coli* proteases. We therefore added a protease inhibitor cocktail (c*O*mplete EDTA-free) to the lysis buffer prior to purification using **IMAC Method 1**. The addition of protease inhibitors did slightly improve the purification, with the number of co-eluting proteins being reduced (**Figure 27C**). This observation suggested that a limited amount of *C*-terminal proteolysis was indeed taking place, so as a precaution protease inhibitors were added to the lysis buffer for all subsequent purifications. Multiple co-eluting proteins remained, however, indicating that these were not degradation products of His_6-ICL1. Various strategies to remove these contaminants were therefore explored.

4.2.1.2. Use of size exclusion chromatography as a polishing step

The use of size exclusion chromatography (SEC) as a polishing step after **IMAC Method 1** was unsuccessful, with the only effect being dilution of the sample (**Figure 29**). Rather than continuing to develop a multi-step protocol, we returned to optimisation of the

IMAC method. Our reasoning was that the methods developed for the Mtb enzymes could later be used as the starting point for the purification of a large number of other mycobacterial enzymes, and a less complex protocol would be preferable given the potential scale of such work.

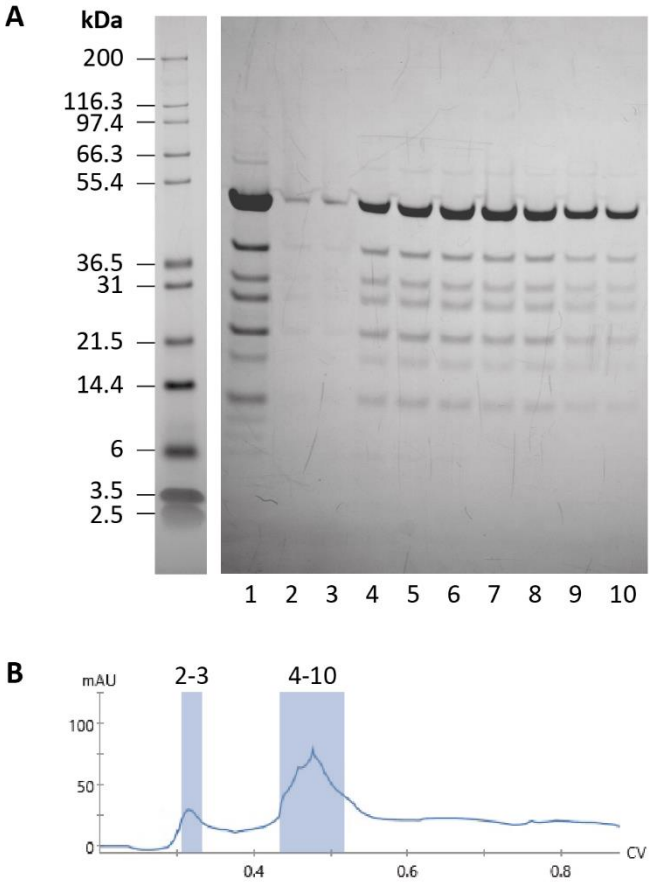


Figure 29: Size exclusion chromatography as a polishing step for His_6-ICL1.
His_6-ICL1 was purified using **IMAC Method 1**, as described in **Figure 26**. Fractions containing the enzyme were pooled (**A -** lane 1) and subjected to SEC, as described in **Chapter 8, Section 8.5.4**. **A – LDS-PAGE analysis of the SEC elution peak.** Fractions from across the SEC elution peak are shown in lanes 2—10, and indicate that SEC was not successful in separating the contaminant proteins from His_6-ICL1. Panels separated by white space were taken from separate images. **B – A_{280} trace of the SEC elution peak.** Blue boxes highlight the fractions used for LDS-PAGE analysis; numbers correspond to the lanes in gel **A**. CV, column volume; mAU, milli-absorbance units. An identical experiment (not shown) was later run using His_6-ICL1 from *M. smegmatis* and produced the same result (but a higher quality A_{280} trace).

4.2.1.3. Use of buffer additives to disrupt protein-protein interactions

Buffer additives such as detergents and glycerol can be used to disrupt non-specific interactions between proteins, facilitating their separation.[510] The Cytiva affinity chromatography troubleshooting guidelines suggest the use of polyoxyethylene (20) sorbitan monolaurate (Tween 20), a non-ionic detergent which has a low absorbance at 280 nm and consequently does not interfere with UV peak detection during FPLC.[511] We therefore considered Tween 20 to be a suitable trial buffer additive. Tween 20 is compatible with the Ni Sepharose 6 Fast Flow resin used in Cytiva HisTrap FF columns at concentrations up to 2% *v*/*v*. The highest compatible concentration, 2% *v*/*v*, was selected on the basis that this would give the clearest indication of the impact of Tween 20 on the His_6-ICL1 purification. If beneficial, other additives could then be tested or the Tween 20 concentration optimised downwards. 2% *v*/*v* Tween 20 was therefore added to both the binding and elution buffers and purification of His_6-ICL1 attempted using **IMAC Method 1**. The use of Tween 20 as a buffer additive did not improve the His_6-ICL1 purification, and the pattern of co-eluting proteins was not altered (**Figure 30A**). This observation suggested that either co-elution of these proteins with His_6-ICL1 was not due to non-specific protein-protein interactions, or that these interactions were not disrupted by the inclusion of Tween 20 in the purification buffers.

Another type of protein-protein interaction that can lead to co-elution is disulfide bonding.[510] To explore disulfide bonding as a possible cause of the observed co-elution, eluate samples containing His_6-ICL1 from previous purification tests were denatured under reducing and non-reducing conditions and analysed by LDS-PAGE (**Figure 30B**). The standard LDS-PAGE sample preparation used in our laboratory includes

denaturation under reducing conditions (50 mM TCEP). Non-reducing conditions were achieved by the omission of TCEP during sample preparation, and resulted in a single band on LDS-PAGE (**Figure 30B**). The finding that the co-eluting proteins were not visible by LDS-PAGE under non-reducing conditions suggested that these proteins may be associated with His_6-ICL1 through disulfide bonding. Were this the case, disruption of these interactions using a reducing agent might facilitate the purification of His_6-ICL1 to homogeneity.

To test this hypothesis, purification of His_6-ICL1 was attempted using **IMAC Method 1** under reducing conditions, including TCEP (0.5 mM) in (a) the lysis buffer; (b) the lysis, binding and elution buffers. Although expensive, TCEP was selected as the reducing agent for these exploratory experiments because it is less toxic than β-mercaptoethanol and more stable than dithiothreitol (DTT) at the run pH (8). Disappointingly, the use of TCEP did not result in separation of His_6-ICL1 from the contaminant proteins by IMAC (**Figure 30C** and **Figure 30D**), suggesting that disulfide bonding was not in fact the cause of the association.

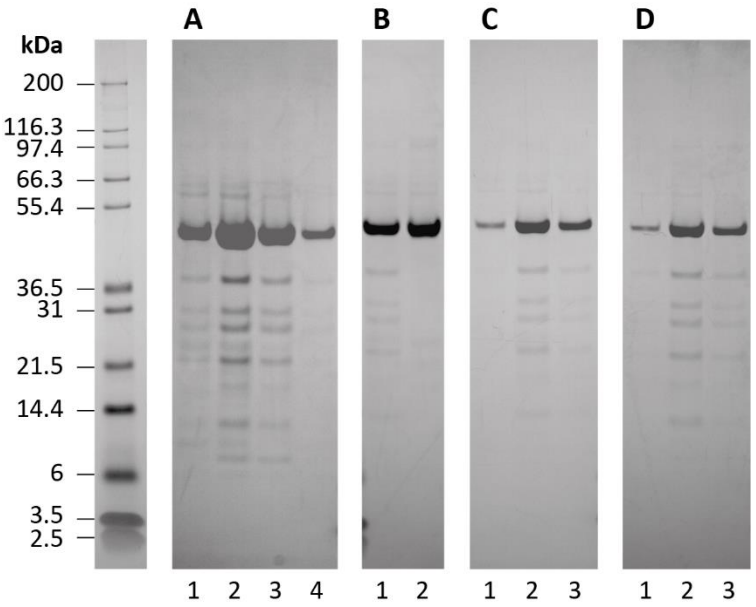


Figure 30: The use of Tween 20 and TCEP as buffer additives.
A – Use of Tween 20 in the binding and elution buffers. His_6-ICL1 was purified using **IMAC Method 1**, as described in **Figure 26** except that 2% *v*/*v* Tween 20 was added to the binding and elution buffers. Lanes 1—4 show the His_6-ICL1 elution peak. **B – LDS-PAGE analysis of His_6-ICL1 under reducing and non-reducing conditions.** Eluate samples containing His_6-ICL1 from previous purification tests were analysed under reducing (50 mM TCEP) and non-reducing (no TCEP) conditions. Multiple bands were visible under reducing conditions (lane 1), whereas His_6-ICL1 appeared homogeneous under non-reducing conditions (lane 2). **C – Use of TCEP in the lysis buffer.** 0.5 mM TCEP was included in the lysis buffer only. His_6-ICL1 was then purified using **IMAC Method 1**, as described in **Figure 26**. Lanes 1—3 show the His_6-ICL1 elution peak. **D – Use of TCEP in the lysis, binding and elution buffers.** 0.5 mM TCEP was included in the lysis buffer. His_6-ICL1 was then purified using **IMAC Method 1**, as described in **Figure 26** except that 0.5 mM TCEP was added to the binding and elution buffers. Lanes 1—3 show the His_6-ICL1 elution peak. Panels separated by white space were taken from separate images.

4.2.1.4. Disruption of protein-nickel interactions during sample application

Since strategies to disrupt protein-protein interactions did not improve the purification of His_6-ICL1, we moved on to considering interactions between *E. coli* proteins and the chromatography medium, specifically the nickel ions. Patterns of surface histidine residues lead some *E. coli* proteins to display high affinity for Ni^{2+} ions.[512,513] Such proteins bind to Ni-nitrilotriacetic acid (Ni-NTA) IMAC resins during sample application and can then co-elute with the hexahistidine-tagged protein of interest.[510] In some cases, contaminant binding to IMAC resin during sample application can be reduced by increasing the concentration of imidazole used to pre-equilibrate the resin, creating

greater competition for Ni^{2+} ions. With this in mind, purification of His_6-ICL1 was again attempted using **IMAC Method 1**, raising the concentration of imidazole in the lysis and binding buffers to 25 mM (from 10 mM).

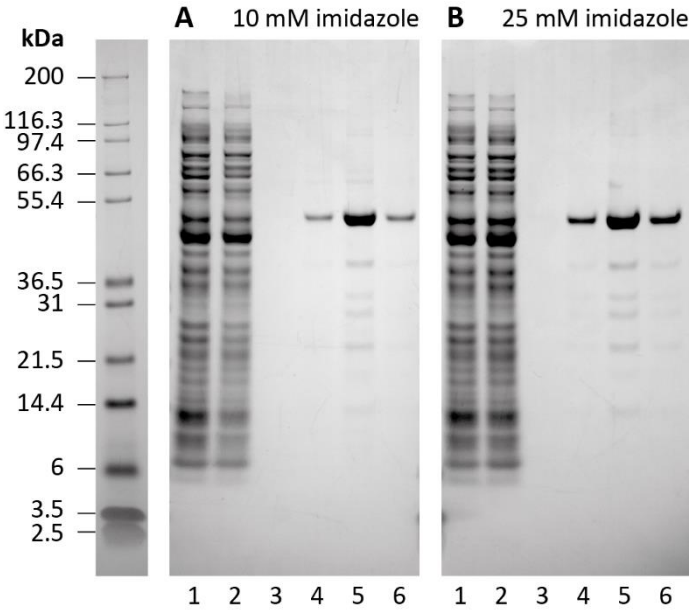


Figure 31: Performing cell lysis and sample application at different imidazole concentrations had no effect on His_6-ICL1 purity.
Cell lysis was performed by sonication using buffer containing either 10 mM or 25 mM imidazole. His_6-ICL1 was then purified using **IMAC Method 1**, as described in **Figure 26**. The imidazole concentration in the binding buffer was the same as in the lysis buffer. **A – Lysis and binding buffers containing 10 mM imidazole.** Lanes 1 and 2, fractions collected during sample application; lane 3, fractions collected during the wash step; lanes 4—6, fractions from the His_6-ICL1 elution peak. **B – Lysis and binding buffers containing 25 mM imidazole.** Lanes 1 and 2, fractions collected during sample application; lane 3, fractions collected during the wash step; lanes 4—6, fractions from the His_6-ICL1 elution peak. Panels separated by white space were taken from separate images.

Increasing the starting concentration of imidazole to 25 mM did not improve the purification (**Figure 31**) – the pattern and relative amounts of contaminant proteins were the same as for a starting imidazole concentration of 10 mM. This observation suggests that neither binding of His_6-ICL1 nor binding of the contaminants to the resin were significantly affected by changing the starting imidazole concentration.

4.2.1.5. Use of anion exchange chromatography as the capture step

As changes to the IMAC method failed to improve the His_6-ICL1 purification, we explored the use of ion exchange chromatography (IEX) to determine whether changing the separation principle could facilitate removal of the contaminant proteins. IEX is suitable for use at any stage of a Capture, Intermediate Purification and Polishing (CIPP) approach to multi-step protein purification.[514] However, protein samples for IEX must be in a low salt buffer, as elution is performed by increasing the ionic strength of the buffer. Consequently, where IEX follows another chromatography step such as IMAC, in which salt is present in the buffer, desalting is required to prepare the sample for IEX. In contrast, IMAC is compatible with higher salt concentrations, meaning that IEX eluates can often be applied directly to IMAC columns. As a result, using IEX before IMAC leads to a more streamlined purification workflow. For this reason, we explored the use of IEX in His_6-ICL1 capture.

Since the predicted pI of His_6-ICL1 is 5.28, the most appropriate starting point for IEX method development was anion exchange chromatography (AIEX) using a strong anion exchanger at pH 8. The pH was selected to match our IMAC method, as this would allow AIEX eluates to be subjected to intermediate purification by IMAC without the need for additional sample handling steps. Tris was selected as the buffer, as it is compatible with His_6-ICL1, AIEX at pH 8, and continuation to intermediate purification by IMAC. A standard AIEX method was written according to the guidelines in the Cytiva IEX handbook and purification of His_6-ICL1 was attempted.[515] LDS-PAGE analysis of the AIEX eluate (**Figure 32**) found that the fractions containing His_6-ICL1 also contained the contaminant proteins seen after IMAC. Having repeatedly failed to separate this pattern

of bands by IMAC, we did not continue to intermediate purification. For the same reason, the alternative sequence of steps – IMAC, desalting, then AIEX – was not considered further.

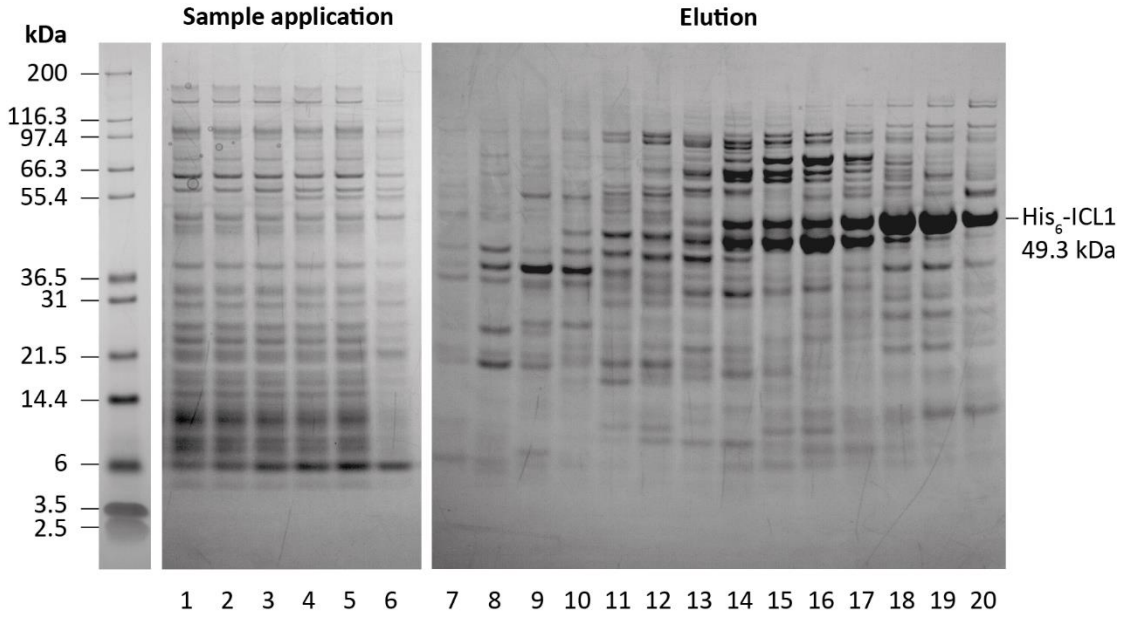


Figure 32: Use of anion exchange chromatography as the capture step in His_6-ICL1 purification.
Cells were resuspended in AIEX lysis buffer (20 mM Tris, pH 8) containing protease inhibitors, then lysed by sonication and centrifuged. The supernatant was filtered and applied to a HiTrap Q HP 1 mL column equilibrated with 20 mM Tris (pH 8). The column was washed with 20 mM Tris (pH 8, 10 CV) prior to gradient elution (0—50% elution buffer over 20 CV; elution buffer, 20 mM Tris, 1 M NaCl, pH 8). Lanes 1—6 show fractions collected during the sample application phase, and indicate that (as expected) His_6-ICL1 bound to the anion exchange resin. Lanes 7—20 show fractions collected during the elution phase. His_6-ICL1 co-eluted with proteins at the same molecular weight as the persistent IMAC contaminants, in addition to other *E. coli* proteins. The pattern is most apparent in lane 19. Panels separated by white space were taken from separate images.

4.2.1.6. Use of alternative *E. coli* expression strains

In the experiments discussed so far, a set of contaminant proteins consistently co-eluted with His_6-ICL1 and could not be removed chromatographically. Since proteolysis of His_6-ICL1 did not appear to be the source of these proteins (**Section 4.2.1.1**), they must have originated from the expression strain, *E. coli* BL21(DE3)Star. Different expression strains were therefore tested to identify whether the problem was unique to the BL21(DE3)Star strain. Four *E. coli* expression strains, NiCo21(DE3), BL21-Gold(DE3),

Rosetta(DE3), and BL21(DE3)pLysS were selected for compatibility with our existing His_6-ICL1 construct and expression protocol. His_6-ICL1 expressed well in all four strains (**Figure 33**).

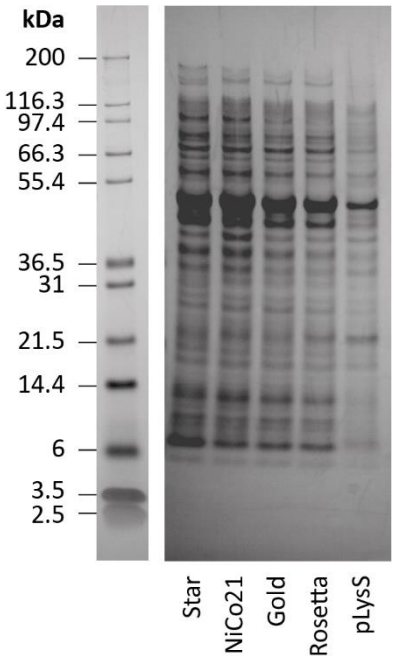


Figure 33: Expression of His_6-ICL1 in different strains of *E. coli*.
Recombinant His_6-ICL1 was expressed in five strains of *E. coli*. In all cases, LDS-PAGE analysis showed that a protein of the expected molecular weight of His_6-ICL1 (49.3 kDa) was present in the soluble fraction of the cell lysate. Star, BL21(DE3)Star; NiCo21, NiCo21(DE3); Gold, BL21-Gold(DE3); Rosetta, Rosetta (DE3); pLysS, BL21(DE3)pLysS. Panels separated by white space were taken from separate images.

Purification of His_6-ICL1 from each strain was then attempted using **IMAC Method 1**. As shown in **Figure 34**, changing the expression strain made little difference to the outcome of the purification – the pattern of contaminant proteins seen in BL21(DE3)Star samples was clearly visible in the NiCo21(DE3), BL21-Gold(DE3), and BL21(DE3)pLysS samples. The pattern also appeared to be present in the Rosetta(DE3) sample, but was less obvious than in the other samples due to co-elution of a large number of other proteins with His_6-ICL1. Consequently, changing the expression strain was discarded as a strategy for improving the His_6-ICL1 purification.

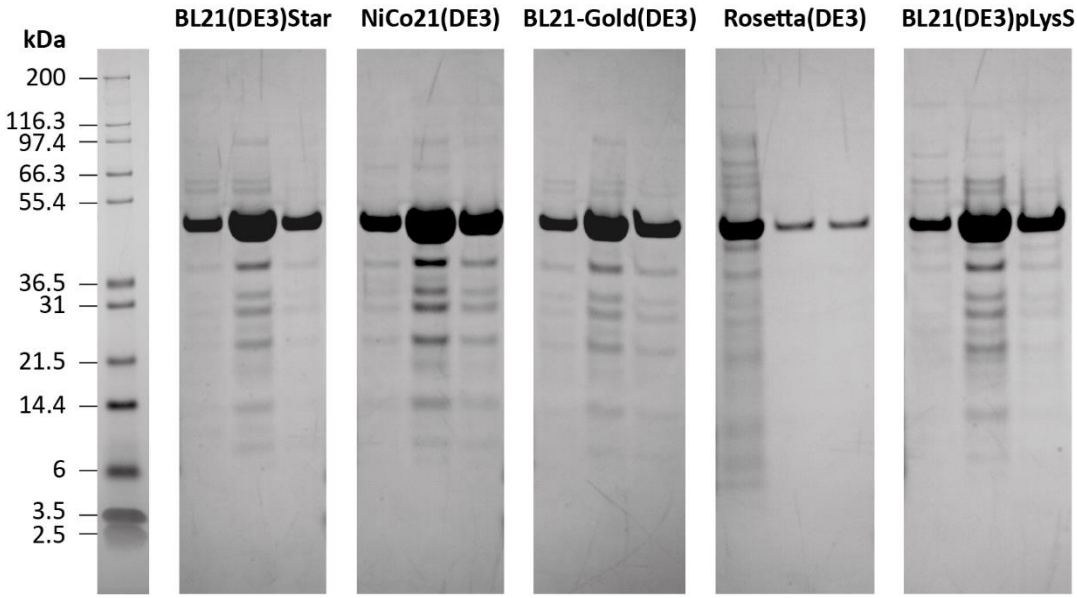


Figure 34: Purification of His_6-ICL1 expressed in different strains of *E. coli*.
In each case, His_6-ICL1 was purified using **IMAC Method 1**, as described in **Figure 26**. Gels are annotated with the *E. coli* strain used to express the enzyme, and the fractions shown correspond to the His_6-ICL1 elution peak. The pattern of co-eluting proteins was consistent between *E. coli* strains, but was less obvious in Rosetta(DE3) owing to the greater background co-elution. Panels separated by white space were taken from separate images or from different parts of the same image.

4.2.1.7. Identification of the contaminant proteins by proteomics

As the pattern of contaminant proteins was consistently and reproducibly observed under all experimental conditions tested, we considered whether their co-elution with His_6-ICL1 was indicative of biological relevance. His_6-ICL1 was again expressed in *E. coli* BL21(DE3)Star and purified using **IMAC Method 1** (**Figure 35A**). On LDS-PAGE analysis, His_6-ICL1 and seven contaminant proteins were clearly visible (**Figure 35B**). These eight bands were excised from the gel and submitted to the Proteomics Science Technology Platform at the Crick, where trypsin digestion and mass spectrometry were performed by David Frith. David then searched the data against the UniProt all species database using Mascot, and provided us with a Scaffold file containing the results. We continued the analysis by considering the hits more abundant than trypsin that were identified as *E. coli* or Mtb proteins. For each band, the mass of each hit was compared to the mass

estimated by LDS-PAGE and hits that were clearly unrealistic were discarded. Putative assignments could then be made for each band (**Table 12**).

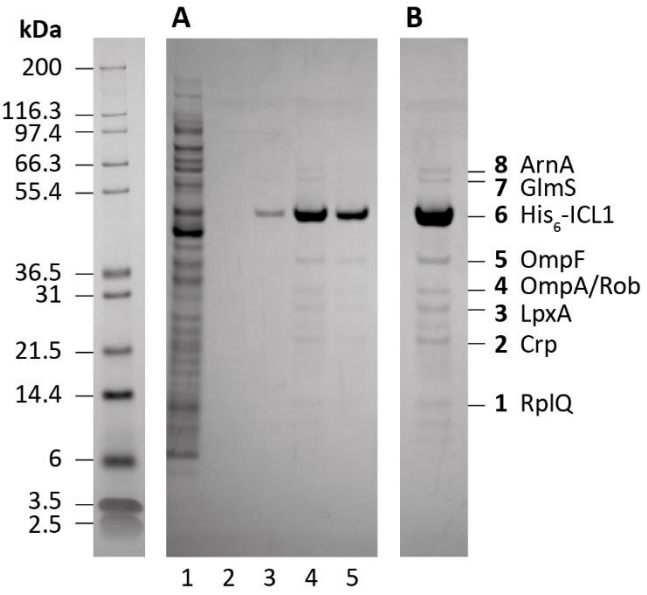


Figure 35: Identification of eluate components by proteomics.
A – Purification of His_6-ICL1. His_6-ICL1 was purified using **IMAC Method 1**, as described in **Figure 26**. Lane 1, sample application; lane 2, wash step; lanes 3—5, His_6-ICL1 elution peak. **B – Eight bands were excised from the gel and identified by proteomics.** Bands are numbered in order of increasing molecular weight and the most likely assignment(s) given. Panels separated by white spaces were taken from separate images or from different parts of the same image.

To begin, we confirmed that band 6 corresponded to Mtb ICL1. As Mtb ICL1 peptides were also abundant in the other seven samples, sequence coverage was analysed for those bands at lower molecular weight than His_6-ICL1 (bands 1—5) to determine whether *C*-terminal proteolysis of His_6-ICL1 during cell lysis or purification was a contributing factor. Both *N*- and *C*-terminal peptides were present in all samples analysed, indicating that proteolytic degradation of His_6-ICL1 did not occur during purification. This result is consistent with our earlier experiments using protease inhibitors (**Section 4.2.1.1**), which suggested that *C*-terminal proteolysis was not responsible for the persistent contaminants at lower molecular weight than His_6-ICL1.

Table 12: Assignment of the eight bands submitted to proteomics.
For each band, the most abundant valid hits arising from Mascot searching of the UniProt all species database are listed. Valid hits were those from either Mtb or *E. coli* and of approximately the correct molecular weight for the band in question. Band numbers correspond to those given in **Figure 35B**. The most likely assignment for each band is listed in bold type. All hits listed were more abundant than trypsin except for RplQ (9 unique peptides compared to 11 unique peptides for trypsin in band 1).

Band	Estimated mass by LDS-PAGE/kDa	Assignment			
		Organism	Protein	Mass/kDa	Number of unique peptides
8	66	***E. coli***	**ArnA**	**74**	**28**
		E. coli	DnaK	69	10
		E. coli	GlmS	67	6
7	55—66	***E. coli***	**GlmS**	**67**	**27**
		E. coli	SdhA	64	16
		E. coli	GroL	57	16
6	47	**Mtb**	**ICL1**	**49**	**71**
5	36—45	***E. coli***	**OmpF**	**39**	**11**
		E. coli	HisB	40	9
4	31—36	***E. coli***	**OmpA**	**37**	**10**
		E. coli	**Rob**	**33**	**10**
		E. coli	YraL	31	7
3	30	***E. coli***	**LpxA**	**37**	**6**
		E. coli	PurU	32	5
		E. coli	RsuA	26	5
		E. coli	DeoR/GntR family	39	5
		E. coli	NanK	30	5
2	25	***E. coli***	**Crp**	**24**	**35**
1	14	***E. coli***	**RplQ**	**14**	**9**

We next proceeded to consider the other bands in more detail, in order of decreasing molecular weight, and assign them to *E. coli* proteins. The most abundant valid hits in band 8 were the bifunctional polymyxin resistance protein ArnA and the chaperone

protein DnaK. Although at 74 kDa, ArnA is a little larger than expected for band 8 (estimated mass 66 kDa), the unique peptide count for ArnA was almost three times higher than for DnaK (69 kDa). ArnA peptides were also seen in four other bands (while DnaK peptides were not), indicating that ArnA was likely a component of the sample. The third most abundant hit for band 8, glucosamine-fructose-6-phosphate aminotransferase (GlmS) was discounted, as a higher number of unique peptides were present in band 7 (27 unique peptides in band 7 compared to 6 in band 8). Band 8 was therefore assigned as primarily containing ArnA. GlmS, meanwhile, was identified as the main protein present in band 7, being more abundant than the other valid hits (the succinate dehydrogenase flavoprotein subunit SdhA and GroL, a chaperonin).

The most abundant valid hits for band 5 were outer membrane protein F (OmpF, 11 unique peptides) and imidazole glycerol-phosphate dehydrogenase (HisB, 9 unique peptides); band 5 was consequently assigned as OmpF. Peptides from outer membrane protein A (OmpA) and right origin-binding protein (Rob) were equally abundant in band 4 (10 unique peptides), so a single assignment could not be made. However, OmpA peptides were also seen in bands 3 and 5, whereas Rob peptides were not, suggesting that OmpA was perhaps overall more abundant in the sample.

UDP-*N*-acetylglucosamine acyltransferase (LpxA) was the most abundant valid hit for band 3. Band 3 was also found to contain peptides from formyltetrahydrofolate deformylase (PurU), pseudouridine synthase (RsuA), a DeoR/GntR family transcriptional regulator, and *N*-acetylmannosamine kinase (NanK). There was only one valid hit for band 2: cAMP-activated global transcriptional regulator, Crp. Finally, band 1 had no valid

hits more abundant than trypsin and was tentatively assigned as 50S ribosomal protein (RplQ), the most abundant *E. coli* protein in this band.

The three most abundant *E. coli* proteins identified – Crp, ArnA, and GlmS – are common IMAC contaminants which bind Ni-NTA resin through clusters of surface-exposed histidine residues.[512,513,516] Since elution of these proteins from Ni-NTA resin requires imidazole concentrations in excess of 55 mM, it is unsurprising that the use of 25 mM imidazole in the lysis buffer (**Section 4.2.1.4**) or in a wash step (**Section 4.2.1.1**) failed to remove them.[513] Importantly, none of the *E. coli* proteins identified are known to be involved in isocitrate metabolism, and there is no obvious metabolic or regulatory connection underlying their co-elution with His_6-ICL1. As these proteins are unlikely to interfere with our proposed His_6-ICL1 assays, the level of purification achieved is likely sufficient for the intended uses of the purified enzyme.[g]

4.2.1.8. Use of higher resolution IMAC medium: development of the final IMAC method

Some *E. coli* proteins, such as GroL, are common IMAC contaminants despite not possessing the surface histidine clusters or metal-binding sites required to engage in binding to nickel ions in Ni-NTA resins. While the mechanisms by which these proteins

[g] In the expression strain *E. coli* NiCo21(DE3), several metal-binding proteins are mutated or tagged in order to reduce contamination of IMAC fractions. These include GlmS, which is mutated such that it no longer binds IMAC resin, and ArnA, which bears a chitin binding domain (CBD) tag and can therefore be removed by chitin affinity chromatography.[512] The effects of these modifications can be seen when comparing our attempts at purifying His_6-ICL1 expressed in BL21(DE3)Star and NiCo21(DE3) (**Figure 34**). Proteomics identified the bands at 55—66 kDa and ~66 kDa in the BL21(DE3)Star gel as GlmS and ArnA respectively. In the NiCo21(DE3) gel, one of these bands is absent, while the other appears at a higher molecular weight. The absent band is presumably due to the GlmS mutant not binding to the IMAC resin, with the band at higher apparent molecular weight being CBD-tagged ArnA. Since the majority of the contaminant proteins are of lower molecular weight than His_6-ICL1, are more abundant than ArnA, and are not metal-binding proteins, we did not consider it worthwhile to attempt removal of CBD-tagged ArnA from His_6-ICL1 purified from NiCo21(DE3) lysate.

co-purify with hexahistidine-tagged proteins have yet to be fully elucidated, it has been theorised that hydrophobic interactions with Sepharose in the chromatography medium may be involved. This hypothesis is supported by the observation of these contaminants in other types of affinity purification that use different Sepharose chemistries; for example, co-purification with GroL has been noted when using Glutathione Sepharose resin.[513] Since GroL was among the His_6-ICL1 contaminants identified in the proteomics experiment discussed above (**Section 4.2.1.7**), we wondered whether protein-Sepharose interactions might be partly responsible for the persistent contamination of His_6-ICL1 with *E. coli* proteins.

As interactions of this type would vary according to the physicochemical properties of the chromatography medium, changing the medium might alter which proteins co-elute with His_6-ICL1. Up to this point, our His_6-ICL1 IMAC method development had employed Cytiva HisTrap FF columns, which contain Ni Sepharose 6 Fast Flow (FF) medium. As an alternative, we selected Cytiva HisTrap HP columns, which are prepacked with Ni Sepharose High Performance (HP) medium. Both FF and HP media comprise a highly cross-linked 6% spherical agarose matrix charged with Ni^{2+} (~15 µmol Ni^{2+}/mL medium), with the key difference between the two being the average particle size in the matrix. The median particle diameter is 90 µm in FF medium and 34 µm in HP medium, meaning that HP medium offers higher chromatographic resolution.[511] Of note, the A_{280} traces (data not shown) and LDS-PAGE analyses of previous His_6-ICL1 purifications did not give any indication (such as peak shouldering or asymmetry) that increasing resolution might facilitate separation of the contaminant proteins from His_6-ICL1. Nonetheless, we felt it

worth exploring the possibility that changing the media might improve His_6-ICL1 purity by reducing non-specific protein-Sepharose interactions.

In addition to changing the column type, we made two modifications to **IMAC Method 1**, resulting in **IMAC Method 3** (**Figure 36**). The column wash step was split into two parts (10 CV 100% buffer A followed by 10 CV 5% buffer B), and a shallower elution gradient was used (over 15 CV rather than 5 CV). Purification of His_6-ICL1 using **IMAC Method 3** resulted in apparently homogeneous protein (**Figure 37A**). However, as the sample was more dilute than many previously analysed during method development, we examined purity by LDS-PAGE at increasing protein concentrations (**Figure 37B**). At higher concentrations of His_6-ICL1, the persistent pattern of contaminants can be faintly seen after staining overnight; these proteins are far less abundant than with **IMAC Method 1**.

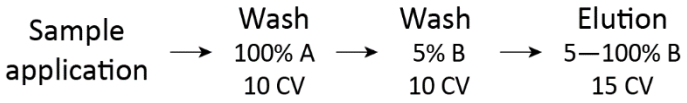


Figure 36: IMAC Method 3.
The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (10 CV), followed by 5% buffer B (10 CV) prior to gradient elution (5—100% B over 15 CV). Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

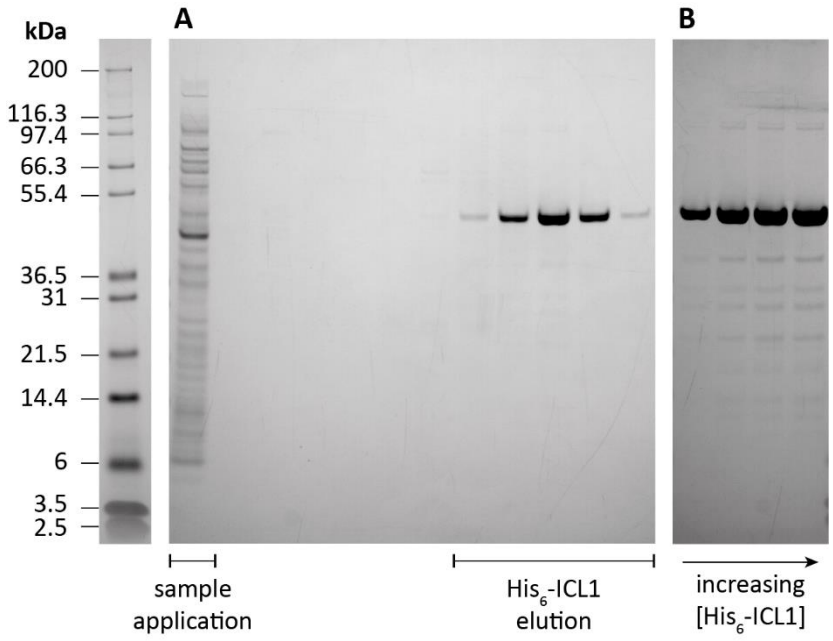


Figure 37: Purification of His_6-ICL1 using IMAC Method 3.
A – Purification. His_6-ICL1 was purified using **IMAC Method 3**, as described in **Figure 36**. **B – LDS-PAGE analysis of the eluate at increasing concentrations.** Increasing concentrations of purified His_6-ICL1 were analysed to check for the presence of co-eluting proteins. While co-eluting proteins were seen at high concentrations of His_6-ICL1, these were far less abundant than with previous methods. Panels separated by white space were taken from separate images.

In conclusion, **IMAC Method 3** (using a HisTrap HP column) yielded higher purity His_6-ICL1 than **IMAC Method 1** (using a HisTrap FF column). Since both the FPLC method and the chromatography medium were altered, it is unclear which change produced this result and whether differences in non-specific protein-Sepharose interactions between FF and HP media might have been relevant. However, the development of **IMAC Method 3** facilitated the purification of His_6-ICL1 to a standard more than adequate for steady-state kinetics experiments, and we determined that no further method development was required. His_6-ICL1 was purified on multiple further occasions during assay development, and **IMAC Method 3** proved to be reproducible in terms of enzyme purity, as assessed by LDS-PAGE (n = 8).

4.2.1.9. Buffer exchange: dialysis vs gravity size exclusion chromatography

After purification using **IMAC Method 3**, His_6-ICL1 was buffer-exchanged in preparation for steady-state kinetics. Buffer exchange was initially performed using dialysis, but in later experiments the method was changed to gravity SEC. Gravity SEC is considerably faster than dialysis (45 minutes compared to 18 hours), and is therefore more compatible with the purification throughput required in the multispecies enzymology study for which the methods developed in this chapter are eventually intended. Additionally, we found that gravity SEC yielded higher quality protein than dialysis, as measured by the A_{280}/A_{250} ratio.[h] The mean A_{280}/A_{250} reading was 2.0 for gravity SEC (n = 4) compared to 1.5 for dialysis (n = 4), suggesting that some protein aggregation occurred during dialysis, but not during gravity SEC.

Enzyme yields were determined after buffer exchange, varied in the range 0.80—5.58 mg His_6-ICL1/g wet cells (n = 8), and were typically in excess of 2.5 mg His_6-ICL1/g wet cells. Unusually low yields were attributable to visibly incomplete cell lysis by sonication. Changing the buffer exchange method made little difference to the yield, with the mean yield being 3.17 mg His_6-ICL1/g wet cells for dialysis (n = 4) and 3.38 mg His_6-ICL1/g wet cells for gravity SEC (n = 4). The yields given above correspond to 4—17 mg His_6-ICL1/L culture (n = 8), lower than the literature value of ~20 mg

[h] Protein near-UV absorbance spectra have a maximum at 280 nm and a minimum at approximately 250 nm. The A_{280}/A_{250} ratio can be used to check for Rayleigh scattering due to the presence of protein aggregates. Rayleigh scattering is proportional to λ^{-4} and is therefore more pronounced at lower wavelengths (λ), meaning that A_{250} increases by more than A_{280}. As a result, protein samples containing aggregates have a lower A_{280}/A_{250} value ($A_{280}/A_{250} < 2$) than unaggregated samples ($A_{280}/A_{250} \geq 2$). Consideration of protein aggregation and the resulting Rayleigh scattering is useful for two reasons. Firstly, aggregation may affect enzyme activity. Secondly, if scattering is extensive, correction of the UV spectrum may be required prior to A_{280}-based protein concentration calculations. Aggregation can alternatively be assessed using the 320—350 nm region of the UV spectrum, where unaggregated small proteins should not absorb light.[555–559]

His_6-ICL1/L culture, but more than satisfactory for our requirements.[318] Having established a reproducible method for the purification of His_6-ICL1 to sufficient purity (as judged by eye from LDS-PAGE gels), we proceeded to ICL assay selection and implementation.

4.2.2. Determination of kinetic parameters for His_6-ICL1

There are two commonly used methods for the spectrophotometric determination of ICL activity, both of which measure the amount of glyoxylate produced by the enzyme (**Figure 38**). The phenylhydrazone assay (**Figure 38A**) measures the increase in absorbance at 324 nm associated with the formation of glyoxylate phenylhydrazone from glyoxylate and phenylhydrazine.[329,517] This assay, which may be run in either a continuous or a discontinuous format, has previously been used to characterise ICL1 from Mtb and *M. smegmatis*.[263,267,334] Alternatively, the glyoxylate produced by ICL can be converted to glycolate by lactate dehydrogenase (LDH) (**Figure 38B**). Concomitant oxidation of the co-substrate, NADH, results in a decrease in the absorbance at 340 nm, which is monitored continuously.[263,518] So that the rate of the coupled reaction is not limiting, LDH and NADH must be present in excess.[519]

Like the phenylhydrazone assay, the LDH assay has been used to characterise a range of ICLs, including Mtb ICL1.[318,319] A potential drawback of both methods is that chromophore instability can cause a drift in the background UV absorbance – phenylhydrazine is unstable at pH 7 and above, while NADH oxidises slowly in solution.[329] However, both assays are faster than the proposed alternative, quantification of succinate production by NMR, making them more suitable for the throughput required later in this project.[329] Ultimately, the LDH assay (**Figure 38B**) was

selected for the characterisation of mycobacterial ICLs, as this method had previously been implemented in our laboratory for the determination of ICL activity in cell lysates.[453]

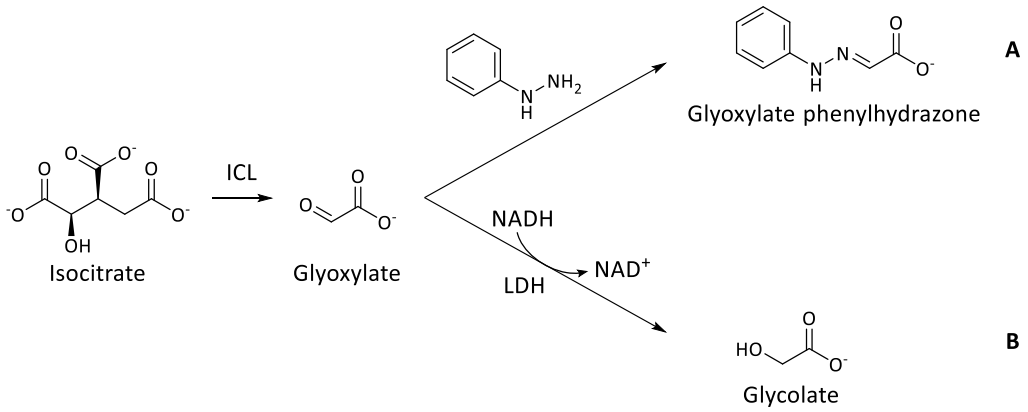


Figure 38: Methods for the spectrophotometric determination of ICL activity.
ICLs convert isocitrate to glyoxylate and succinate; the enzyme activity can be determined by quantifying glyoxylate production. **A – Phenylhydrazone assay.** Glyoxylate reacts with phenylhydrazine to form glyoxylate phenylhydrazone, resulting in an increase in the absorbance at 324 nm. **B – LDH-coupled assay.** LDH converts glyoxylate to glycolate. Concomitant oxidation of the co-substrate, NADH, results in a decrease in the absorbance at 340 nm. ICL, isocitrate lyase; LDH, lactate dehydrogenase.

Accordingly, His_6-ICL1 activity was determined by coupling glyoxylate production to LDH, using a protocol based on the literature methods.[318,363] Full experimental procedures are provided in **Chapter 8**. Briefly, each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 25—1500 µM and reactions were initiated with 30 nM His_6-ICL1 (the enzyme concentration used in the literature[318]). Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation (**Table 13** and **Figure 39**).

Table 13: Kinetic parameters for His_6-ICL1.
Michaelis-Menten kinetic parameters determined by coupling His_6-ICL1 activity to LDH (**Figure 39**). Errors are reported as symmetrical 95% confidence intervals.

v_{max} (µM min^{-1})	K_M(DL-isocitrate) (µM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
40 ± 2	390 ± 59	22 ± 1	$5.7\times10^4 \pm 9\times10^3$

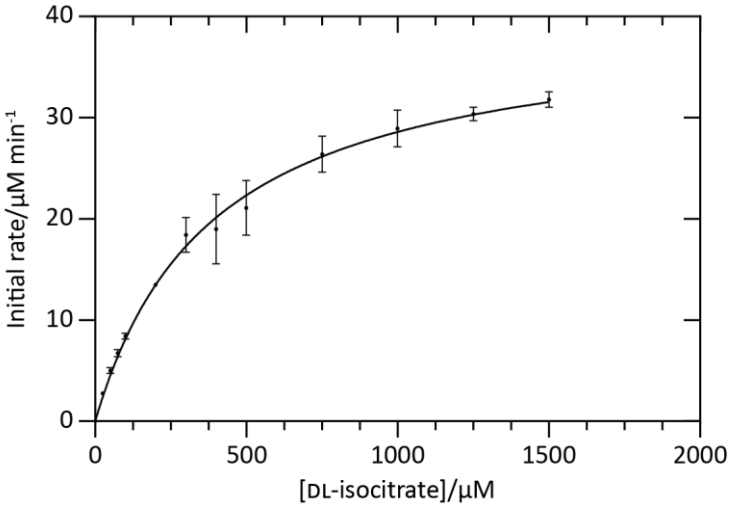


Figure 39: Steady-state kinetics of His_6-ICL1.
Initial velocity data for His_6-ICL1 at isocitrate concentrations in the range 25—1500 µM. Reactions were carried out as described in **Chapter 8**, and were performed in quadruplicate. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation (solid line).

Comparison of these parameters to the literature data is of limited value, as they are influenced by the exact form of the enzyme used (for example the type and position of affinity tags) and the assay conditions (such as temperature). Indeed, as presented in **Chapter 1**, the reported parameters for Mtb ICL1 vary widely. Considering only those reports where DL-isocitrate was used as the substrate, values for K_M(DL-isocitrate) vary in the range 45 µM—1.14 mM, while k_{cat} varies in the range 4.3—9.8 s^{-1} and k_{cat}/K_M varies in the range 8.6×10^3—1.2×10^5 M^{-1} s^{-1}.[284,305,311,363] Thus, the values we obtained for K_M(DL-isocitrate) (390 ± 59 µM) and k_{cat}/K_M ($5.7\times10^4 \pm 9\times10^3$ M^{-1} s^{-1}) fall within the previously reported range. Our higher k_{cat} value (22 ± 1 s^{-1}) is most likely the result of the assay temperature (37 °C; literature 27—30 °C, where specified[305,311]) combined

with our enzyme handling strategy. Our His_6-ICL1 purifications and assays were completed within 36 hours of cell lysis, keeping samples at 4 °C to minimise the loss of enzymatic activity that can occur on freezing. In contrast, literature preparations of ICL1 were typically frozen prior to kinetic analysis.

4.3. Isocitrate lyase 2

4.3.1. Cloning and expression of full-length ICL2-His_6FLAG

AceA, the gene encoding ICL2, was amplified from Mtb CDC 1551 genomic DNA by PCR and assembled into the pNIC-CTHF vector, which provides a TEV cleavable *C*-terminal His_6FLAG tag. The resulting plasmid was used to transform *E. coli* BL21(DE3)Star and protein expression was performed using the same protocol as for Mtb His_6-ICL1 above (**Section 4.2.1**). Recombinant ICL2 expressed well and was soluble (**Figure 40**).

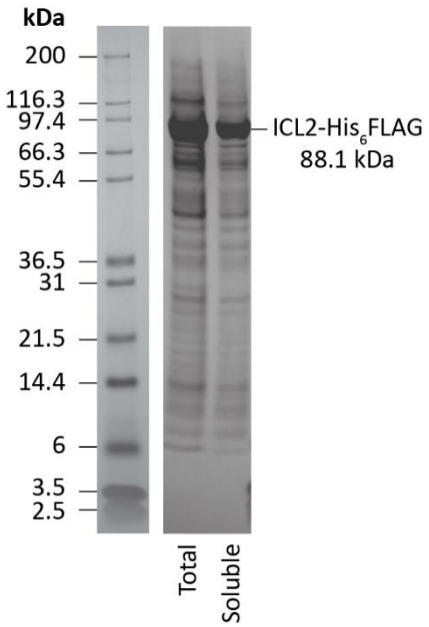


Figure 40: Expression of ICL2-His_6FLAG.
Recombinant ICL2-His_6FLAG was expressed in *E. coli* BL21(DE3)Star as described in **Chapter 8**. LDS-PAGE analysis showed that a protein of the expected molecular weight of ICL2-His_6FLAG (88.1 kDa) was present in both the complete cell lysate (**Total**) and in the soluble fraction (**Soluble**). Panels separated by white space were taken from separate images.

4.3.2. Purification of full-length ICL2-His$_6$FLAG

4.3.2.1. Initial attempts at purification using IMAC

As for His$_6$-ICL1, **IMAC Method 1** (**Figure 26**) was used as the starting point for method development. While this method was successful in enriching ICL2-His$_6$FLAG, the enzyme co-purified with a number of other proteins of lower molecular weight (**Figure 41A**). This observation was in contrast to the literature purification of truncated ICL2-His$_6$, in which homogeneous enzyme was obtained in one step by IMAC.[317]

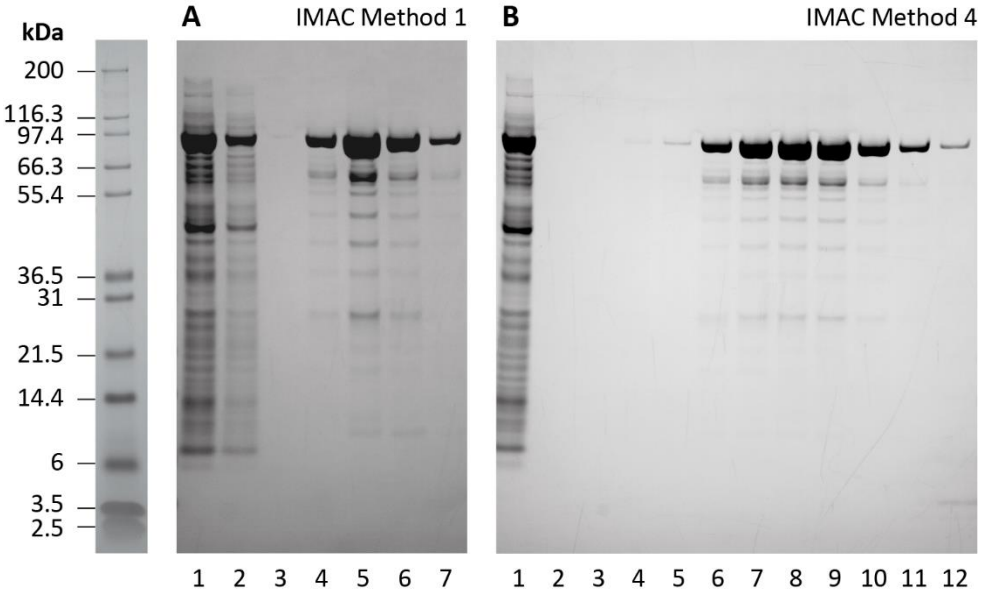


Figure 41: Initial attempts at purification of ICL2-His$_6$FLAG by IMAC.
A – Purification using IMAC Method 1. ICL2-His$_6$FLAG was purified using **IMAC Method 1**, as described in **Figure 26**. Lanes 1 and 2, sample application; lane 3, wash step; lanes 4—7, ICL2-His$_6$FLAG elution peak. **B – Purification using IMAC Method 4.** ICL2-His$_6$FLAG was purified using **IMAC Method 4**, as described in **Figure 42**. Lane 1, sample application; lanes 2—4, wash step; lanes 5—12, ICL2-His$_6$FLAG elution peak. Panels separated by white space were taken from separate images.

Comparing **IMAC Method 1** to the literature method, one notable difference is the steepness of the elution gradient – the imidazole concentration increases by 100 mM/CV in **IMAC Method 1**, but just 8.5 mM/CV in the literature method. The elution phase in **IMAC Method 1** was therefore extended to 30 CV (**Figure 42**), corresponding to an increase in the imidazole concentration of 16.7 mM/CV.

Purification of ICL2-His_6FLAG was then attempted using the revised method, **IMAC Method 4**. A single symmetrical protein peak was observed during the elution phase, and analysis of its constituent fractions by LDS-PAGE (**Figure 41B**) revealed the same pattern of contaminant proteins observed with **IMAC Method 1**. The concentrations of these co-eluting proteins track with the concentration of ICL2-His_6FLAG, suggesting that further modification of the elution gradient would not be a useful strategy for their elimination.

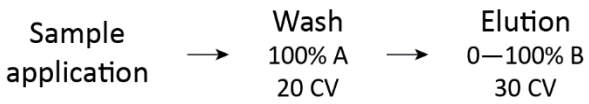


Figure 42: IMAC Method 4.
The sample was applied to a HisTrap FF 1 mL column equilibrated with buffer A. The column was washed with buffer A (20 CV) prior to gradient elution (0—100% B over 30 CV). Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

4.3.2.2. Use of anion exchange chromatography as a polishing step

Rather than making further modifications to the IMAC method, we explored whether the use of IEX as a polishing step would yield higher purity ICL2-His_6FLAG. As the predicted pI of ICL2-His_6FLAG is 5.80, AIEX was identified as a suitable initial approach. Capture of ICL2-His_6FLAG was performed using **IMAC Method 1** (**Figure 43**, lanes 1—5). Fractions containing ICL2-His_6FLAG were pooled and desalted (**Figure 43**, lane 6), then subjected to AIEX using the method previously developed for His_6-ICL1 (**Section 4.2.1.5**). However, the contaminant proteins co-eluted with ICL2-His_6FLAG (**Figure 43**, lanes 7—12). Comparing IMAC and AIEX fractions with similar ICL2-His_6FLAG concentrations, we concluded that the additional step did not substantially improve the purity of ICL2-His_6FLAG.

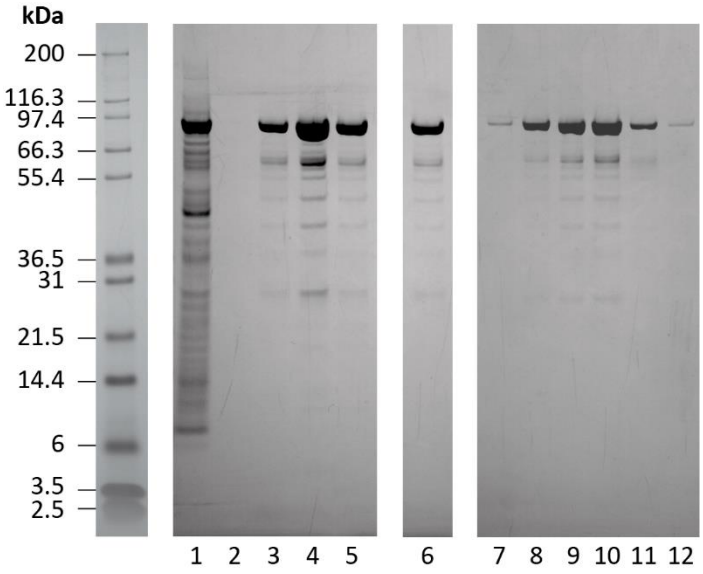


Figure 43: Use of anion exchange chromatography as a polishing step for ICL2-His$_6$FLAG.
ICL2-His$_6$FLAG was purified using **IMAC Method 1**, as described in **Figure 26** (lanes 1—5: lane 1, sample application; lane 2, wash step; lanes 3—5, ICL2-His$_6$-FLAG elution peak). Fractions containing the enzyme were pooled and desalted (lane 6). The desalted sample was applied to a HiTrap Q HP 1 mL column equilibrated with 20 mM Tris (pH 8). The column was washed with 20 mM Tris (pH 8, 10 CV) prior to gradient elution (0—50% elution buffer over 20 CV; elution buffer, 20 mM Tris, 1 M NaCl, pH 8). Lanes 7—12 show fractions collected during the elution phase. Comparing samples at similar protein concentrations (lanes 6 and 10) shows that the purity of ICL2-His$_6$FLAG was not improved after AIEX. Panels separated by white space were taken from separate images or from different parts of the same image.

4.3.2.3. Use of higher resolution chromatography medium: purification of ICL2-His$_6$FLAG using IMAC Method 3

Purification of ICL2-His$_6$FLAG was next attempted using **IMAC Method 3**, the protocol optimised for His$_6$-ICL1 (**Section 4.2.1.8**). While enzyme purity was slightly improved compared to previous tests, the biggest difference observed was in binding of the enzyme to the medium (**Figure 44**). When FF medium was used (**Section 4.3.2.1**), ICL2-His$_6$FLAG binding was incomplete, as evidenced by the presence of the enzyme in the fractions collected during sample application. The yields of ICL2-His$_6$FLAG calculated later (**Section 4.3.2.4**) were consistently well below the theoretical binding capacity of the medium (40 mg His$_6$-tagged protein/mL resin), indicating that the column had not been saturated. Switching to HP medium significantly reduced the amount of ICL2-His$_6$FLAG seen in the fractions collected during sample application (**Figure 44**).

Since the preceding expression and sample preparation steps were unaltered, the ICL2-His_6FLAG content of the samples is likely to have been similar. Additionally, the sample application phase is consistent between methods and the two media types have the same theoretical binding capacity, indicating that the observed difference in ICL2-His_6FLAG binding was due to the physical properties of the media.

As greater binding of the recombinant enzyme to the medium results in a higher potential yield, **IMAC Method 3** (which uses HP medium) was selected as the preferred method for the purification of ICL2-His_6FLAG. The enzyme purity was considered adequate for steady-state kinetics, so method development was not continued further. ICL2-His_6FLAG was purified multiple times over the course of assay implementation and enzyme characterisation, and **IMAC Method 3** proved to be reproducible in terms of enzyme purity, as assessed by LDS-PAGE (n = 8).

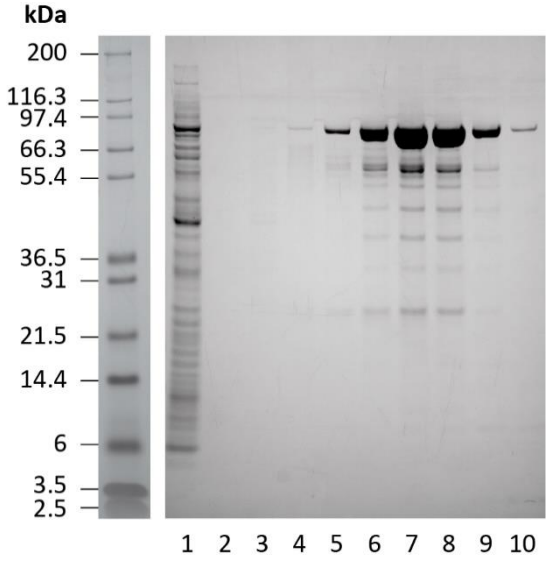


Figure 44: Purification of ICL2-His_6FLAG using IMAC Method 3.
ICL2-His_6FLAG was purified using **IMAC Method 3**, as described in **Figure 36**. Lane 1, sample application; lanes 2—3, wash steps; lanes 4—10, ICL2-His_6FLAG elution peak. Panels separated by white space were taken from separate images.

4.3.2.4. Buffer exchange: dialysis vs gravity SEC

After purification using **IMAC Method 3**, ICL2-His_6FLAG was buffer-exchanged in preparation for steady-state kinetics. As for His_6-ICL1 (**Section 4.2.1.9**), buffer exchange was initially performed using dialysis, with the method subsequently being changed to gravity SEC. Again, we noted that gravity SEC yielded higher quality protein than dialysis, as measured by the A_{280}/A_{250} ratio – the mean A_{280}/A_{250} reading was 5.0 for gravity SEC (n = 6) compared to 1.9 for dialysis (n = 2). Enzyme yields were determined after buffer exchange, varied in the range 9.0—22.3 mg ICL2-His_6FLAG/g wet cells (n = 8), and were typically towards the lower end of this range. As enzyme yields were not reported in the literature, no comparison can be made.[278,305,317]

4.3.3. Determination of kinetic parameters for full length ICL2-His_6FLAG

4.3.3.1. Optimisation of the LDH-coupled assay for ICL2-His_6FLAG

Earlier, we selected the LDH assay for the characterisation of mycobacterial ICLs and used this approach to determine kinetic parameters for His_6-ICL1 (**Section 4.2.2**). To begin the characterisation of ICL2-His_6FLAG, we employed the protocol optimised for His_6-ICL1. Initial trials indicated that – as expected from the literature – ICL2-His_6FLAG had a much higher K_M(DL-isocitrate) than His_6-ICL1. Accordingly, the range of isocitrate concentrations used in the assay was altered from 0.025—1.5 mM (His_6-ICL1) to 0.1—4 mM (ICL2-His_6FLAG). Additionally, ICL2-His_6FLAG was less active than His_6-ICL1, resulting in noisy progress curves, particularly at the lower isocitrate concentrations tested. To resolve this, the concentration of ICL2-His_6FLAG was raised from 30 nM to 200 nM (**Figure 45A**).

However, the Michaelis-Menten fit of the initial rates was poor at both enzyme concentrations tested, with the replicates test for lack of fit identifying statistically significant deviation from the model (30 nM ICL2-His_6FLAG: $p = 0.0036$ and $p = 0.0033$, independent experiments; 200 nM ICL2-His_6FLAG: $p = 0.0018$).[i] Examination of the residual plots (**Figure 45B**) revealed that this deviation was likely to have been caused by the technical setup of the assay. The residuals for the first six isocitrate concentrations exhibit a downward trend. After a sharp jump upwards, the downward trend repeats for the following five samples. The spectrophotometer used to run the assay was equipped with a six-position sample changer, and reactions within a replicate were run in order of increasing isocitrate concentration, meaning that the break in the pattern coincides with reloading of the sample changer. We therefore considered it likely that the deviation from the Michaelis-Menten model was the product of inconsistent thermal equilibration within replicates, as equilibration times varied from 5—17.5 minutes depending on the position of the sample in the run. This hypothesis was confirmed by randomising the samples within each replicate (**Figure 45C** and **Figure 45D**). The Michaelis-Menten fit of the resulting initial rates was good, as assessed by the replicates test for lack of fit ($p = 0.91$), enabling the determination of kinetic parameters for ICL2-His_6FLAG. These will be presented in the next section.

[i] The replicates test for lack of fit is a numerical evaluation of the residual pattern. It compares the scatter of replicates (the random error inherent in the data) to the average distance of the data points from the curve in order to establish whether the observed pattern of data points represents Gaussian variation around the curve. The null hypothesis for the lack-of-fit test is that the model is correct and there is no lack of fit. A small p value suggests that the data follow a different model; $p \leq 0.05$ is typically considered to indicate a statistically significant lack of fit.[560]

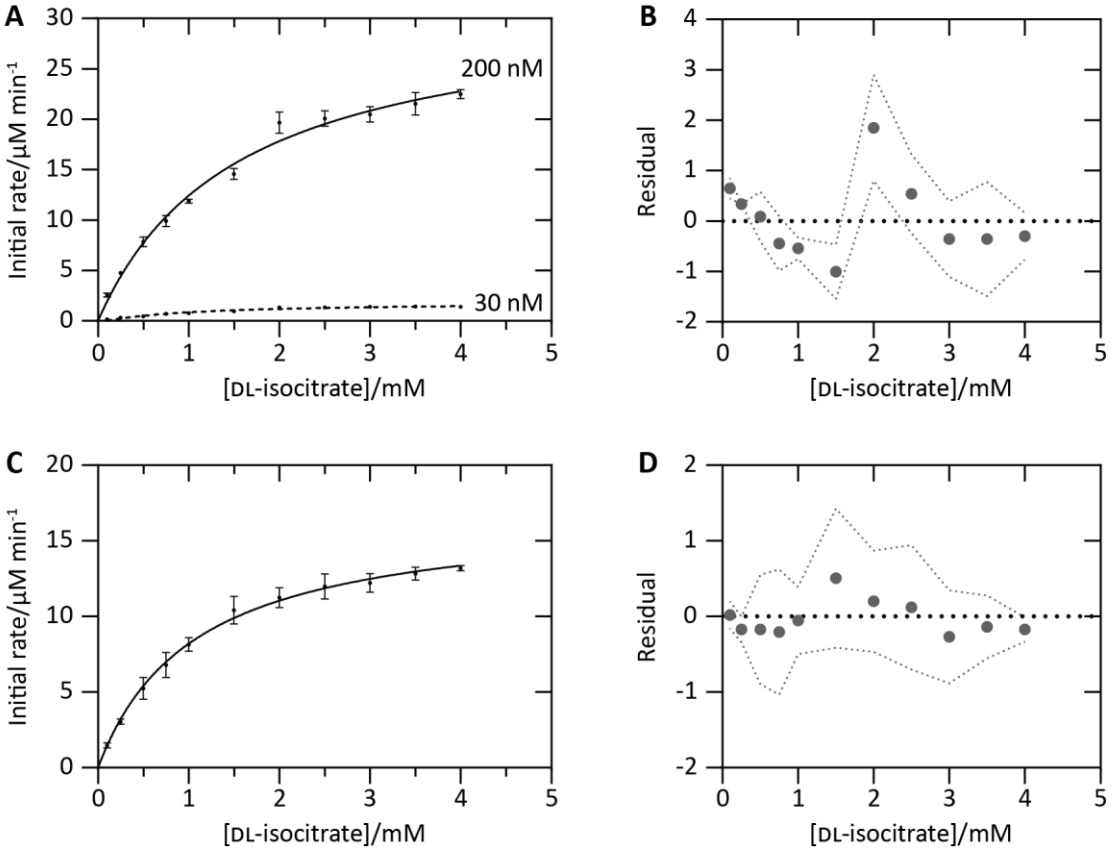


Figure 45: Optimisation of the LDH-coupled assay for ICL2-His$_6$FLAG.
A – Steady-state kinetics of ICL2-His$_6$FLAG at 30 nM and 200 nM. Reactions were performed in triplicate and were initiated with either 30 nM (dashed line) or 200 nM (solid line) ICL2-His$_6$FLAG. As for His$_6$-ICL1, each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—4 mM. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. The replicates test for lack of fit identified statistically significant deviation from the model (30 nM ICL2-His$_6$FLAG: p = 0.0036 and p = 0.0033, independent experiments; 200 nM ICL2-His$_6$FLAG: p = 0.0018). **B – Residual plots identified a systematic deviation from the Michaelis-Menten fit.** The residual plot for the curve in panel **A** obtained using 200 nM ICL2-His$_6$FLAG is shown in this panel; the residual plots for the assays at 30 nM ICL2-His$_6$FLAG exhibited the same zig-zag pattern. The mean (data points) and standard deviation (dotted lines) of the residuals were plotted. **C – Steady-state kinetics of ICL2-His$_6$FLAG at 200 nM, randomising the samples within each replicate.** Reactions were initiated with 200 nM ICL2-His$_6$FLAG and contained the same components as in panel **A**. The assay was performed in triplicate and the order of reactions within each replicate was randomised. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation (solid line). The fit was good, as assessed by the replicates test for lack of fit (p = 0.91), indicating that thermal equilibration was likely the cause of the deviation seen in panels **A** and **B**. **D – Residual plot for the curve in panel C.** The mean (data points) and standard deviation (dotted lines) of the residuals were plotted. The zig-zag pattern evident in panel **B** has been lost.

As discussed earlier (**Chapter 1**), Ac-CoA activates ICL2 by binding to the *C*-terminal domain, triggering a significant conformational change that moves the active site loop into a more catalytically relevant position. While it has been proposed that this

activation may be important for the control of flux through the glyoxylate shunt, it is also possible that the activated (Ac-CoA-bound) form of the enzyme is the usual state of ICL2 *in bacterio*.[305] The optimised assay was therefore performed in the presence and absence of 25 µM Ac-CoA. To summarise, each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—4 mM and reactions were initiated with 200 nM ICL2-His_6FLAG. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation (**Figure 46A** and **Table 14**).

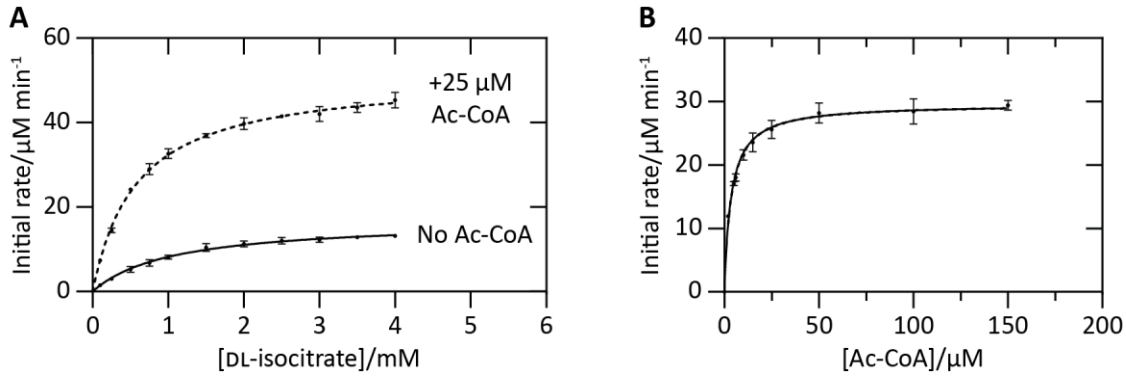


Figure 46: Steady-state kinetics of ICL2-His_6FLAG.
A – Initial velocity data for ICL2-His_6FLAG in the presence and absence of Ac-CoA. The optimised assay was performed in the presence (dashed line) and absence (solid line) of 25 µM Ac-CoA. Each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—4 mM. Reactions were performed in triplicate and were initiated with 200 nM ICL2-His_6FLAG. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. **B – Determination of K_M(Ac-CoA).** Each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and isocitrate (5 mM). The Ac-CoA concentration was varied in the range 2—150 µM. Reactions were performed in triplicate and were initiated with 100 nM ICL2-His_6FLAG. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation.

Next, the assay was adjusted to determine K_M(Ac-CoA). Each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and isocitrate (5 mM). The Ac-CoA concentration was varied in the range 2—150 µM and reactions were initiated with 100 nM ICL2-His_6FLAG. K_M(Ac-CoA) was obtained by fitting the

resultant initial rates to the Michaelis-Menten equation (**Figure 46B**) and was found to be 3.5 ± 0.5 μM, comparable to the literature value of 2.9 ± 0.5 μM.[305]

Table 14: Kinetic parameters for ICL2-His$_6$FLAG and His$_6$-ICL2.
A – Kinetic parameters for ICL2-His$_6$FLAG in the presence and absence of 25 μM Ac-CoA. These parameters were obtained from the Michaelis-Menten fits shown in **Figure 46A**. The assay was performed at 37 °C. Errors are reported as symmetrical 95% confidence intervals. **B – Literature kinetic parameters for His$_6$-ICL2 in the presence and absence of 25 μM Ac-CoA.**[305] These parameters were obtained in an NMR assay at 27 °C and are presented here to facilitate comparison to the data obtained in our UV-vis assay.

A	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
No Ac-CoA	17 ± 1	1100 ± 180	1.4 ± 0.1	$1.3\times10^3 \pm 2\times10^2$
+25 μM Ac-CoA	51 ± 1	580 ± 48	4.3 ± 0.1	$7.4\times10^3 \pm 6\times10^2$

B	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
No Ac-CoA	54 ± 3.9	606 ± 96	0.45 ± 0.1	$7.6\times10^2 \pm 6.8\times10^1$
+25 μM Ac-CoA	55 ± 3.2	125 ± 27	4.6 ± 0.3	$3.7\text{x}10^4 \pm 6.2\text{x}10^3$

As expected, ICL2-His$_6$FLAG was activated by Ac-CoA (**Figure 46A**). A 5.5-fold increase in k_{cat}/K_M was observed in the presence of 25 μM Ac-CoA, resulting from the tripling of k_{cat} and the halving of K_M(DL-isocitrate) (**Table 14A**). However, the degree of activation was lower than reported in the literature, where addition of 25 μM Ac-CoA produced a 50-fold increase in k_{cat}/K_M (**Table 14B**).[305] Additionally, in contrast to the literature, activated ICL2-His$_6$FLAG did not exceed the catalytic efficiency recorded for His$_6$-ICL1.[305]

The lower than expected activation by Ac-CoA was most likely the result of differences in the protein purification and assay conditions. We did not evaluate whether ICL2-His$_6$FLAG co-purified with Ac-CoA. We felt it important to explore one difference in particular: the position of the affinity tag. While the literature study used His$_6$-ICL2, our initial work was performed before publication of the crystal structure revealed the

importance of the *C*-terminal domain, and we selected a *C*-terminal affinity tag in accordance with the literature that existed at the time.[305,317] Given that activation involves significant movement of the *C*-terminal domain, it is possible that the *C*-terminal His_6FLAG tag could interfere with the new interactions between ICL2 monomers that occur following Ac-CoA binding, affecting the conformational state of the tetramer and reducing the extent of activation. To test whether this was the case, we cleaved the affinity tag.

4.3.3.2. Cleavage of the *C*-terminal His_6FLAG tag does not improve ICL2 activity

To explore the effect of the *C*-terminal His_6FLAG tag on ICL2 activation by Ac-CoA, ICL2-His_6FLAG was purified as before (**Section 4.3.2.3**, **Figure 47A**) and the His_6FLAG tag was cleaved by overnight incubation with TEV protease.

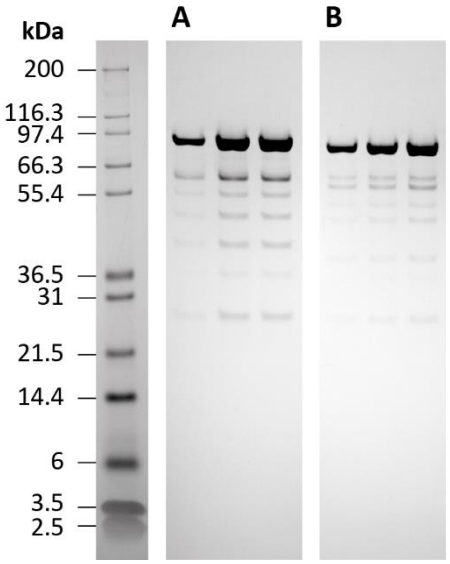


Figure 47: TEV cleavage of ICL2-His_6FLAG.
A – Purification of ICL2-His_6FLAG. ICL2-His_6FLAG was purified using **IMAC Method 3**, as described in **Figure 36**. The fractions shown are from the ICL2-His_6FLAG elution peak. **B – Recovery of tag-cleaved ICL2 by IMAC.** After overnight cleavage with TEV protease, ICL2 was recovered by IMAC in 82% yield. The fractions shown are from the ICL2 elution peak. Panels separated by white space were taken from separate images or from different parts of the same image.

ICL2 was recovered by IMAC in 82% yield (**Figure 47B**) and characterised in the presence and absence of 25 µM Ac-CoA using the assay developed for ICL2-His$_6$FLAG above (**Section 4.3.3.1**) (**Figure 48** and **Table 15**).

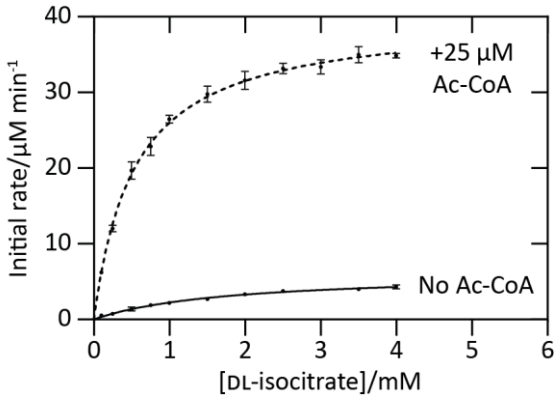


Figure 48: Steady-state kinetics of tag-cleaved ICL2 in the presence and absence of 25 µM Ac-CoA. The optimised assay was performed in the presence (dashed line) and absence (solid line) of 25 µM Ac-CoA. Each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—4 mM. Reactions were performed in triplicate and were initiated with 200 nM ICL2-His$_6$FLAG. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation.

Table 15: Kinetic parameters for tag-cleaved ICL2 in the presence and absence of 25 µM Ac-CoA. These parameters were obtained from the Michaelis-Menten fits shown in **Figure 48**. Errors are reported as symmetrical 95% confidence intervals.

	v_{max} (µM min^{-1})	K_M(DL-isocitrate) (µM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
No Ac-CoA	6.2 ± 0.5	1800 ± 300	0.51 ± 0.04	$2.9\times10^2 \pm 5\times10^1$
+25 µM Ac-CoA	40 ± 0.9	540 ± 45	3.3 ± 0.07	$6.2\times10^3 \pm 5\times10^2$

In the absence of Ac-CoA, the tag-cleaved ICL2 was ~4.5 times less active than ICL2-His$_6$FLAG (comparing the k_{cat}/K_M values). This activity loss was likely the consequence of the extended purification time required to achieve tag cleavage. Addition of 25 µM Ac-CoA rescued ICL2 activity, with the enzyme reaching a catalytic efficiency comparable to that of ICL2-His$_6$FLAG. Under our assay conditions, the kinetic

parameters for ICL2 and ICL2-His_6FLAG in the presence of 25 µM Ac-CoA are essentially the same, suggesting that there is no benefit to cleaving the *C*-terminal His_6FLAG tag prior to determining kinetic parameters for mycobacterial ICL2s expressed bearing this tag. Of note, the use of an *N*-terminal tag would eliminate any possibility of the extent of ICL2 activation being limited by the residues remaining after tag cleavage. Indeed, had the role of the *C*-terminal domain been known at the start of our work, we would likely have selected an *N*-terminal tag for ICL2 expression (as noted above, decisions were made in accordance with the literature available at the time).

4.3.4. Removal of the *C*-terminal domain renders ICL2 kinetically insensitive to Ac-CoA

Since activation of ICL2 by Ac-CoA occurs through binding to the *C*-terminal domain of the enzyme, ICL2 lacking the *C*-terminal domain should be kinetically insensitive to added Ac-CoA. To verify this, a truncated form of *aceA* (encoding amino acids 1—605 of ICL2) was amplified from Mtb CDC 1551 genomic DNA by PCR and assembled into the pNIC-CTHF vector. The resulting plasmid was used to transform *E. coli* BL21(DE3)Star and protein expression was performed using the same protocol as for His_6-ICL1 and ICL2-His_6FLAG. The truncated enzyme, ICL2(T605)-His_6FLAG, expressed well and was soluble (**Figure 49A**). ICL2(T605)-His_6FLAG lacks the *C*-terminal domain and should therefore be kinetically insensitive to added Ac-CoA. To confirm this, ICL2(T605)-His_6FLAG was purified using **IMAC Method 3** (**Figure 49B**). After buffer exchange by gravity SEC, the yield was determined to be 13.2 mg ICL2(T605)-His_6FLAG/g wet cells.

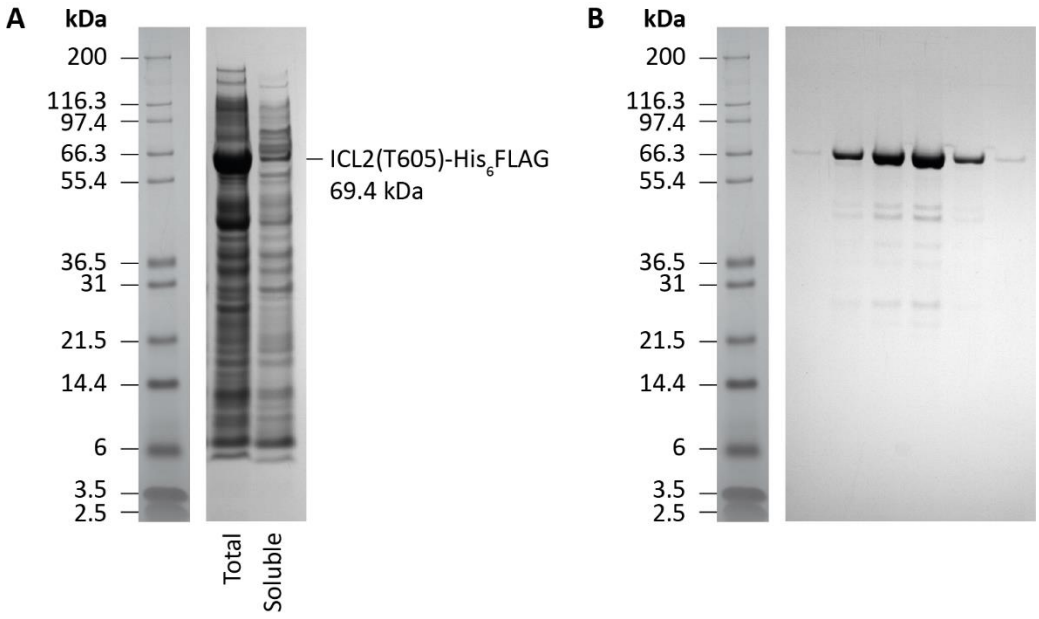


Figure 49: Expression and purification of ICL2(T605)-His$_6$FLAG.
A – Expression of ICL2(T605)-His$_6$FLAG. Recombinant ICL2(T605)-His$_6$FLAG was expressed in *E. coli* BL21(DE3)Star as described in **Chapter 8**. LDS-PAGE analysis showed that a protein of the expected molecular weight of ICL2(T605)-His$_6$FLAG (69.4 kDa) was present in both the complete cell lysate (**Total**) and in the soluble fraction (**Soluble**). **B – Purification of ICL2(T605)-His$_6$FLAG.** ICL2(T605)-His$_6$FLAG was purified using **IMAC Method 3**, as described in **Figure 36**. All fractions shown are from the ICL2(T605)-His$_6$FLAG elution peak. Panels separated by white space were taken from separate images.

Next, the enzyme was characterised in the presence and absence of 25 µM Ac-CoA using the assays developed for ICL2-His$_6$FLAG above (**Section 4.3.3.1**). As expected, ICL2(T605)-His$_6$FLAG was kinetically insensitive to 25 µM Ac-CoA (**Figure 50** and **Table 16**), corroborating the crystallographic conclusion that activation of ICL2 by Ac-CoA occurs through binding to the *C*-terminal domain.[305]

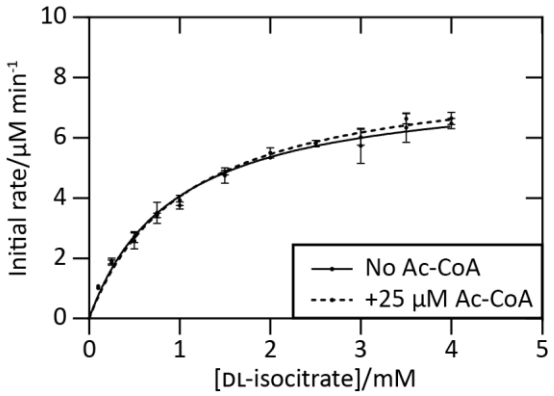


Figure 50: Steady-state kinetics of ICL2(T605)-His$_6$FLAG.
The optimised assay was performed in the presence (dashed line) and absence (solid line) of 25 μM Ac-CoA. Each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 μM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—4 mM. Reactions were performed in triplicate and were initiated with 200 nM ICL2-His$_6$FLAG. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation.

Table 16: Kinetic parameters for ICL2(T605)-His$_6$FLAG.
These parameters were obtained from the Michaelis-Menten fits shown in **Figure 50**. Errors are reported as symmetrical 95% confidence intervals.

	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
No Ac-CoA	7.9 ± 0.4	940 ± 140	0.66 ± 0.03	$7.0\times10^2 \pm 1\times10^2$
+25 μM Ac-CoA	8.4 ± 0.4	1100 ± 150	0.70 ± 0.03	$6.5\times10^2 \pm 1\times10^2$

4.4. Isocitrate dehydrogenase 2

4.4.1. Expression of IDH2-His$_6$FLAG

A construct comprising *icd2* in the pNIC-CTHF vector was prepared by Raphael Reinbold (University of Oxford). After sequence verification, this construct was used to transform *E. coli* BL21(DE3)Star and protein expression was performed using a protocol adapted from the literature.[363] As expected, recombinant IDH2-His$_6$FLAG expressed well and was soluble (**Figure 51**).

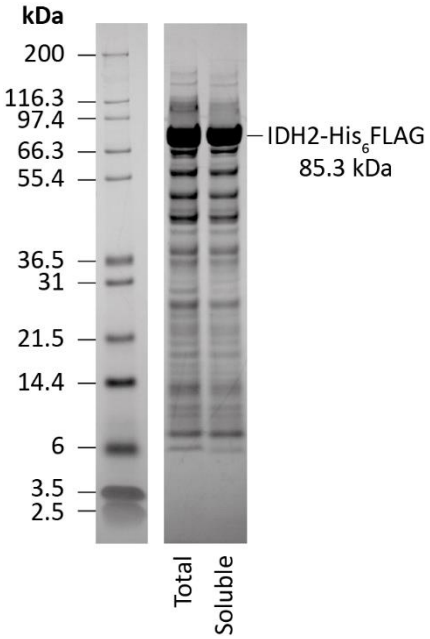


Figure 51: Expression of IDH2-His$_6$FLAG.
Recombinant IDH2-His$_6$FLAG was expressed in *E. coli* BL21(DE3)Star as described in **Chapter 8**. LDS-PAGE analysis showed that a protein of the expected molecular weight of IDH2-His$_6$FLAG (85.3 kDa) was present in both the complete cell lysate (**Total**) and in the soluble fraction (**Soluble**). Panels separated by white space were taken from separate images.

4.4.2. Purification of IDH2-His$_6$FLAG

Purification of IDH2-His$_6$FLAG was first attempted using **IMAC Method 1**. As noted for ICL2-His$_6$FLAG (**Section 4.3.2.3**), binding of the enzyme to FF medium was incomplete, and a substantial amount of IDH2-His$_6$FLAG was present in the fractions collected during sample application (**Figure 52A**). Consequently, we next attempted purification using **IMAC Method 3** and HP medium, since this strategy proved successful for ICL2-His$_6$FLAG. Indeed, changing to HP medium significantly reduced the amount of IDH2-His$_6$FLAG seen in the fractions collected during sample application (**Figure 52B**). As the enzyme purity was considered adequate for steady-state kinetics, **IMAC Method 3** with HP medium was selected as the preferred method for the purification of IDH2-His$_6$FLAG, and method development was not continued further. IDH2-His$_6$FLAG was purified multiple times over the course of assay implementation and enzyme

characterisation, and **IMAC Method 3** proved to be reproducible in terms of enzyme purity, as assessed by LDS-PAGE (n = 5).

After purification using **IMAC Method 3**, IDH2-His_6FLAG was buffer-exchanged in preparation for steady-state kinetics. Dialysis was used for the first experiment, but was subsequently replaced by gravity SEC. As for the ICLs, we noted that gravity SEC yielded higher quality protein than dialysis, as measured by the A_{280}/A_{250} ratio – the mean A_{280}/A_{250} reading was 3.1 for gravity SEC (n = 4) compared to 2.1 for dialysis (n = 1). Enzyme yields were determined after buffer exchange and varied in the range 10.9—21.9 mg IDH2-His_6FLAG/g wet cells (n = 5).

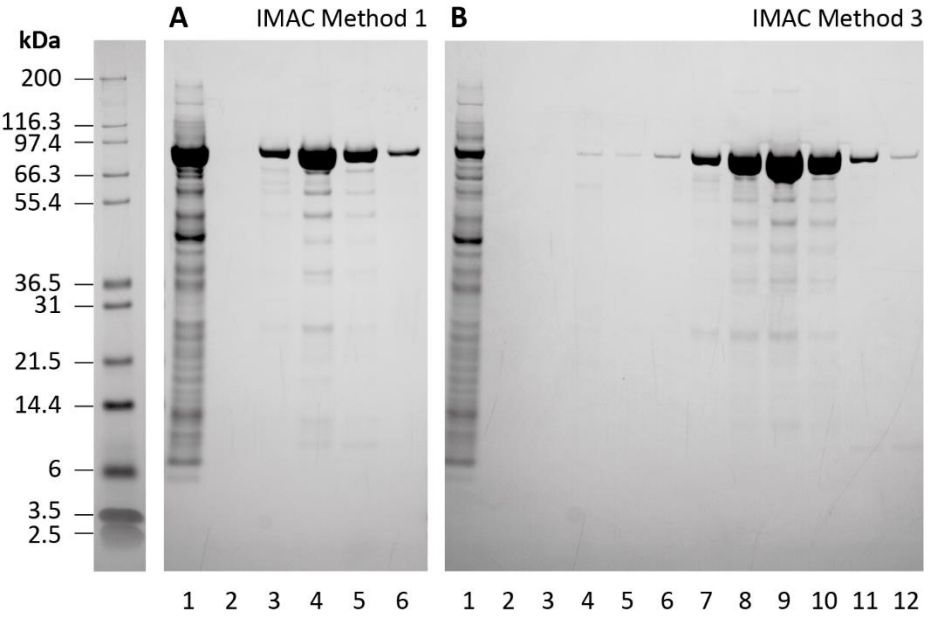


Figure 52: Purification of IDH2-His_6FLAG.
A – Purification of IDH2-His_6FLAG using IMAC Method 1. IDH2-His_6FLAG was purified using **IMAC Method 1**, as described in **Figure 26**. Lane 1, sample application; lane 2, wash step; lanes 3—6, IDH2-His_6FLAG elution peak. **B – Purification of IDH2-His_6FLAG using IMAC Method 3.** IDH2-His_6FLAG was purified using **IMAC Method 3**, as described in **Figure 36**. Lane 1, sample application; lanes 2—3, wash steps; lanes 4—12, IDH2-His_6FLAG elution peak. Panels separated by white space were taken from separate images.

4.4.3. Determination of kinetic parameters for IDH2-His_6FLAG

4.4.3.1. Method development

As discussed earlier (**Chapter 1**), both Mtb IDHs are $NADP^+$-dependent enzymes. Their activity can therefore be determined spectrophotometrically by measuring the increase in absorbance at 340 nm resulting from the reduction of $NADP^+$ (**Figure 53**).

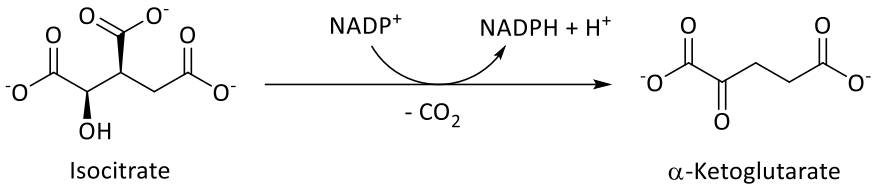


Figure 53: IDH assay reaction.
IDHs decarboxylate isocitrate to α-ketoglutarate. Concomitant reduction of the co-substrate, $NADP^+$, results in an increase in the absorbance at 340 nm, facilitating the spectrophotometric determination of IDH activity.

First, saturating concentrations were determined for each of the two co-substrates, isocitrate and $NADP^+$. In both cases, 2 mM was identified as a suitable saturating concentration for use in subsequent assays. We next attempted to determine kinetic parameters with respect to isocitrate. Briefly, each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). The isocitrate concentration was varied in the range 20—800 µM and reactions were initiated with 10 nM IDH2-His_6FLAG. However, the Michaelis-Menten fit of the initial rates (**Figure 54A**) was poor, with the replicates test for lack of fit identifying statistically significant deviation from the model ($p < 0.0001$). Examination of the residual plot (**Figure 54B**) revealed a zig-zag pattern, analogous to that seen for ICL2-His_6FLAG (**Section 4.3.3.1**). The residuals for the first six isocitrate concentrations exhibit an upward trend. After a sharp jump downwards, the upward trend repeats for the following six samples. As before, reactions within a replicate were run in order of increasing isocitrate concentration, suggesting that the

deviation from the Michaelis-Menten model was again the product of inconsistent thermal equilibration within replicates.

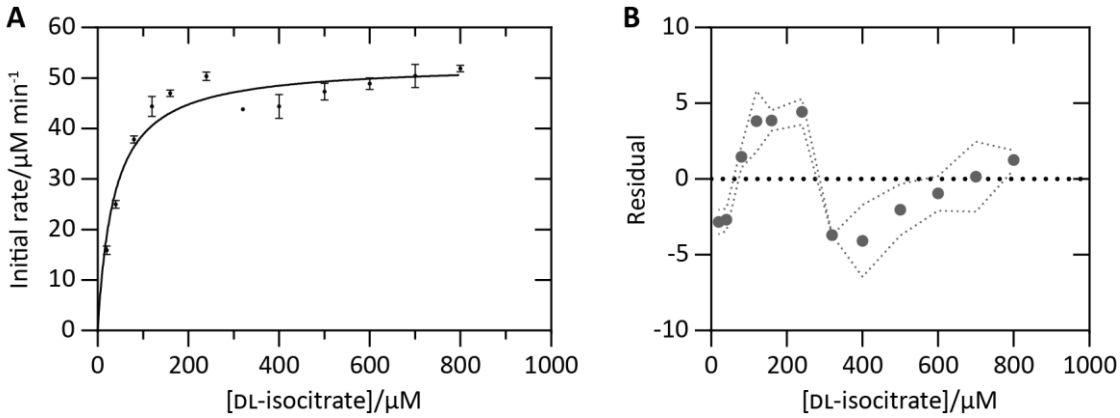


Figure 54: IDH2-His$_6$FLAG assay development: initial attempt.
A – Steady-state kinetics of IDH2-His$_6$FLAG, varying the concentration of isocitrate in the range 20—800 µM. Reactions were performed in triplicate and were initiated with 10 nM IDH2-His$_6$FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. The replicates test for lack of fit identified statistically significant deviation from the model ($p < 0.0001$). **B – The residual plot identified a systematic deviation from the Michaelis-Menten fit.** This panel shows the residual plot for the curve in panel **A**. The mean (data points) and standard deviation (dotted lines) of the residuals were plotted, and exhibit a zig-zag pattern analogous to that seen for ICL2-His$_6$FLAG.

This issue was addressed using the same approach as for ICL2-His$_6$FLAG, namely the randomisation of samples within replicates (**Figure 55**). Additionally, the minimum equilibration time was extended from 5 minutes to 7 minutes, and the enzyme concentration was lowered to 5 nM to improve the capture of initial rates at low isocitrate concentrations. (False sigmoidal kinetics were sometimes observed at higher concentrations of IDH2-His$_6$FLAG due to reactions at low isocitrate concentrations reaching completion in the time between initiation and the commencement of absorbance readings.) Finally, additional reactions at low isocitrate concentrations were included to improve the number of data points around the expected value of K_M(DL-isocitrate). The Michaelis-Menten fit of the resulting initial rates was good, as

assessed by the replicates test for lack of fit (p = 0.86), indicating that the above alterations to the assay protocol had successfully resolved the issue.

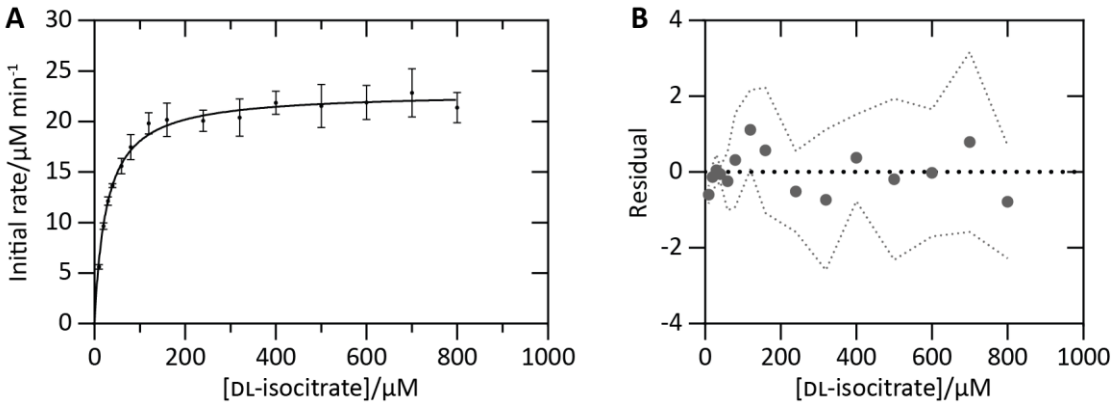


Figure 55: IDH2-His$_6$FLAG assay development: the effect of sample randomisation.
A – Steady-state kinetics of IDH2-His$_6$FLAG, varying the concentration of isocitrate in the range 20—800 µM. Reactions were performed in triplicate and were initiated with 5 nM IDH2-His$_6$FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. **B – Residual plot for the curve in panel A.** The mean (data points) and standard deviation (dotted lines) of the residuals were plotted.

4.4.3.2. Determination of kinetic parameters with respect to isocitrate and $NADP^+$

Next, kinetic parameters were determined with respect to each of the two co-substrates, isocitrate and $NADP^+$. Full experimental procedures are provided in **Chapter 8**. Briefly, to determine kinetic parameters with respect to isocitrate, each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). The isocitrate concentration was varied in the range 10—800 µM. To determine kinetic parameters with respect to $NADP^+$, each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and isocitrate (2 mM). The $NADP^+$ concentration was varied in the range 10 µM—2 mM. In both cases, reactions were initiated with 5 nM IDH2-His$_6$FLAG and samples were randomised within each replicate. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation (**Figure 56** and **Table 17**). While the literature data for Mtb IDH2 are limited, our

K_M(DL-isocitrate) (27 ± 3.9 µM) is comparable to the literature value (20 ± 1 µM), while our K_M(NADP$^+$) (42 ± 6.8 µM) is larger than the literature value (19.6 ± 6 µM).[379] However, as discussed for Mtb ICL1 above (**Section 4.2.2**), differences in protein preparation and assay conditions limit the utility of such comparisons.

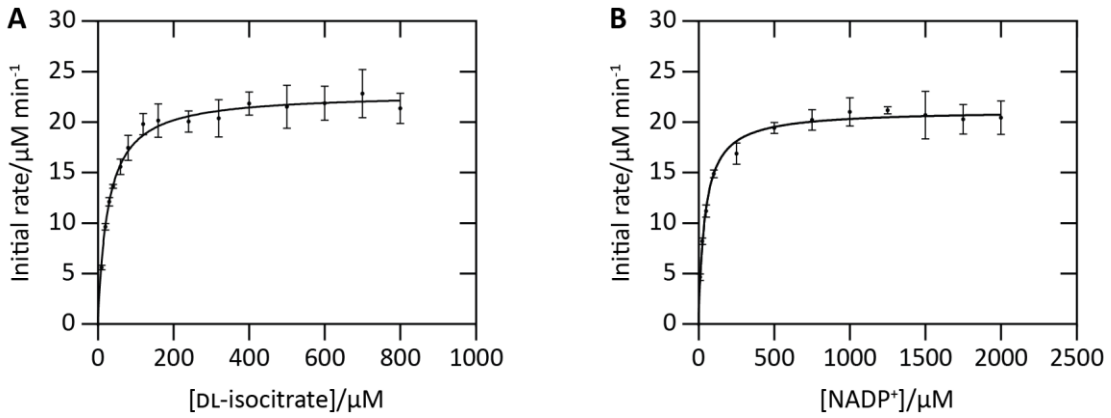


Figure 56: Steady-state kinetics of IDH2-His$_6$FLAG.
A – Varying the concentration of isocitrate in the range 20—800 µM. Reactions were performed in triplicate and were initiated with 5 nM IDH2-His$_6$FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and NADP$^+$ (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. **B – Varying the concentration of NADP$^+$ in the range 10 µM—2 mM.** Reactions were performed in triplicate and were initiated with 5 nM IDH2-His$_6$FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and isocitrate (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation.

Table 17: Kinetic parameters for IDH2-His$_6$FLAG.
These parameters were obtained from the Michaelis-Menten fits shown in **Figure 56**. Errors are reported as symmetrical 95% confidence intervals.

Substrate	v_{max} (µM min^{-1})	K_M (µM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
DL-isocitrate	23 ± 0.7	27 ± 3.9	76 ± 2	$2.8\times10^6 \pm 4\times10^5$
NADP$^+$	21 ± 0.6	42 ± 6.8	71 ± 2	$1.7\times10^6 \pm 3\times10^5$

4.5. Isocitrate dehydrogenase 1

4.5.1. Expression of IDH1-His$_6$FLAG

A construct comprising *icd1* in the pNIC-CTHF vector was prepared by Raphael Reinbold (University of Oxford). The gene was codon-optimised for translation in *E. coli*. After sequence verification, this construct was used to transform *E. coli* BL21(DE3)Star and protein expression was performed using the same protocol as for Mtb IDH2-His$_6$FLAG above (**Section 4.4.1**). As expected from the literature, expression of recombinant IDH1-His$_6$FLAG was lower than expression of recombinant IDH2-His$_6$FLAG.[363] Nonetheless, soluble protein was obtained (**Figure 57**).

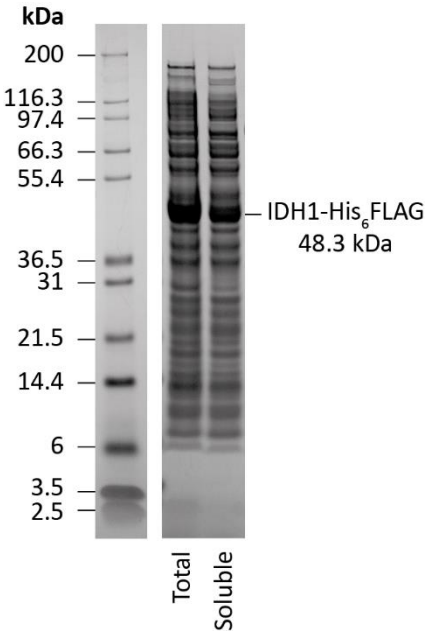


Figure 57: Expression of codon-optimised IDH1-His$_6$FLAG.
Recombinant IDH1-His$_6$FLAG was expressed in *E. coli* BL21(DE3)Star as described in **Chapter 8**. LDS-PAGE analysis showed that a protein of the expected molecular weight of IDH1-His$_6$FLAG (48.3 kDa) was present in both the complete cell lysate (**Total**) and in the soluble fraction (**Soluble**). Panels separated by white space were taken from separate images.

4.5.2. Purification of IDH1-His$_6$FLAG

4.5.2.1. Initial attempt at purification

Purification of IDH1-His$_6$FLAG was first attempted using **IMAC Method 3**, as this method proved to be the most effective for the purification of His$_6$-ICL1, ICL2-His$_6$FLAG and

IDH2-His_6FLAG. IDH1-His_6FLAG was successfully enriched (**Figure 58A**), but in contrast to the other enzymes discussed in this chapter, a shouldered peak was observed during A_{280} monitoring of the elution phase. LDS-PAGE analysis (**Figure 58A**) confirmed that the concentrations of the co-purifying proteins did not correlate with the concentration of IDH1-His_6FLAG, indicating that IMAC method optimisation would likely be a useful strategy for improving the purity of the enzyme. Fractions containing IDH1-His_6FLAG were pooled and desalted, then subjected to AIEX using the method previously developed for His_6-ICL1 (**Section 4.2.1.5**). LDS-PAGE analysis of the elution phase (**Figure 58B**) showed that AIEX was partially successful in separating the sample components.

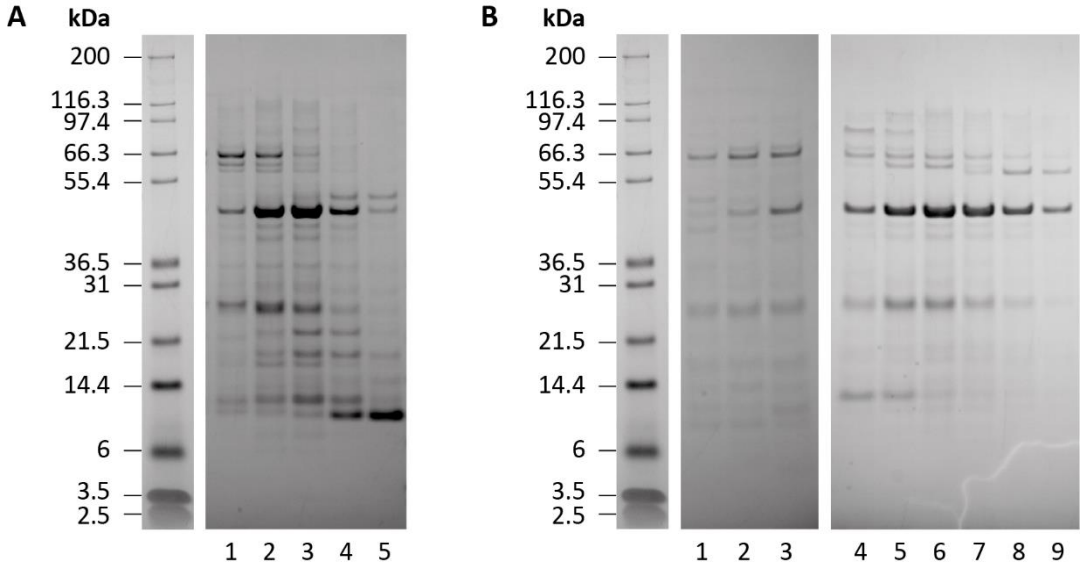


Figure 58: Purification of IDH1-His_6FLAG: IMAC Method 3 followed by AIEX.
A – Purification of IDH1-His_6FLAG using IMAC Method 3. IDH1-His_6FLAG capture was performed using **IMAC Method 3**, as described in **Figure 36**. Lanes 1—5 show the IDH1-His_6FLAG elution peak. **B – Further purification of IDH1-His_6FLAG using AIEX.** IMAC fractions containing IDH1-His_6FLAG were pooled and desalted, then applied to a HiTrap Q HP 1 mL column equilibrated with 20 mM Tris (pH 8). The column was washed with 20 mM Tris (pH 8, 10 CV) prior to gradient elution (0—50% elution buffer over 20 CV; elution buffer, 20 mM Tris, 1 M NaCl, pH 8). Lanes 1—9 show the IDH1-His_6FLAG elution peak. Panels separated by white space were taken from separate images.

Overall, the initial experiment suggested that a multi-step purification workflow – IMAC followed by desalting and AIEX – would likely be effective for the purification of IDH1-His_6FLAG.

4.5.2.2. Optimisation of the IMAC method

Having established that some of the contaminant proteins could likely be separated from IDH1-His_6FLAG by IMAC, we proceeded to IMAC method optimisation. The two wash steps in **IMAC Method 3** (100% buffer A and 5% buffer B) were extended to 15 CV in duration, and a third wash step was added (15% buffer B, 5 CV). Additionally, elution was split into two phases, with the first elution gradient running from 15—70% buffer B over 20 CV and the second elution gradient running from 70—100% buffer B over 5 CV. Finally, the fraction volume was reduced to 0.5 mL (from 1 mL) during the first elution gradient. The resulting method, **IMAC Method 5** (**Figure 59**), was then used to purify IDH1-His_6FLAG (**Figure 60A**). Those fractions containing the enzyme and contaminant proteins which the initial experiment indicated could be removed by AIEX were pooled, desalted and further purified by AIEX. No changes were made to the AIEX method.

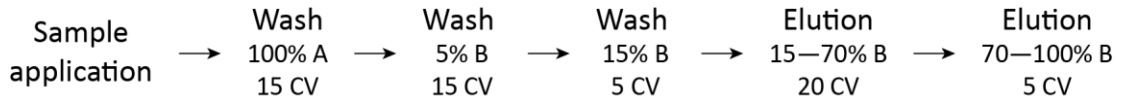


Figure 59: IMAC Method 5.
The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (15 CV), 5% buffer B (15 CV) and 15% buffer B (5 CV). Gradient elution was split into two steps: 15—70% B over 20 CV, followed by 70—100% B over 5 CV. Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

The revised first elution phase of **IMAC Method 5** (a shallower elution gradient combined with a smaller fraction volume) facilitated separation of the contaminant proteins from IDH1-His_6FLAG such that the enzyme was almost homogeneous after AIEX

(**Figure 60B**). However, the third wash step (15% buffer B, 5 CV) prompted elution of IDH1-His_6FLAG to begin earlier than intended, resulting in incomplete separation of the enzyme from the major contaminant protein at ~66 kDa. As fractions containing this contaminant were not pooled for AIEX, this would have reduced the yield of IDH1-His_6FLAG (enzyme yields were not determined at this stage).

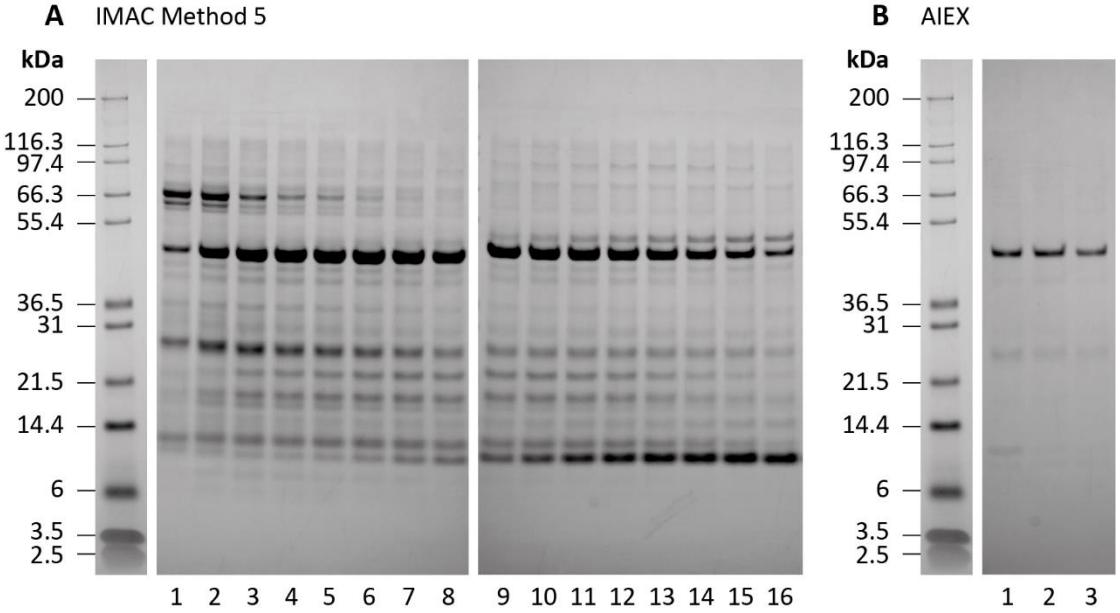


Figure 60: Purification of IDH1-His_6FLAG: IMAC Method 5 followed by AIEX.
A – Purification of IDH1-His_6FLAG using IMAC Method 5. IDH1-His_6FLAG capture was performed using **IMAC Method 5**, as described in **Figure 59**. Lanes 1—16 show the IDH1-His_6FLAG elution peak. **B – Further purification of IDH1-His_6FLAG using AIEX.** IMAC fractions containing IDH1-His_6FLAG were pooled and desalted, then applied to a HiTrap Q HP 1 mL column equilibrated with 20 mM Tris (pH 8). The column was washed with 20 mM Tris (pH 8, 10 CV) prior to gradient elution (0—50% elution buffer over 20 CV; elution buffer, 20 mM Tris, 1 M NaCl, pH 8). Lanes 1—3 show the IDH1-His_6FLAG elution peak. Panels separated by white space were taken from separate images.

Consequently, the next step in IMAC method optimisation was the removal of the third wash step to create **IMAC Method 6** (**Figure 61**). After purification with **IMAC Method 6**, fractions containing IDH1-His_6FLAG and contaminant proteins removable by AIEX were pooled, desalted, and further purified by AIEX. Removal of the third wash step did improve separation of IDH1-His_6FLAG from the 66 kDa contaminant protein (**Figure 62A**), and once again the enzyme was almost homogeneous after AIEX (**Figure 62B**). We

therefore considered the combination of **IMAC Method 6** and AIEX to be a suitable strategy for the purification of IDH1-His_6FLAG.

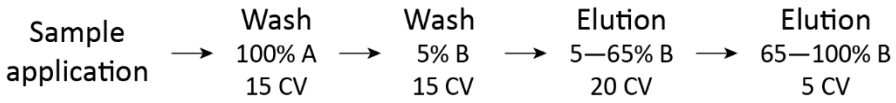


Figure 61: IMAC Method 6.
The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (15 CV), followed by 5% buffer B (15 CV). Gradient elution was split into two steps: 5—65% B over 20 CV, followed by 65—100% B over 5 CV. Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

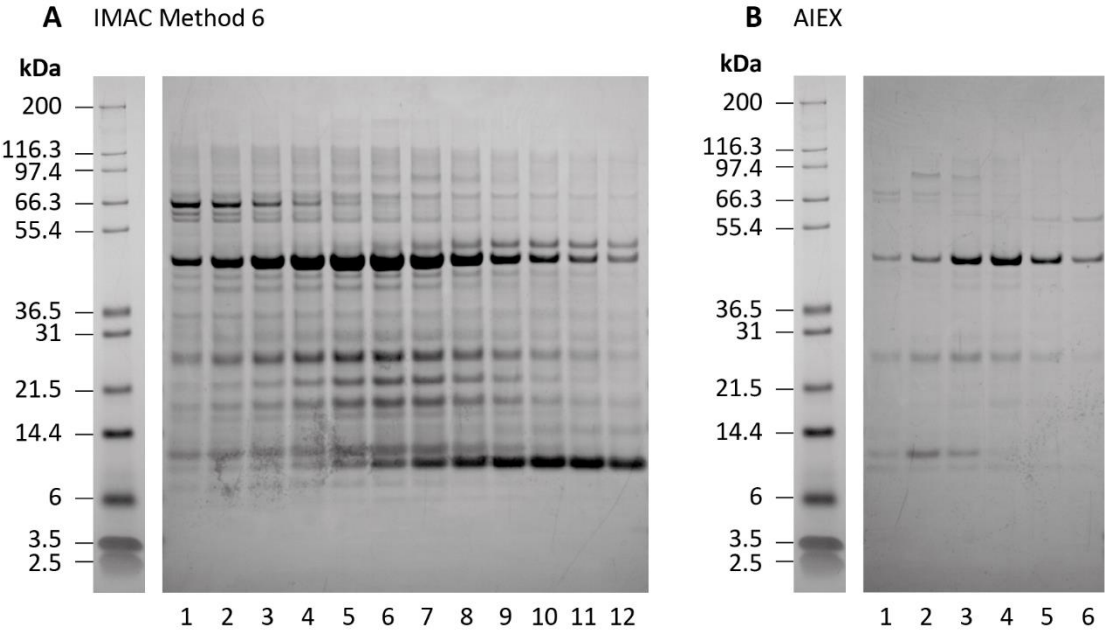


Figure 62: Purification of IDH1-His_6FLAG: IMAC Method 6 followed by AIEX.
A – Purification of IDH1-His_6FLAG using IMAC Method 5. IDH1-His_6FLAG capture was performed using **IMAC Method 6**, as described in **Figure 61**. Lanes 1—12 show the IDH1-His_6FLAG elution peak. **B – Further purification of IDH1-His_6FLAG using AIEX.** IMAC fractions containing IDH1-His_6FLAG were pooled and desalted, then applied to a HiTrap Q HP 1 mL column equilibrated with 20 mM Tris (pH 8). The column was washed with 20 mM Tris (pH 8, 10 CV) prior to gradient elution (0—50% elution buffer over 20 CV; elution buffer, 20 mM Tris, 1 M NaCl, pH 8). Lanes 1—6 show the IDH1-His_6FLAG elution peak. Protein purity was not quantified. Panels separated by white space were taken from separate images.

4.5.2.3. Stabilisation of purified IDH1-His_6FLAG

Having reached a satisfactory level of purity, AIEX fractions containing IDH1-His_6FLAG were pooled and buffer-exchanged by dialysis. During subsequent protein concentration measurements, we found that the protein quality was poor (A_{280}/A_{250} = 1), and that strategies to remove aggregates, such as filtration and centrifugation, were not beneficial. Instead, we sought to prevent deterioration at an earlier stage. Returning to the literature, we noted that His_6-IDH1 had previously been found to be unstable, but could be stabilised by addition of 5 mM $MnCl_2$ to the SEC buffer. The stabilising effect of $MnCl_2$ was attributed to Mn^{2+} binding at the IDH1 dimer interface.[363] Consequently, the buffer exchange method was changed to gravity SEC and 5 mM $MnCl_2$ was included in the SEC buffer. Additionally, all steps of the purification (IMAC, desalting, AIEX, and gravity SEC) were completed on the same day to remove the possibility of aggregation during standing at 4 °C overnight. Contrary to the literature report, inclusion of 5 mM $MnCl_2$ in the SEC buffer did not stabilise the enzyme.

IDH1-His_6FLAG was eventually stabilised (A_{280}/A_{250} = 3.5) by spiking the AIEX fractions containing the enzyme with $MnCl_2$ (final [$MnCl_2$] 5 mM) prior to gravity SEC. 5 mM $MnCl_2$ was also included in the SEC buffer. The optimised conditions yielded 0.08 mg IDH1-His_6FLAG/g wet cells (0.32 mg IDH1-His_6FLAG/L culture, n = 1), a result comparable to the recent literature (0.3 mg IDH1/L culture).[363] Having successfully purified and stabilised IDH1-His_6FLAG, we next proceeded to steady-state kinetics.

4.5.3. Determination of kinetic parameters for IDH1-His_6FLAG

Characterisation of IDH1-His_6FLAG was performed using the assays developed for IDH2-His_6FLAG (**Section 4.4.3**). Further method development was not required. As for

IDH2-His_6FLAG, kinetic parameters were determined with respect to each of the two co-substrates, isocitrate and $NADP^+$, by fitting initial rates to the Michaelis-Menten equation (**Figure 63** and **Table 18**). Comparing our calculated catalytic efficiencies to those calculated from the literature (**Table 19**), we find that both k_{cat}/K_M(DL-isocitrate) and $k_{cat}/K_M(NADP^+)$ are within error.[363]

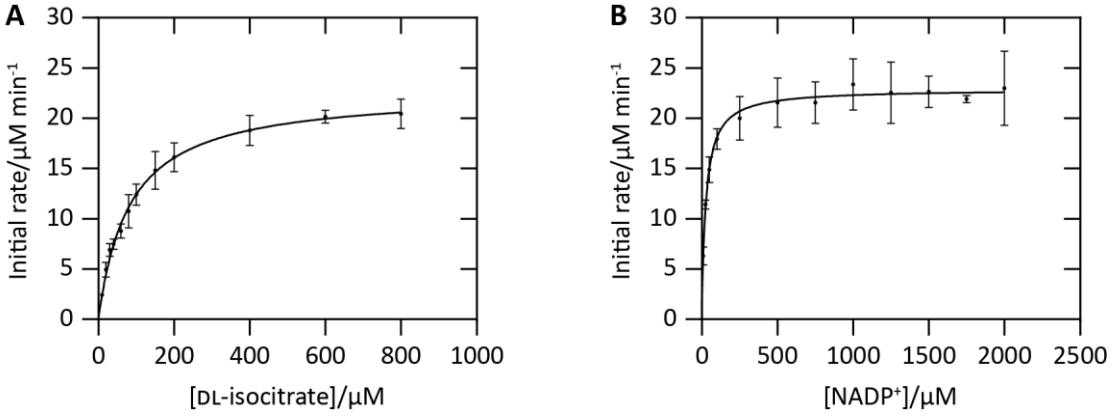


Figure 63: Steady-state kinetics of IDH1-His_6FLAG.
A – Varying the concentration of isocitrate in the range 20—800 μM. Reactions were performed in triplicate and were initiated with 5 nM IDH1-His_6FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. **B – Varying the concentration of $NADP^+$ in the range 10 μM—2 mM.** Reactions were performed in triplicate and were initiated with 5 nM IDH1-His_6FLAG. Each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and isocitrate (2 mM). Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation.

Table 18: Kinetic parameters for IDH1-His_6FLAG
These parameters were obtained from the Michaelis-Menten fits shown in **Figure 63**. Errors are reported as symmetrical 95% confidence intervals.

Substrate	v_{max} (μM min^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
DL-isocitrate	23 ± 1	83 ± 12	76 ± 4	$9.0\times10^5 \pm 1\times10^5$
$NADP^+$	23 ± 0.9	27 ± 6.7	76 ± 3	$2.9\times10^6 \pm 7\times10^5$

Table 19: Literature kinetic parameters for IDH1-His$_6$FLAG.
Values for K_M and k_{cat} were obtained from the literature and used to calculate k_{cat}/K_M.[363]

Substrate	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
DL-isocitrate	50 ± 7	33 ± 2	$7.0\times10^5 \pm 1.0\times10^5$
$NADP^+$	15 ± 3	33 ± 2	$2.2\times10^6 \pm 4.6\times10^5$

4.6. Kinetic comparison of the isocitrate branch point enzymes from Mtb CDC 1551

Having characterised each of the four isocitrate branch point enzymes from Mtb CDC 1551 individually, we are now in a position to compare their ability to metabolise isocitrate. For ease of comparison, the kinetic parameters with respect to DL-isocitrate obtained in the preceding sections are summarised in **Table 20** and plotted in **Figure 64**. Under saturating conditions, the IDHs turn over isocitrate three times faster than His$_6$-ICL1, which is in turn 5—15 times faster than ICL2-His$_6$FLAG, depending on the activation state of the latter (**Figure 64A**). The IDHs also have a much higher affinity for isocitrate than the ICLs (**Figure 64B**). For example, K_M(DL-isocitrate) is 15 times lower for IDH2-His$_6$FLAG than for His$_6$-ICL1. Consequently, the IDHs can be considered more catalytically efficient than the ICLs (**Figure 64C**).

Table 20: Kinetic parameters for the Mtb enzymes obtained in the preceding sections.
His$_6$-ICL1, **Figure 39** and **Table 13**; ICL2-His$_6$FLAG, **Figure 46** and **Table 14**; IDH2-His$_6$FLAG, **Figure 56** and **Table 17**; IDH1-His$_6$FLAG, **Figure 63** and **Table 18**. The kinetic parameters for ICL2-His$_6$FLAG obtained in the presence of 25 μM Ac-CoA are indicated by * next to the enzyme name. Errors are reported as symmetrical 95% confidence intervals.

	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
His$_6$-ICL1	40 ± 2	390 ± 59	22 ± 1	$5.7\times10^4 \pm 9\times10^3$
ICL2-His$_6$FLAG	17 ± 1	1100 ± 180	1.4 ± 0.1	$1.3\times10^3 \pm 2\times10^2$
ICL2-His$_6$FLAG*	51 ± 1	580 ± 48	4.3 ± 0.1	$7.4\times10^3 \pm 6\times10^2$
IDH2-His$_6$FLAG	23 ± 0.7	27 ± 3.9	76 ± 2	$2.8\times10^6 \pm 4\times10^5$
IDH1-His$_6$FLAG	23 ± 1	83 ± 12	76 ± 4	$9.0\times10^5 \pm 1\times10^5$

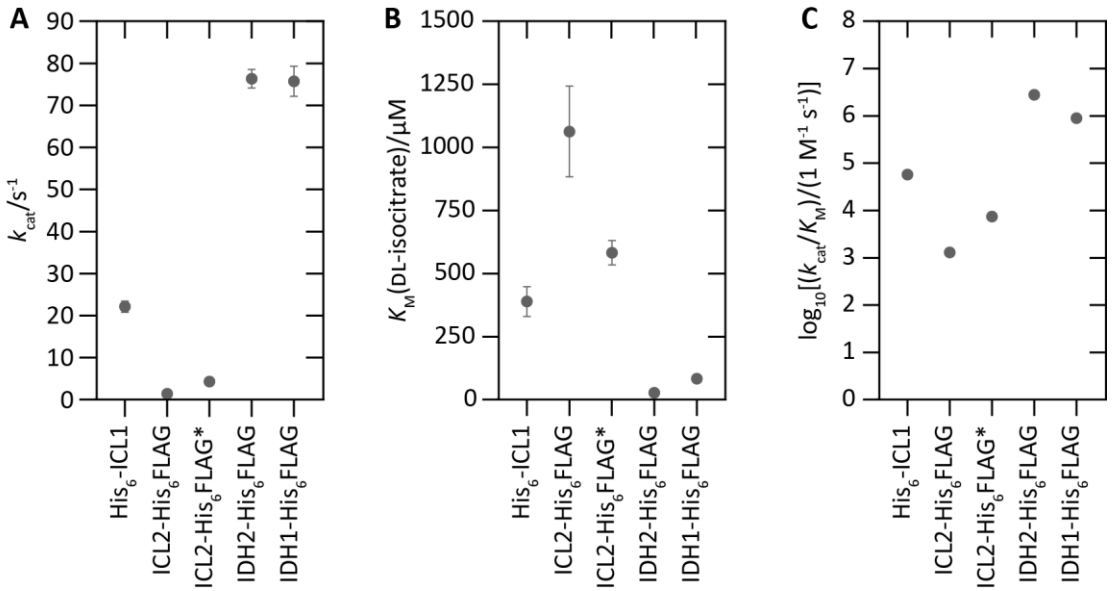


Figure 64: Comparison of the kinetic parameters obtained for the Mtb enzymes.
These plots show the kinetic parameters listed in **Table 20**. ICL2-His$_6$FLAG*, enzyme in the presence of 25 μM Ac-CoA. **A – k_{cat}.** Under saturating conditions, IDH2-His$_6$FLAG and IDH1-His$_6$FLAG turn over isocitrate three times faster than His$_6$-ICL1. His$_6$-ICL1 turns over isocitrate 15 times faster than ICL2-His$_6$FLAG in the absence of Ac-CoA and five times faster than Ac-CoA-activated ICL2-His$_6$FLAG. **B – K_M(DL-isocitrate).** The IDHs have a much higher affinity for isocitrate than the ICLs. For example, K_M(DL-isocitrate) is 15 times lower for IDH2-His$_6$FLAG than for His$_6$-ICL1. **C – k_{cat}/K_M.** Catalytic efficiencies were plotted on a $\log_{10}$ scale. The combination of higher k_{cat} values and lower K_M(DL-isocitrate) values mean that the IDHs can be considered more catalytically efficient than the ICLs.

Of course, these comparisons reflect saturating isocitrate concentrations *in vitro*. Flux partitioning between the four enzymes *in bacterio* would depend on the isocitrate pool size, which has yet to be quantified in Mtb.[j] Nonetheless, the kinetic parameters obtained in this study provide a useful guide to the theoretical metabolic capacity of the isocitrate branch point in Mtb CDC 1551. As we encountered in our discussion of His$_6$-ICL1 above (**Section 4.2.2**), enzyme preparation and assay conditions can have a substantial impact on the recorded values of kinetic parameters. Critically, the methodology developed in this chapter minimises the variability in experimental

[j] We planned to determine the concentration of isocitrate in Mtb (and other mycobacterial species) through quantitative metabolomics experiments using our Ion Mobility Q-TOF LC-MS instrument. The use of ion mobility spectrometry would allow the resolution and separate quantification of isocitrate and citrate (structural isomers which co-elute from the liquid chromatography columns we typically use for metabolomics).[561,562] Due to time constraints, we were unable to progress beyond preliminary experimental design.

conditions between enzymes as far as possible. To our knowledge, this is the first study in which all four isocitrate metabolising enzymes from Mtb have been characterised simultaneously. Our data therefore constitute the most directly comparable set of kinetic parameters for the complete Mtb isocitrate branch point recorded to date, and will allow more accurate modelling of flux through the isocitrate branch point under different enzyme and substrate concentrations.

4.7. Chapter conclusion

In this chapter, we expressed the four isocitrate metabolising enzymes from Mtb CDC 1551 in *E. coli* BL21(DE3)Star, developed methods for their purification and implemented assays to facilitate their kinetic characterisation. While protein expression proved straightforward, purification required more extensive method development, particularly in the case of His_6-ICL1. Ultimately, all four enzymes – His_6-ICL1, ICL2-His_6FLAG, IDH2-His_6FLAG and IDH1-His_6FLAG – were successfully purified in sufficient yields for steady-state kinetics. The final purification strategy for His_6-ICL1, ICL2-His_6FLAG and IDH2-His_6FLAG consisted of a single IMAC step (**IMAC Method 3**), while IDH1-His_6FLAG required an alternative IMAC method (**IMAC Method 6**) followed by desalting and AIEX. Purification was followed by buffer exchange, with gravity SEC consistently proving superior to dialysis for the maintenance of protein quality.

Assay implementation also proved relatively straightforward. ICLs were characterised by coupling glyoxylate production to LDH, while IDH activity was measured directly, providing a complete set of kinetic parameters for the four Mtb enzymes. Additionally, we confirmed that ICL2 is activated by Ac-CoA and, using a truncated form of the

enzyme, corroborated the crystallographic finding that this activation involves binding of Ac-CoA to the *C*-terminal domain of ICL2.[305]

Due to careful design of the protein purification and assay conditions, the kinetic parameters obtained for the Mtb enzymes are directly comparable to each other and – assuming successful method translation – will be directly comparable to the parameters we later obtain for the homologous enzymes from other mycobacterial species. Thus, by establishing methods for the expression, purification and characterisation of mycobacterial isocitrate metabolising enzymes in our laboratory, we have laid the foundations for the comparative exploration of isocitrate metabolism kinetics in other mycobacterial species. Such explorations are the focus of **Chapter 5**.

Chapter 5. A multispecies comparative enzymology study of the mycobacterial isocitrate branch point

5.1. Chapter introduction

Metabolic flexibility is a critical aspect of the Mtb response to the host niche, with control of flux through the isocitrate branch point being of particular importance. Since host and non-host niches exert different demands on the isocitrate branch point due to their distinct organic carbon content, changes in isocitrate metabolism may be associated with mycobacterial niche adaptation. Specifically, host-adapted mycobacterial species may have evolved a more efficient glyoxylate shunt as a result of the selection pressure exerted by the host niche (**Chapter 1**).

The structure of the isocitrate branch point varies between mycobacterial species, and may consist of two, three, or four enzymes. However, the presence or absence of individual enzymes does not correlate with mycobacterial ecology (**Chapter 2**). Our next question is whether the basic kinetic parameters of isocitrate metabolising enzymes vary between mycobacterial species. For example, do ICLs from host-adapted mycobacterial species metabolise isocitrate more efficiently than ICLs from soil mycobacteria? Are the kinetics of IDH2 different in species possessing and lacking IDH1? To facilitate the comparison of isocitrate metabolism kinetics at the genus level, we established methods for the expression, purification and characterisation of mycobacterial isocitrate metabolising enzymes in our laboratory, using the enzymes from Mtb CDC 1551 (**Chapter 4**).

A feasible comparative enzymology study requires capturing the diversity of the genus in a realistic number of species. Accordingly, we will select a subset of our Mycobacterial

Species Library (**Chapter 1, Section 1.4**) which provides phylogenetic coverage of the genus and includes different isocitrate branch point structures and ecological phenotypes. Next, we will use the methods developed in **Chapter 4** to clone, express, purify and kinetically characterise the isocitrate metabolising enzymes from the selected species. Finally, comparison of these enzymes to each other and to the Mtb CDC 1551 enzymes characterised in **Chapter 4** will reveal whether isocitrate metabolism kinetics vary across the genus *Mycobacterium*. If changes in isocitrate metabolism kinetics form part of mycobacterial niche adaptation, kinetic and ecological phenotypes will be associated. Otherwise, any variation observed will be correlated with the evolutionary history of the genus.

Detailed experimental procedures for the work described in this chapter are presented in **Chapter 8**. Throughout this chapter, the terms 'fast growers' and 'slow growers' will be used to refer to the major phylogenetic divisions in the genus *Mycobacterium*, rather than the absolute growth rate of any given species *in vitro* (as discussed in **Chapter 1, Section 1.2**).[478] Unless otherwise specified, 'Mtb' refers to Mtb CDC 1551. Biologically significant variation is defined as a $\log_{10}$ difference between enzymes. This significance threshold was selected for consistency with future comparative microbiology and -omics experiments.

5.2. Selection of mycobacterial species for comparison to Mtb

At the time of species selection, our Mycobacterial Species Library contained 32 species in addition to Mtb. **Figure 65** shows the isocitrate branch point composition of these species, as determined in **Chapter 2**. The Mycobacterial Species Library includes the three common branch point structures identified in **Chapter 2**: four enzymes (ICL1, ICL2,

IDH1 and IDH2); three enzymes (ICL1, ICL2 and IDH2); and two enzymes (ICL1 and IDH2). Within the library, clustering of the three branch point structures is predominantly phylogenetic (**Figure 65**). As a result, the library species can be split into five groups by considering their branch point composition and phylogenetic growth rate, namely: slow growers with 4 enzymes; slow growers with 3 enzymes; fast growers with 4 enzymes; fast growers with 3 enzymes; and fast growers with 2 enzymes.

To ensure that the conclusions of the study are based on genus-level patterns rather than species-specific phenotypes, the species set should include at least two members of each group. Consequently, some species must be included in our study because they are the only two species in the library that belong to a particular group. Specifically, *M. hiberniae* and *M. terrae* are the only two slow growers with 3 enzymes, and *M. fallax* and *M. salmoniphilum* are the only fast growers with two enzymes. By the same rule, we must exclude the fast growers with four enzymes from the study. Our library only includes two species from this group, *M. holsaticum* and *M. mageritense*, and the *M. holsaticum icd2* sequence deposited in GenBank is truncated, meaning that we will not be able to design primers for the amplification of this gene using the available sequencing data.

Thus far, we have selected four species to compare to Mtb: *M. hiberniae*, *M. terrae*, *M. fallax*, and *M. salmoniphilum*. We next need to choose species to represent the slow growers with 4 enzymes and the fast growers with 3 enzymes, while ensuring adequate ecological and phylogenetic coverage of the genus in the final species set. To avoid the number of enzymes becoming too large, the total number of species should be limited to around ten.

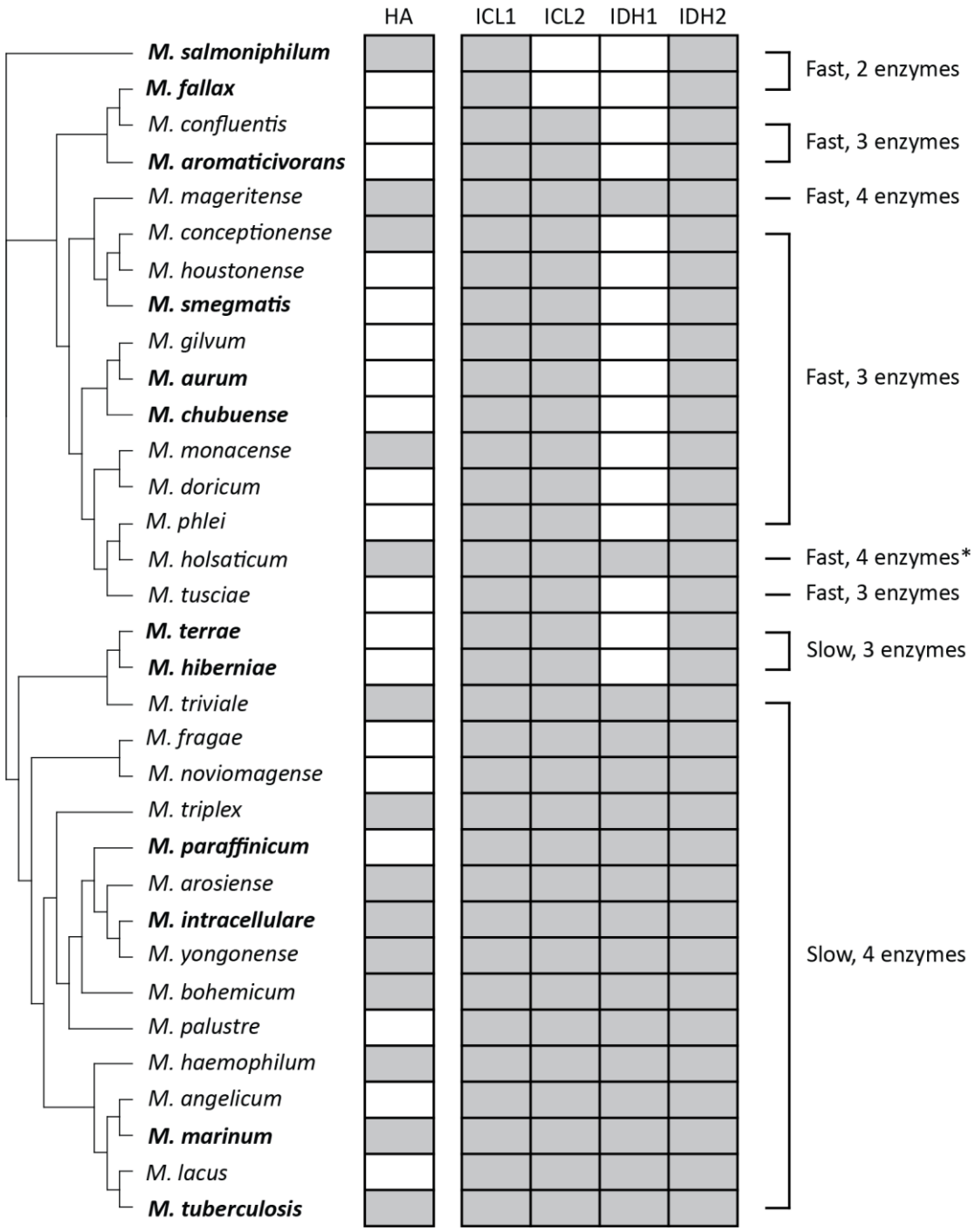


Figure 65: Analysis of host adaptation and isocitrate branch point composition in the 2018 Mycobacterial Species Library.

The library species can be split into five groups by considering their phylogenetic growth rate (fast or slow) and isocitrate branch point composition (2, 3, or 4 enzymes). Species selected for study are named in bold type. **HA – host adaptation.** A literature analysis of mycobacterial host adaptation was completed by Acely Garza-Garcia. Grey rectangles indicate species which display some degree of host adaptation. White rectangles indicate species considered not to be host-adapted. **ICL1, ICL2, IDH1, IDH2 – presence/absence of isocitrate metabolising enzymes.** Isocitrate branch point composition data are from **Chapter 2**. Grey rectangles indicate that the enzyme in question is present; white rectangles indicate that the enzyme in question is absent. * The *M. holsaticum icd2* sequence deposited in GenBank is truncated, so it is not clear whether this species possesses a full length IDH2. The phylogenetic tree used in this figure was modified from the 2022 Mycobacterial Species Library phylogenetic tree generated by Acely Garza-Garcia to show only those species present in the 2018 Mycobacterial Species Library.

M. marinum, *M. intracellulare*, and *M. paraffinicum* were selected to represent the slow growers with 4 enzymes. As discussed in **Chapter 1, Section 1.4.3**, *M. marinum* occupies both host (fish) and non-host (water) niches and is commonly used as a model organism for Mtb.[116,240] *M. intracellulare* is a member of the *M. avium* complex (MAC). Represented in our library by *M. arosiense*, *M. intracellulare*, and *M. yongonense*, the MAC is a group of 12 mycobacterial species which are collectively the most frequent cause of non-tuberculous mycobacterial infections in humans.[520] MAC species are also found in non-host niches such as soil and water.[116] *M. paraffinicum* was selected to add ecological diversity, being closely related to the MAC but less host-adapted than might be predicted from its phylogenetic position (**Figure 65**). This species is typically isolated from oil-contaminated soil and is only occasionally identified as the cause of human disease.[521,522]

M. smegmatis, *M. aromaticivorans*, *M. aurum*, and *M. chubuense* were selected to represent the fast growers with 3 enzymes. Isocitrate metabolism has been partly characterised in *M. smegmatis*, and this species is frequently compared to Mtb (**Chapter 1, Section 1.4.3**).[240,267] *M. aromaticivorans* is an environmental saprophytic species isolated from oil-contaminated soils, and is able to metabolise polyaromatic hydrocarbons.[470] *M. aurum* and *M. chubuense* were added to the species set to improve phylogenetic coverage of the fast growers. Both species were originally isolated from soil and can degrade various aliphatic pollutants.[523–526]

In total, eleven mycobacterial species (marked in bold in **Figure 65**) were selected for comparison to Mtb CDC 1551: *M. aromaticivorans*, *M. aurum*, *M. chubuense*, *M. fallax*, *M. hiberniae*, *M. intracellulare*, *M. marinum*, *M. paraffinicum*, *M. salmoniphilum*,

M. smegmatis, and *M. terrae*.[k] Accordingly, we set out to prepare and kinetically characterise the 34 enzymes (11 ICL1s, 9 ICL2s, 3 IDH1s, and 11 IDH2s) that compose the isocitrate branch points of these species using the methods developed for the Mtb enzymes in **Chapter 4**.

5.3. Preparation of mycobacterial isocitrate metabolising enzymes

5.3.1. Cloning

*Icl*s, *aceA*s, *icd2*s, and *icd1*s were amplified from the appropriate mycobacterial templates by PCR. The PCR conditions were modified as necessary to enable the successful amplification of each template, typically by changing the polymerase, annealing temperature, and/or extension time. Detailed PCR methods and the optimised conditions for the amplification of each gene are presented in **Chapter 8**. As for the Mtb enzymes, *icl*s were assembled into the pNIC-NTH vector (for expression of proteins bearing an *N*-terminal His_6 tag), while *aceA*s, *icd2*s, and *icd1*s were assembled into the pNIC-CTHF vector (for expression of proteins bearing a *C*-terminal His_6FLAG tag). In this way, we were able to prepare constructs encoding 32 of our 34 target genes – all except *icl* and *icd2* from *M. paraffinicum*.

Our inability to clone *M. paraffinicum icl* and *icd2* likely resulted from the fact that *M. paraffinicum* M11, the strain for which the whole-genome sequence is available, is

[k] As discussed above, the aim of the species selection process was to achieve phylogenetic coverage of the genus while ensuring adequate representation of the various isocitrate branch point structures, phylogenetic growth rates, and ecological profiles of interest. Consequently, structural variation and sequence identity were not considered at this stage. The mean pairwise protein sequence identity between the enzymes we selected was: ICL1s, 92% (83—98%); ICL2s, 82% (77—94%); IDH2s, 82% (78—92%); IDH1s, 89% (88—92%).

not the Mycobacterial Species Library strain (DSM-44181). Comparison of the *aceA* and *icd1* sequences from *M. paraffinicum* M11 and DSM-44181 revealed many SNPs between the two strains. Similar inter-strain variation in *icl* and *icd2* – specifically SNPs in the primer-binding regions of these genes – could prevent effective amplification of *icl* and *icd2* from DSM-44181 template DNA using primers designed for M11. Alternative routes to access *M. paraffinicum icl* and *icd2*, such as the use of synthetic genes from M11 or the sequencing of DSM-44181, were not pursued due to time constraints. For the same reason, we did not substitute another suitable species for *M. paraffinicum* following expansion of the Mycobacterial Species Library.

5.3.2. Expression

The constructs generated in **Section 5.3.1** were used to transform *E. coli* BL21(DE3)Star or BL21-Gold(DE3). BL21(DE3)Star was used in preference, as we had successfully used this strain to express the Mtb enzymes, with BL21-Gold(DE3) being used when BL21(DE3)Star was not available. Consistent with our experience using these two strains for the preparation of Mtb CDC 1551 His_6-ICL1 (**Chapter 4**), the choice of expression strain had no noticeable downstream impact. Protein expression was then performed using the protocols developed for the Mtb enzymes (**Chapter 4**). The recombinant His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs expressed well and were soluble. As for the Mtb proteins, expression of IDH1-His_6FLAGs was lower than expression of the other enzymes, but soluble protein was obtained in each case. Consequently, while 100 mL cultures yielded sufficient His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs for basic characterisation, 1 L cultures were required for IDH1-His_6FLAGs.

5.3.3. Purification

In **Chapter 4**, we developed purification strategies and chromatographic methods for each of the four isocitrate metabolising enzymes from Mtb CDC 1551 (**Figure 66**). To summarise, method development for His_6-ICL1, ICL2-His_6FLAG, and IDH2-His_6FLAG converged on a single strategy (**Figure 66A**): purification using **IMAC Method 3** (**Figure 66C**) followed by buffer exchange using gravity SEC. The purification and subsequent stabilisation of IDH1-His_6FLAG required further method development, resulting in the multi-step strategy shown in **Figure 66B**. We found that IDH1-His_6FLAG could be purified using **IMAC Method 6** (**Figure 66D**) followed by desalting and AIEX, but that it was necessary to stabilise the enzyme prior to buffer exchange using gravity SEC. Stabilisation was achieved through the addition of $MnCl_2$ to both the AIEX eluate and the gravity SEC buffer.

The methods developed for the Mtb enzymes were applied to the purification of the recombinant isocitrate metabolising enzymes from other mycobacterial species. His_6-ICL1s, ICL2-His_6FLAGs and IDH2-His_6FLAGs were purified using **IMAC Method 3** followed by buffer exchange using gravity SEC (**Figure 66A** and **Figure 66C**). This strategy proved highly successful, with all enzymes being obtained in sufficient yields and with appropriate purity for steady-state kinetics (His_6-ICL1s, **Figure 67**; ICL2-His_6FLAGs, **Figure 68**; IDH2-His_6FLAGs, **Figure 69**).

Enzyme yields were determined after buffer exchange and were comparable to those achieved for the Mtb enzymes (His_6-ICL1s, 0.69—12.70 mg/g wet cells; ICL2-His_6FLAGs, 4.39—26.57 mg/g wet cells; IDH2-His_6FLAGs, 6.03—20.64 mg/g wet cells. Individual

enzyme yields are given in the corresponding figure captions and in **Chapter 8** (**Table 48**, **Table 49**, and **Table 50**).

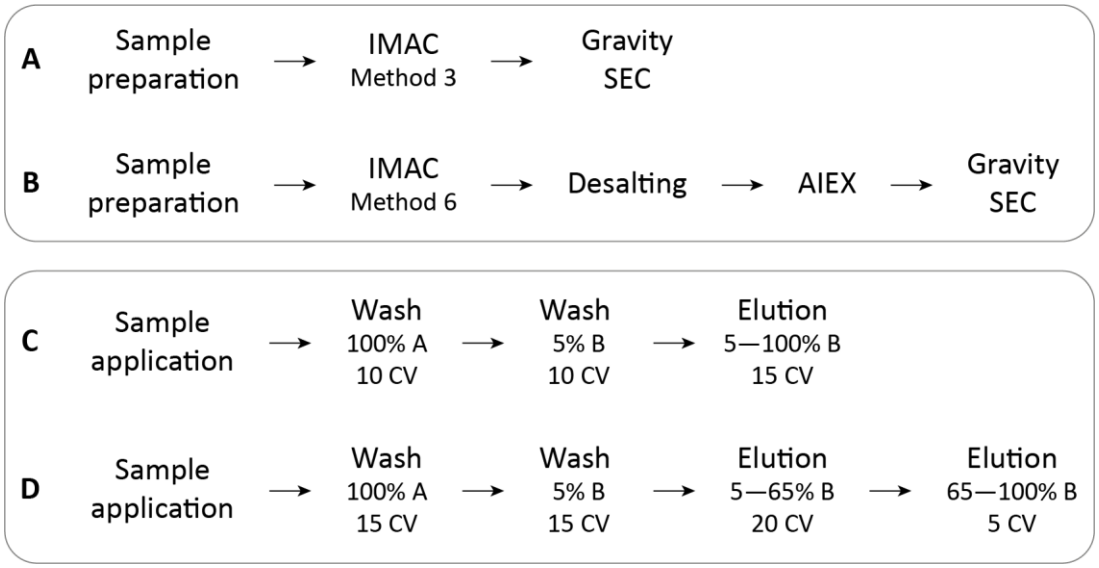


Figure 66: Strategies of the purification of mycobacterial isocitrate metabolising enzymes.
A – Purification of His_6-ICL1s, ICL2-His_6FLAGs and IDH2-His_6FLAGs. His_6-ICL1s, ICL2-His_6FLAGs and IDH2-His_6FLAGs were purified using **IMAC Method 3**, then buffer exchanged using gravity SEC. **B – Purification of IDH1-His_6FLAGs.** IDH1-His_6FLAGs were captured using **IMAC Method 6**. Fractions containing the enzyme were pooled and desalted, then further purified by AIEX. AIEX fractions containing the enzyme were spiked with 5 mM $MnCl_2$ prior to buffer exchange using gravity SEC. 5 mM $MnCl_2$ was included in the gravity SEC buffer. **C – IMAC Method 3.** The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (10 CV), followed by 5% buffer B (10 CV) prior to gradient elution (5—100% B over 15 CV). **D – IMAC Method 6.** The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (15 CV), followed by 5% buffer B (15 CV). Gradient elution was split into two steps: 5—65% B over 20 CV, followed by 65—100% B over 5 CV. **For IMAC Methods 3 & 6:** Buffer A (binding buffer): 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8; buffer B (elution buffer): 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8. CV, column volume.

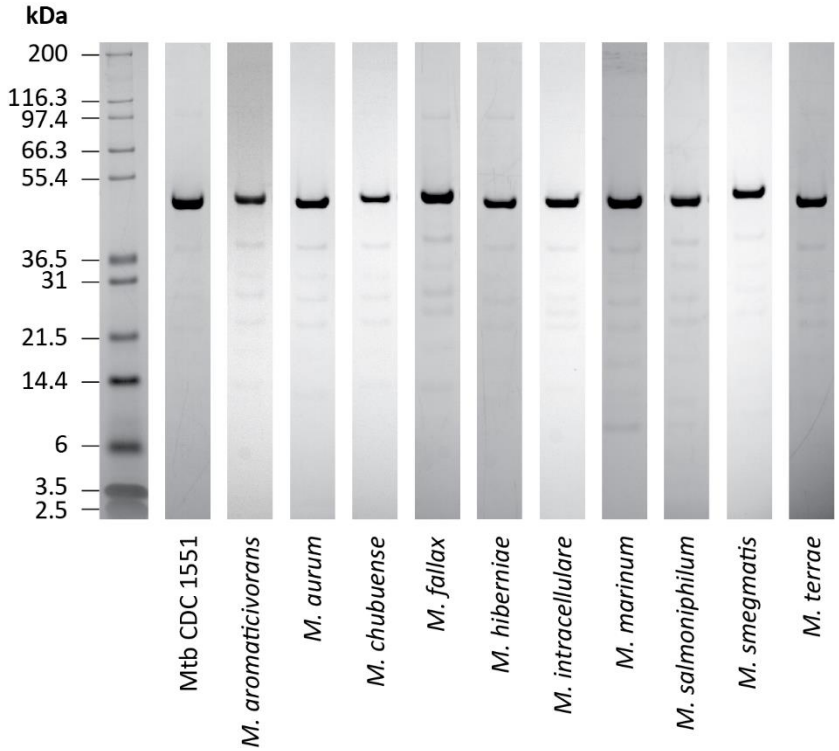


Figure 67: Purification of mycobacterial His_6-ICL1s.
His_6-ICL1s were purified using **IMAC Method 3** followed by buffer exchange using gravity SEC, as shown in **Figure 66**. His_6-ICL1 from Mtb CDC 1551, which was purified in **Chapter 4**, is shown for comparison. **Expected His_6-ICL1 sizes:** Mtb CDC 1551, 49.25 kDa; *M. aromaticivorans*, 49.28 kDa; *M. aurum*, 49.25 kDa; *M. chubuense*, 49.16 kDa; *M. fallax*, 49.48 kDa; *M. hiberniae*, 49.12 kDa; *M. intracellulare*, 49.16 kDa; *M. marinum*, 49.45 kDa; *M. salmoniphilum*, 48.86 kDa; *M. smegmatis*, 49.26 kDa; *M. terrae*, 48.98 kDa. **His_6-ICL1 yields (mg protein/g wet cells):** Mtb CDC 1551, 0.80—5.58 (n = 8); *M. aromaticivorans*, 7.79 (n = 1); *M. aurum*, 9.13 (n = 1); *M. chubuense*, 12.70 (n = 1); *M. fallax*, 6.88 (n = 1); *M. hiberniae*, 5.33—5.99 (n = 2); *M. intracellulare*, 8.23 (n = 1); *M. marinum*, 0.69 (n = 1); *M. salmoniphilum*, 2.52—3.89 (n = 3); *M. smegmatis*, 8.98 (n = 1); *M. terrae*, 1.58 (n = 1). Panels separated by white space were taken from separate images.

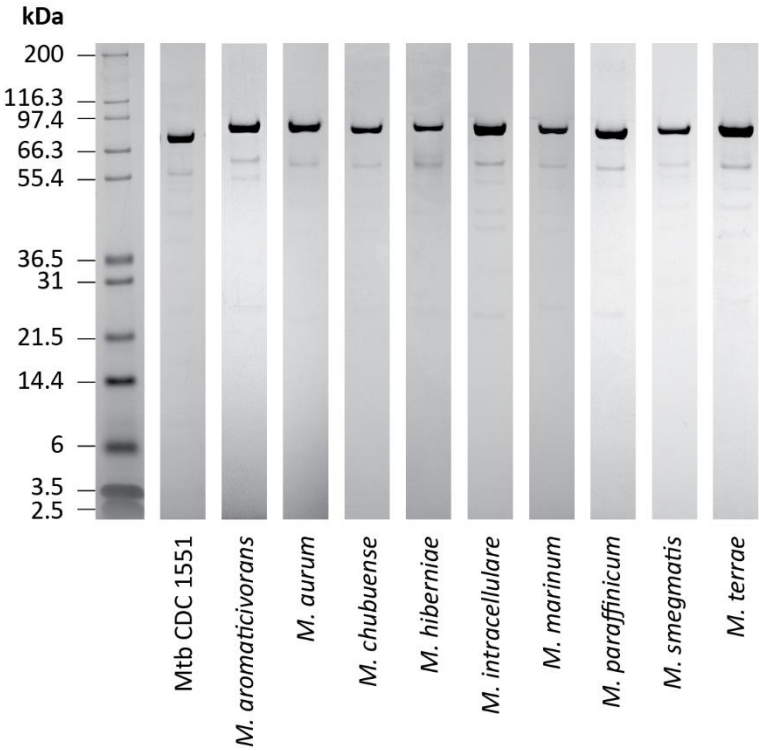


Figure 68: Purification of mycobacterial ICL2-His$_6$FLAGs.
ICL2-His$_6$FLAGs were purified using **IMAC Method 3** followed by buffer exchange using gravity SEC, as shown in **Figure 66**. ICL2-His$_6$FLAG from Mtb CDC 1551, which was purified in **Chapter 4**, is shown for comparison. **Expected ICL2-His$_6$FLAG sizes:** Mtb CDC 1551, 88.09 kDa; *M. aromaticivorans*, 88.10 kDa; *M. aurum*, 87.95 kDa; *M. chubuense*, 88.06 kDa; *M. hiberniae*, 87.23 kDa; *M. intracellulare*, 87.84 kDa; *M. marinum*, 88.01 kDa; *M. paraffinicum*, 87.80 kDa; *M. smegmatis*, 88.10 kDa; *M. terrae*, 87.25 kDa. **ICL2-His$_6$FLAG yields (mg protein/g wet cells):** Mtb CDC 1551, 8.97—28.28 (n = 8); *M. aromaticivorans*, 4.39 (n = 1); *M. aurum*, 25.12 (n = 1); *M. chubuense*, 6.35 (n = 1); *M. hiberniae*, 22.73 (n = 1); *M. intracellulare*, 26.57 (n = 1); *M. marinum*, 16.00—23.24 (n = 2); *M. paraffinicum*, 24.33 (n = 1); *M. smegmatis*, 24.76 (n = 1); *M. terrae*, 9.22 (n = 1). Panels separated by white space were taken from separate images.

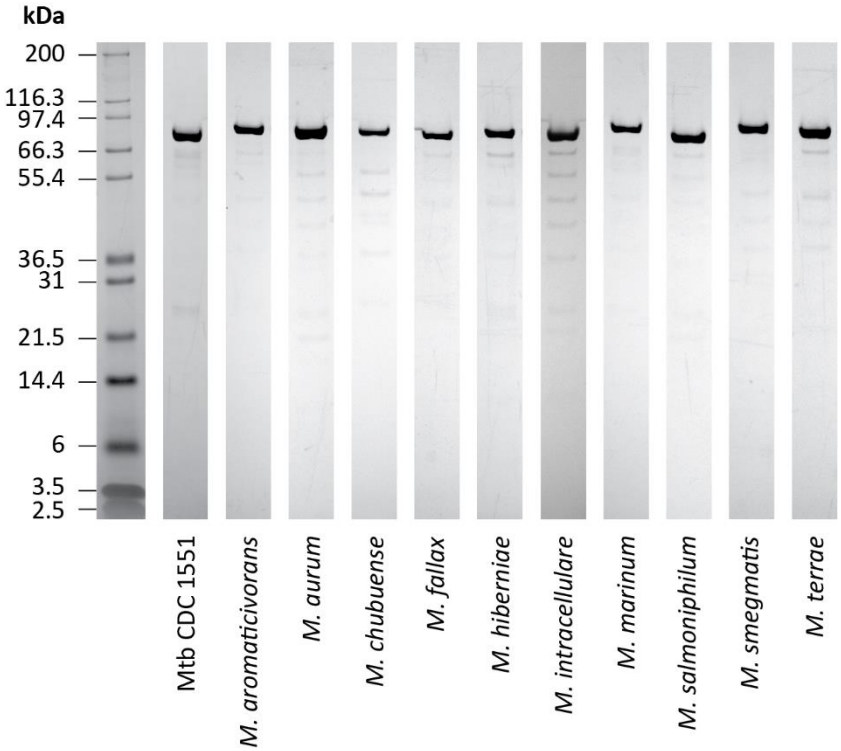


Figure 69: Purification of mycobacterial IDH2-His$_6$FLAGs.
IDH2-His$_6$FLAGs were purified using **IMAC Method 3** followed by buffer exchange using gravity SEC, as shown in **Figure 66**. IDH2-His$_6$FLAG from Mtb CDC 1551, which was purified in **Chapter 4**, is shown for comparison. **Expected IDH2-His$_6$FLAG sizes:** Mtb CDC 1551, 85.32 kDa; *M. aromaticivorans*, 84.79 kDa; *M. aurum*, 85.33 kDa; *M. chubuense*, 85.72 kDa; *M. fallax*, 85.12 kDa; *M. hiberniae*, 84.94 kDa; *M. intracellulare*, 85.57 kDa; *M. marinum*, 85.18 kDa; *M. salmoniphilum*, 84.68 kDa; *M. smegmatis*, 85.38 kDa; *M. terrae*, 84.85 kDa. **IDH2-His$_6$FLAG yields (mg protein/g wet cells):** Mtb CDC 1551, 10.86—21.88 (n = 5); *M. aromaticivorans*, 8.90 (n = 1); *M. aurum*, 1.57 (n = 1); *M. chubuense*, 12.14 (n = 1); *M. fallax*, 11.38 (n = 1); *M. hiberniae*, 8.32 (n = 1); *M. intracellulare*, 14.04 (n = 1); *M. marinum*, 9.89 (n = 1); *M. salmoniphilum*, 6.03 (n = 1); *M. smegmatis*, 0.15—11.18 (n = 3); *M. terrae*, 20.64 (n = 1). Panels separated by white space were taken from separate images.

IDH1-His$_6$FLAGs were purified using **IMAC Method 6** followed by desalting and AIEX (**Figure 66B** and **Figure 66D**). In each case, AIEX fractions containing the enzyme of interest were spiked with $MnCl_2$ and buffer exchange was performed using gravity SEC. The three IDH1-His$_6$FLAGs were thus obtained in sufficient yields and with appropriate purity for steady-state kinetics (**Figure 70**). Enzyme yields were determined after buffer exchange, and were higher than that achieved for Mtb IDH1-His$_6$FLAG, varying in the range 0.59—1.28 mg/g wet cells (for individual enzyme yields, see **Chapter 8, Table 51**).

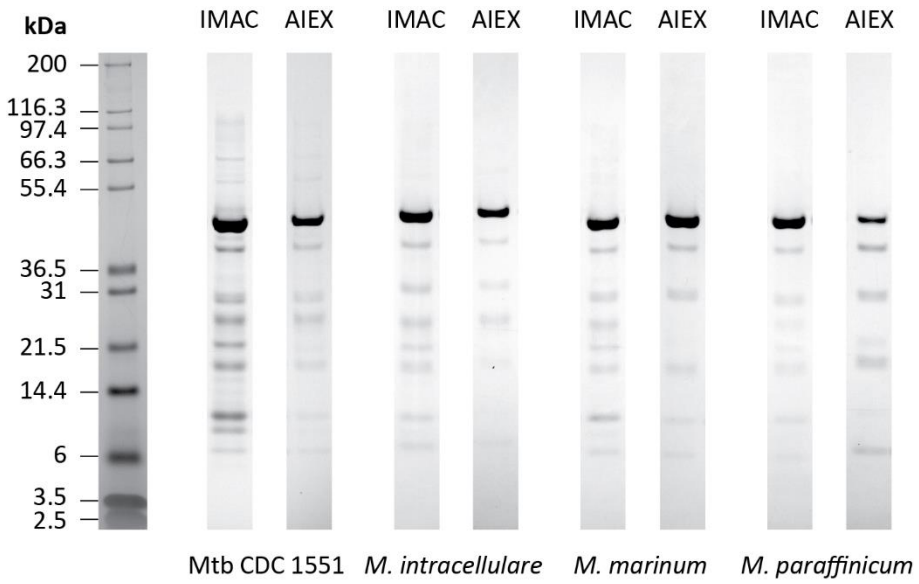


Figure 70: Purification of mycobacterial IDH1-His$_6$FLAGs.
IDH1-His$_6$FLAGs were purified using **IMAC Method 6** followed by desalting, AIEX, and gravity SEC, as shown in **Figure 66**. IDH1-His$_6$FLAG from Mtb CDC 1551, which was purified in **Chapter 4**, is shown for comparison. **Expected IDH1-His$_6$FLAG sizes:** Mtb CDC 1551, 48.28 kDa; *M. intracellulare*, 48.90 kDa; *M. marinum*, 48.42 kDa; *M. paraffinicum*, 48.55 kDa. **IDH1-His$_6$FLAG yields (mg protein/g wet cells):** Mtb CDC 1551, 0.08 (n = 1); *M. intracellulare*, 0.59 (n = 1); *M. marinum*, 0.68 (n = 1); *M. paraffinicum*, 1.28 (n = 1). **Predicted isoelectric points (pI):** Mtb CDC 1551, 5.23; *M. intracellulare*, 5.25; *M. marinum*, 5.78; *M. paraffinicum*, 5.61. The variation in pI between species was responsible for the observed differences in the pattern of contaminants remaining after AIEX. Panels separated by white space were taken from separate images. Protein purity was not quantified.

5.4. Kinetic characterisation and comparison of ICL1s

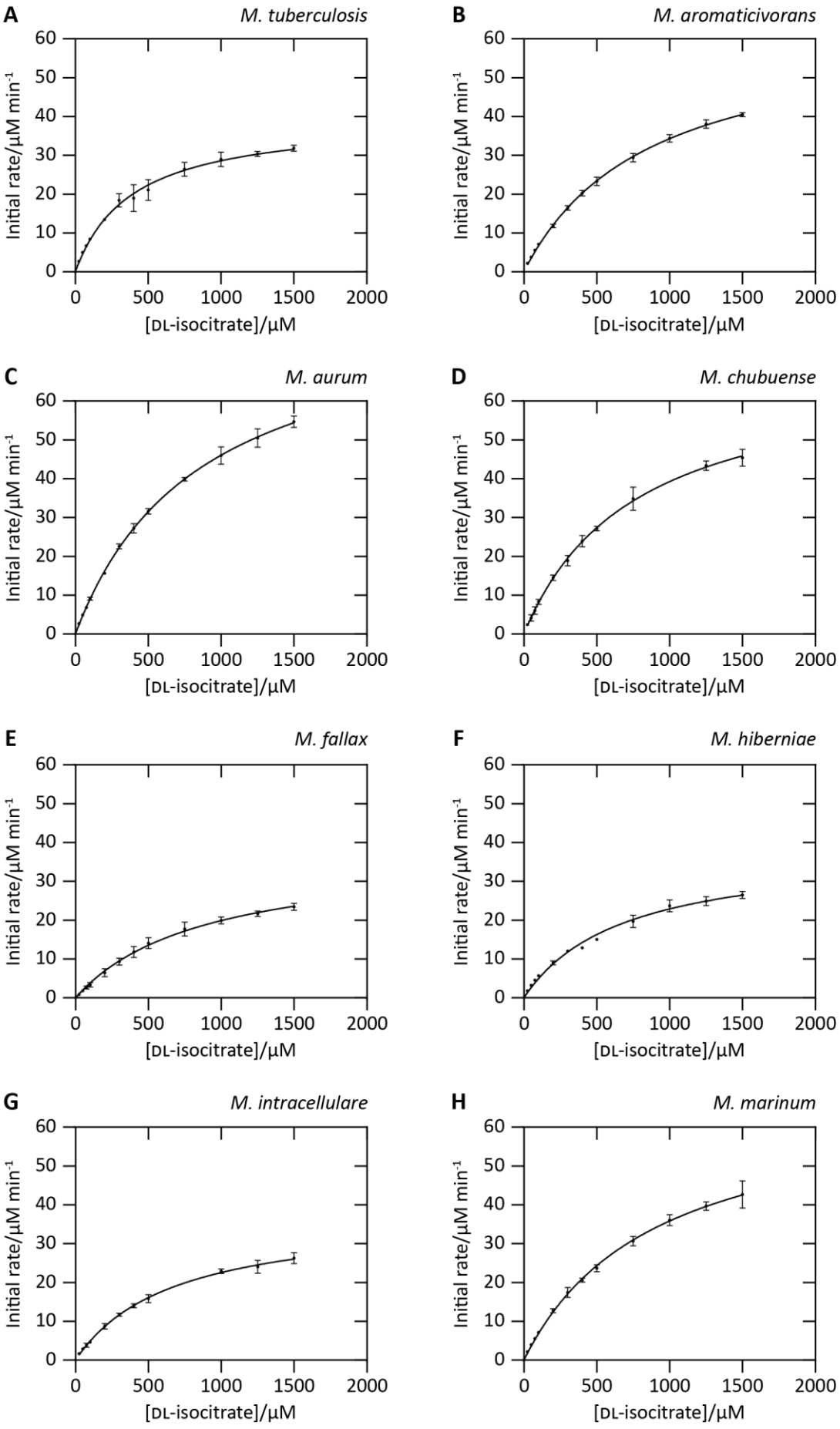
5.4.1. Determination of kinetic parameters for ICL1s

Mycobacterial ICL1s were characterised using the LDH-coupled assay previously optimised for Mtb CDC 1551 ICL1 (**Chapter 4, Section 4.2.2**). All ten newly characterised enzymes were active, and assay re-optimisation was not required. Briefly, each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 25—1500 µM and reactions were initiated with 30 nM ICL1, as for the Mtb enzyme. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The initial velocity plots for each enzyme are shown in **Figure 71**, and the kinetic

parameters thus obtained are summarised in **Table 21**. In both cases, the data for Mtb CDC 1551 ICL1 acquired in **Chapter 4** are presented to aid comparison.

Table 21: Kinetic parameters for mycobacterial ICL1s.
ICL1 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The kinetic parameters for Mtb CDC 1551 ICL1 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} **(μM min^{-1})**	K_M**(DL-isocitrate) (μM)**	k_{cat} **(s^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**
Mtb CDC 1551	40 ± 2	390 ± 59	22 ± 1	$5.7\times10^4 \pm 9\times10^3$
M. aromaticivorans	64 ± 2	860 ± 63	35 ± 1	$4.1\times10^4 \pm 3\times10^3$
M. aurum	86 ± 4	860 ± 70	48 ± 2	$5.5\times10^4 \pm 5\times10^3$
M. chubuense	69 ± 4	760 ± 90	39 ± 2	$5.0\times10^4 \pm 7\times10^3$
M. fallax	38 ± 3	890 ± 140	21 ± 2	$2.3\times10^4 \pm 4\times10^3$
M. hiberniae	38 ± 3	680 ± 100	21 ± 1	$3.1\times10^4 \pm 5\times10^3$
M. intracellulare	38 ± 2	670 ± 78	21 ± 1	$3.1\times10^4 \pm 4\times10^3$
M. marinum	68 ± 4	900 ± 110	38 ± 3	$4.2\times10^4 \pm 6\times10^3$
M. salmoniphilum	40 ± 2	460 ± 49	22 ± 1	$4.7\times10^4 \pm 5\times10^3$
M. smegmatis	36 ± 1	410 ± 39	20 ± 1	$4.9\times10^4 \pm 5\times10^3$
M. terrae	57 ± 5	820 ± 130	32 ± 3	$3.8\times10^4 \pm 7\times10^3$


A
M. tuberculosis
B
M. aromaticivorans
C
M. aurum
D
M. chubuense
E
M. fallax
F
M. hiberniae
G
M. intracellulare
H
M. marinum
Initial rate/μM min-1
[DL-isocitrate]/μM
0
10
20
30
40
50
60
0
500
1000
1500
2000

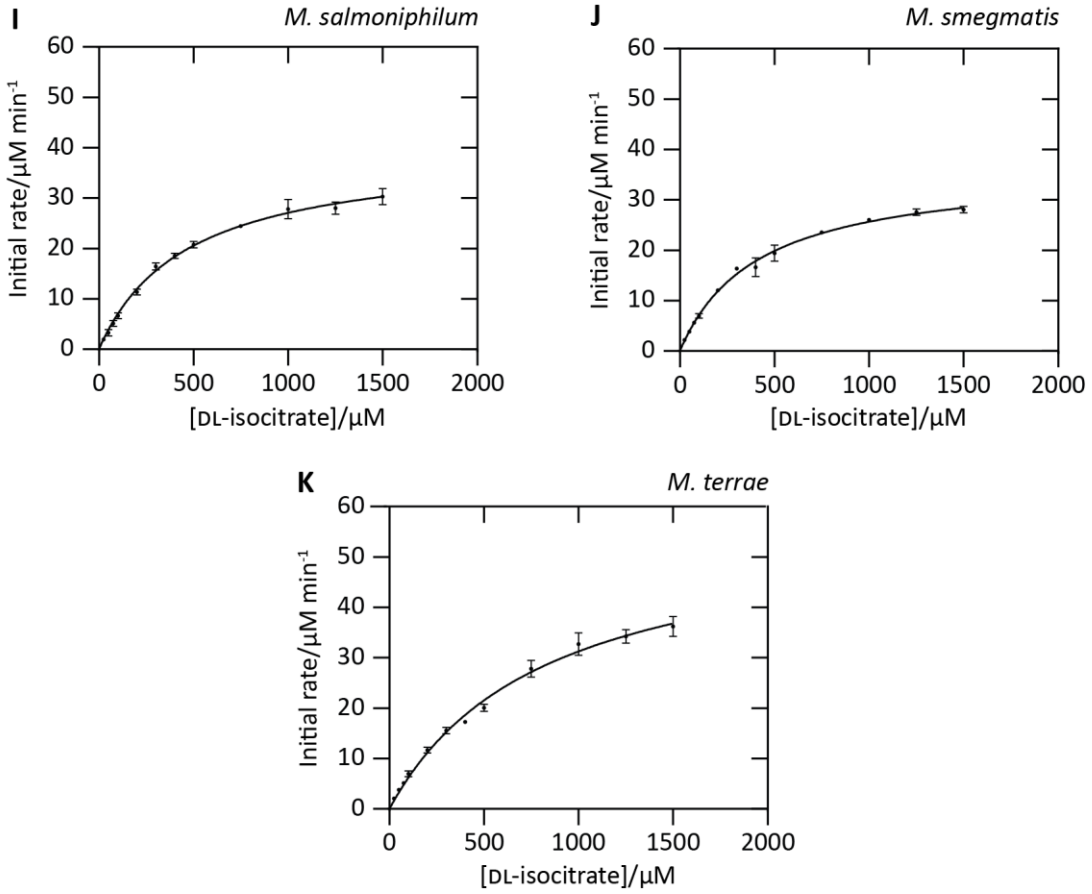


Figure 71: Steady-state kinetics of mycobacterial ICL1s.
Initial velocity data for mycobacterial ICL1s at isocitrate concentrations in the range 25—1500 µM. Reactions were carried out as described above and in **Chapter 8**, and were performed in triplicate or in quadruplicate. Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation (solid lines). The resulting kinetic parameters are presented in **Table 21**. The initial velocity data for Mtb CDC 1551 ICL1 determined in **Chapter 4** are presented for comparison (plot **A**).

5.4.2. Comparison of mycobacterial ICL1s

Having characterised eleven ICL1s, we can now compare their kinetic parameters. The values for K_M(DL-isocitrate) and k_{cat} shown in **Table 21** are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 72A** and **Figure 72B** respectively. Although the ICL1s from Mtb, *M. smegmatis*, and *M. salmoniphilum* exhibit a higher affinity for isocitrate than the other ICL1s assayed, K_M(DL-isocitrate) falls within 0.25 $\log_{10}$ of the mean (700 µM) for all species except Mtb. The total variation in K_M(DL-isocitrate) falls within a 0.5 $\log_{10}$ range and is therefore not biologically significant

(**Section 5.1**). Similarly, there is no biologically significant variation in k_{cat} between ICL1s – for all species, k_{cat} falls within 0.25 $\log_{10}$ of the mean (28.9 s^{-1}). Of note, the ICL1s from phylogenetically and ecologically diverse species (Mtb, *M. intracellulare*, *M. salmoniphilum*, *M. fallax*, *M. hiberniae*, and *M. smegmatis*) have almost identical k_{cat} values. Closely related species, such as *M. terrae* and *M. hiberniae*, often have different k_{cat} values.

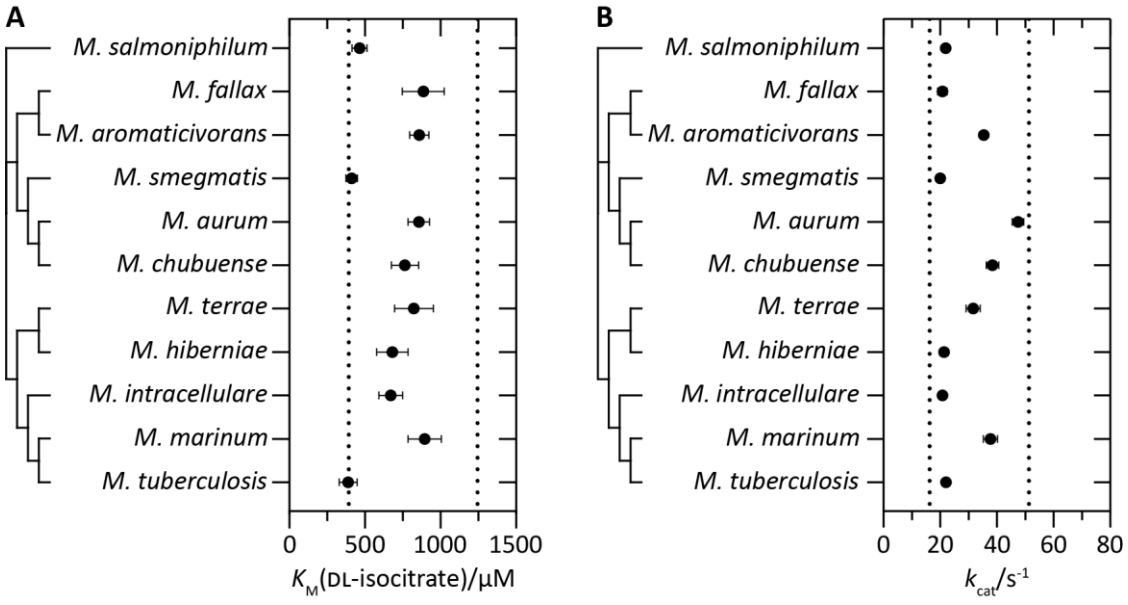


Figure 72: Comparison of K_M(DL-isocitrate) and k_{cat} for eleven mycobacterial ICL1s.
Values for K_M(DL-isocitrate) and k_{cat} are those obtained as described above and presented in **Table 21**. There is no biologically significant variation in K_M(DL-isocitrate) or k_{cat} between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). **A – Comparison of K_M(DL-isocitrate).** The dotted lines (at 394 and 1245 μM) indicate ±0.25 $\log_{10}$ of the mean (700 μM). **B – Comparison of k_{cat}.** The dotted lines (at 16.3 and 51.4 s^{-1}) indicate ±0.25 $\log_{10}$ of the mean (28.9 s^{-1}). Trees are presented in an unscaled format.

The values for k_{cat}/K_M shown in **Table 21** are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 73**. For all eleven species studied, $\log_{10}[k_{cat}/K_M]$ falls within 0.25 $\log_{10}$ of the mean (4.61), indicating that ICL1s do not exhibit biologically significant variation in their catalytic efficiency with respect to isocitrate. Overall, therefore, we did not observe any biologically significant variation in

mycobacterial ICL1 kinetics. Where subtle variation between species was observed, it was neither phylogenetically nor ecologically associated.

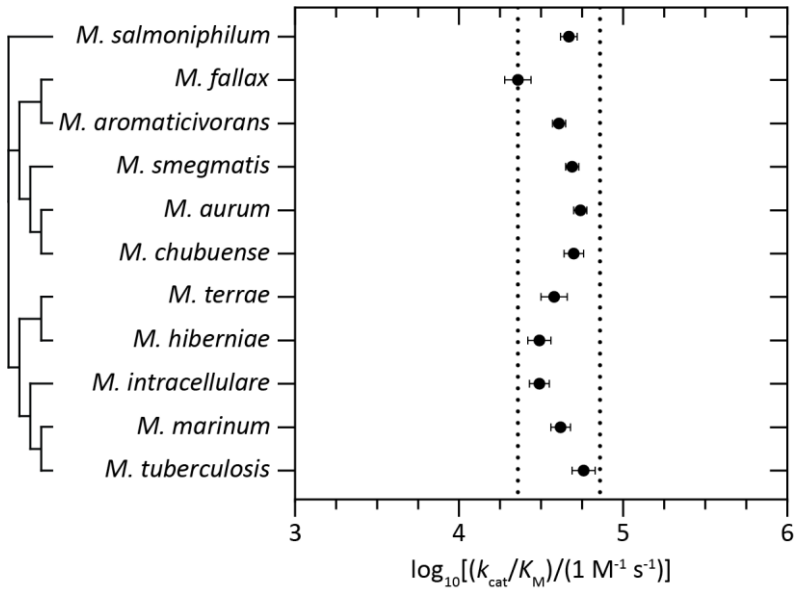


Figure 73: Comparison of $\log_{10}[k_{cat}/K_M]$ for eleven mycobacterial ICL1s.
Values for k_{cat}/K_M are those obtained as described above and presented in **Table 21**, plotted on a $\log_{10}$ scale to aid the identification of biologically significant variation. The dotted lines (at 4.36 and 4.86) indicate ±0.25 $\log_{10}$ of the mean (4.61). There is no biologically significant variation in $\log_{10}[k_{cat}/K_M]$ between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). The tree is presented in an unscaled format.

5.5. Kinetic characterisation and comparison of ICL2s

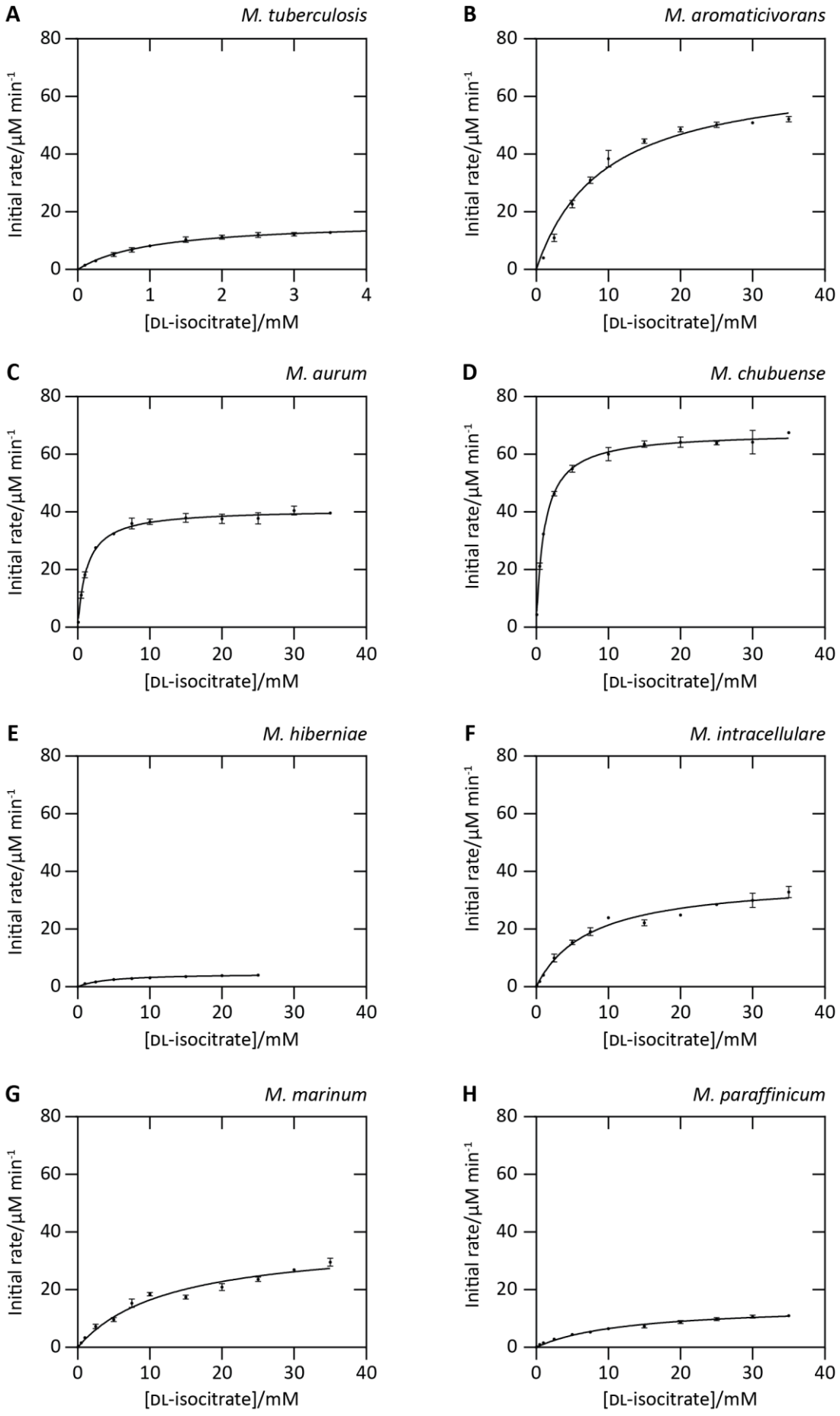
5.5.1. Determination of kinetic parameters for ICL2s

The LDH-coupled assay optimised for Mtb CDC 1551 ICL2 in **Chapter 4, Section 4.3.3** was used as a starting point for the characterisation of mycobacterial ICL2s. As the majority of the new enzymes were more active than Mtb CDC 1551 ICL2, in most assays the enzyme concentration was reduced from 200 nM to 50 nM. Two-thirds of the new ICL2s had substantially higher values for K_M(DL-isocitrate) than Mtb CDC 1551 ICL2, so the range of isocitrate concentrations used was extended to include concentrations up to 35 mM. Assays were performed in the absence of Ac-CoA.

To summarise, each reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), and LDH (110 U/mL). The isocitrate concentration was varied in the range 0.1—35 mM, except for *M. aromaticivorans* ICL2 (1—35 mM isocitrate) and *M. hiberniae* ICL2 (1—25 mM isocitrate). Reactions were initiated with 50 nM ICL2, except for *M. terrae* ICL2, where 200 nM enzyme was used. Assays were performed in triplicate, and the order of reactions within each replicate was randomised. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The initial velocity plots for each enzyme are shown in **Figure 74**, and the kinetic parameters thus obtained are summarised in **Table 22**. In both cases, the data for Mtb CDC 1551 ICL2 acquired in **Chapter 4** are presented to aid comparison.

Table 22: Kinetic parameters for mycobacterial ICL2s.
ICL2 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The kinetic parameters for Mtb CDC 1551 ICL2 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} **(µM min^{-1})**	K_M**(DL-isocitrate) (mM)**	k_{cat} **(s^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**
Mtb CDC 1551	17 ± 1	1.1 ± 0.18	1.4 ± 0.1	$1.3\times10^3 \pm 2\times10^2$
M. aromaticivorans	69 ± 4	9.4 ± 1.6	23 ± 1	$2.4\times10^3 \pm 5\times10^2$
M. aurum	41 ± 0.8	1.3 ± 0.17	14 ± 0.3	$1.0\times10^4 \pm 1\times10^3$
M. chubuense	68 ± 1	1.1 ± 0.10	23 ± 0.3	$2.0\times10^4 \pm 2\times10^3$
M. hiberniae	4.6 ± 0.2	4.3 ± 0.79	1.5 ± 0.1	$3.5\times10^2 \pm 7\times10^1$
M. intracellulare	37 ± 3	7.5 ± 1.6	12 ± 0.8	$1.7\times10^3 \pm 4\times10^2$
M. marinum	37 ± 4	13 ± 3.2	12 ± 1	$1.0\times10^3 \pm 3\times10^2$
M. paraffinicum	15 ± 1	13 ± 2.3	4.9 ± 0.4	$3.9\times10^2 \pm 8\times10^1$
M. smegmatis	56 ± 1	0.75 ± 0.12	19 ± 0.5	$2.5\times10^4 \pm 4\times10^3$
M. terrae	19 ± 2	13 ± 3.0	1.6 ± 0.2	$1.3\times10^2 \pm 3\times10^1$


A
M. tuberculosis
B
M. aromaticivorans
C
M. aurum
D
M. chubuense
E
M. hiberniae
F
M. intracellulare
G
M. marinum
H
M. paraffinicum
Initial rate/µM min-1
[DL-isocitrate]/mM

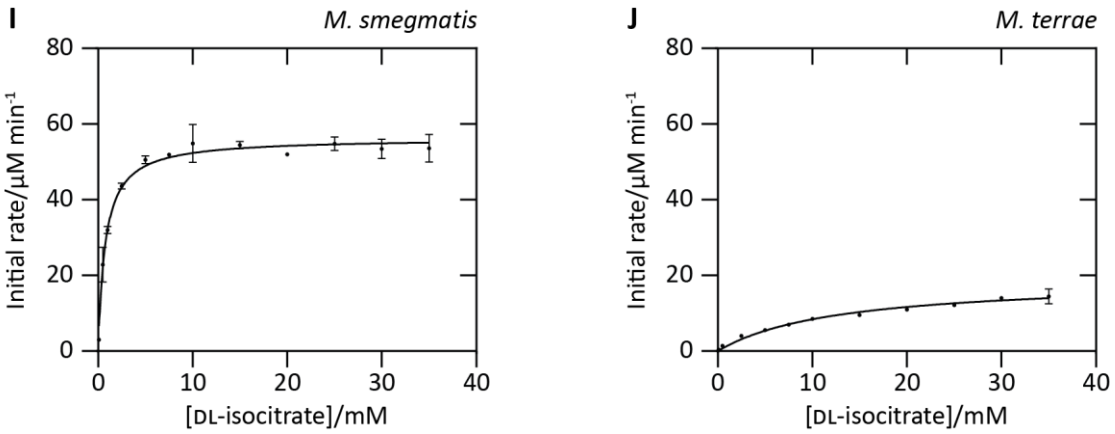


Figure 74: Steady-state kinetics of mycobacterial ICL2s.
Initial velocity data for mycobacterial ICL2s. The concentration of isocitrate was varied in the range 0.1—35 mM, except for *M. aromaticivorans* ICL2 (plot **B**, 1—35 mM isocitrate) and *M. hiberniae* ICL2 (plot **E**, 1—25 mM isocitrate). Reactions were carried out as described above and in **Chapter 8**, and were performed in triplicate. 50 nM ICL2 was used for initiation, except for *M. terrae* (plot **J**, 200 nM ICL2). Initial rates were extracted, converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation (solid lines). The resulting kinetic parameters are presented in **Table 22**. The initial velocity data for Mtb CDC 1551 ICL2 (with 50 nM ICL2 at isocitrate concentrations in the range 0.1—4 mM) determined in **Chapter 4** are presented for comparison (plot **A**).

5.5.2. Comparison of mycobacterial ICL2s

Plotting the kinetic parameters of our ten ICL2s (**Table 22**) against the phylogenetic tree of the corresponding mycobacterial species (K_M(DL-isocitrate), **Figure 75**; k_{cat}, **Figure 76**; k_{cat}/K_M, **Figure 77**), it is immediately apparent that ICL2s are more kinetically variable than ICL1s. **Figure 75** shows that ICL2s exhibit biologically significant variation in K_M(DL-isocitrate) – there is a $\log_{10}$ difference between *M. marinum* (highest K_M(DL-isocitrate)) and *M. smegmatis* (lowest K_M(DL-isocitrate)). Mtb ICL2 has an unusually low K_M(DL-isocitrate) for its phylogenetic position. Also of note is the cluster of fast-growing species – *M. smegmatis*, *M. aurum*, and *M. chubuense* – for which K_M(DL-isocitrate) is similar to Mtb ICL2, but low relative to the other ICL2s assayed. This pattern suggests that affinity for isocitrate has evolved on at least two separate occasions.

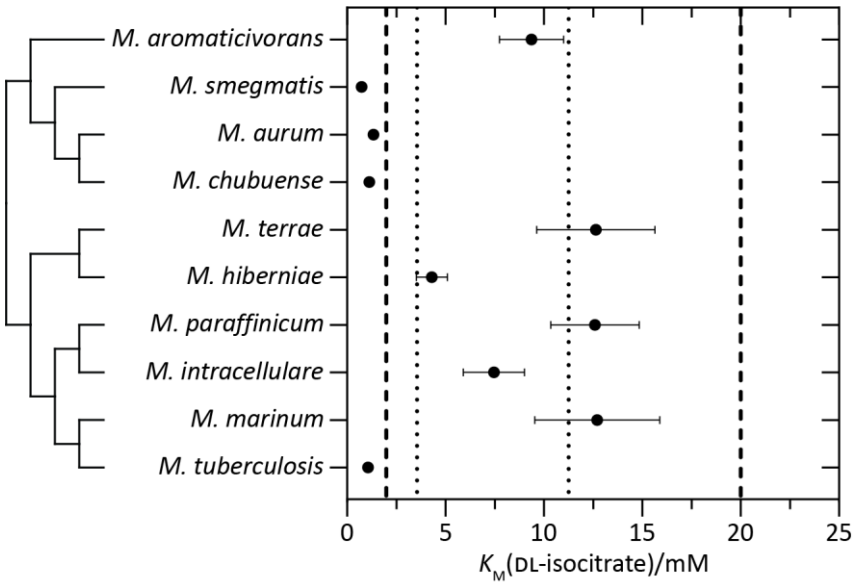


Figure 75: Comparison of K_M(DL-isocitrate) for ten mycobacterial ICL2s.
Values for K_M(DL-isocitrate) are those obtained as described above and presented in **Table 22**. The dotted lines (at 3.6 and 11.3 mM) indicate ±0.25 $\log_{10}$ of the mean (6.3 mM). The dashed lines (at 2.0 and 20.0 mM) indicate ±0.5 $\log_{10}$ of the mean. There is biologically significant variation in K_M(DL-isocitrate) between the enzymes tested; there is a $\log_{10}$ difference in K_M(DL-isocitrate) between *M. marinum* (highest K_M of the enzymes tested) and *M. smegmatis* (lowest K_M of the enzymes tested). The tree is presented in an unscaled format.

ICL2s also exhibit biologically significant variation in k_{cat} (**Figure 76**) – there is a 1.3 $\log_{10}$ difference between *M. aromaticivorans* (highest k_{cat}) and Mtb (lowest k_{cat}). The ICL2s from fast growing mycobacterial species turn over isocitrate more rapidly than those from slow growing species. Turnover of isocitrate is particularly slow in *M. terrae* and *M. hiberniae*, which are members of the same clade. The least active ICL2, that from Mtb, has a lower k_{cat} than might be predicted from its phylogenetic position. These results indicate that variation in ICL2 k_{cat} is phylogenetically rather than ecologically associated, and that changes in k_{cat} have occurred on multiple occasions during ICL2 evolution.

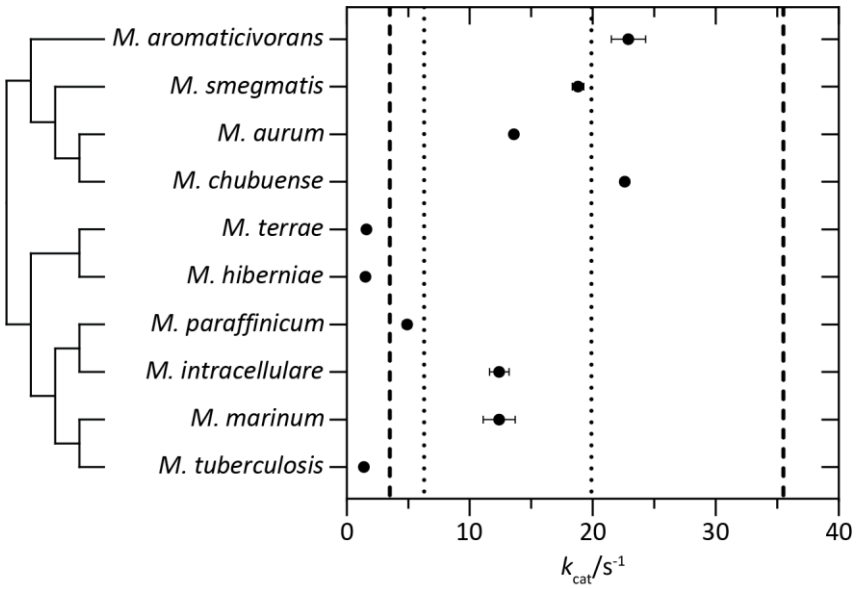


Figure 76: Comparison of k_{cat} for ten mycobacterial ICL2s.
Values for k_{cat} are those obtained as described above and presented in **Table 22**. The dotted lines (at 6.3 and 19.9 s^{-1}) indicate ±0.25 $\log_{10}$ of the mean (11.2 s^{-1}). The dashed lines (at 3.5 and 35.5 s^{-1}) indicate ±0.5 $\log_{10}$ of the mean. There is biologically significant variation in k_{cat} between the enzymes tested; there is a 1.3 $\log_{10}$ difference in k_{cat} between *M. aromaticivorans* (highest k_{cat} of the enzymes tested) and *M. tuberculosis* (lowest k_{cat} of the enzymes tested). The tree is presented in an unscaled format.

Finally, **Figure 77** shows that ICL2s exhibit biologically significant variation in their catalytic efficiency with respect to isocitrate, and that this variation is phylogenetically associated. The ICL2s from fast-growing mycobacterial species have higher values of k_{cat}/K_M than the ICL2s from slow growers. Of note, the ICL2s with similar catalytic efficiencies to Mtb ICL2 (*M. aromaticivorans*, *M. intracellulare,* and *M. marinum*) have higher values for both k_{cat} and K_M(DL-isocitrate) than the Mtb enzyme. In conclusion, ICL2s display biologically significant variation in all three kinetic parameters analysed. This variation is phylogenetically rather than ecologically associated. The evolutionary implications of this observation will be discussed later (**Section 5.9.1**).

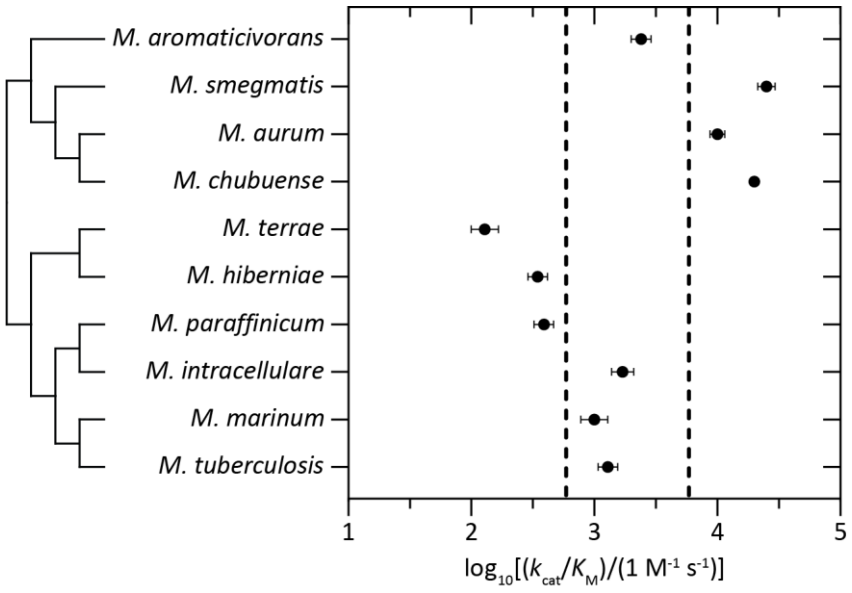


Figure 77: Comparison of $\log_{10}[k_{cat}/K_M]$ for ten mycobacterial ICL2s.
Values for k_{cat}/K_M are those obtained as described above and presented in **Table 22**, plotted on a $\log_{10}$ scale to aid the identification of biologically significant variation. The dashed lines (at 2.77 and 3.77) indicate ±0.5 $\log_{10}$ of the mean (3.27). There is biologically significant variation in $\log_{10}[k_{cat}/K_M]$ between the enzymes tested; there is a 2.28 $\log_{10}$ difference in catalytic efficiency between the most (*M. smegmatis*) and least (*M. terrae*) efficient enzymes tested. The tree is presented in an unscaled format.

5.6. Kinetic characterisation and comparison of IDH2s

5.6.1. Determination of kinetic parameters for IDH2s

Mycobacterial IDH2s were characterised using the assay previously optimised for Mtb CDC 1551 IDH2 (**Chapter 4, Section 4.4.3**); assay re-optimisation was not required. All eleven enzymes were found to be active, and kinetic parameters were determined with respect to each of the two co-substrates, isocitrate and $NADP^+$. To determine kinetic parameters with respect to isocitrate, each reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), and $NADP^+$ (2 mM). The isocitrate concentration was varied in the range 10—800 µM. To determine kinetic parameters with respect to $NADP^+$, each reaction mixture contained HEPES (100 mM, pH 7.5, $MnCl_2$ (5 mM), and isocitrate (2 mM). The $NADP^+$ concentration was varied in the range 10 µM—2 mM. In both cases, reactions were initiated with 5 nM IDH2 and samples were randomised within each

replicate. As some IDH2s displayed substrate inhibition kinetics, kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation, or to a substrate inhibition model, as appropriate. The kinetic parameters obtained by varying the concentration of isocitrate are presented in **Table 23**; those obtained by varying the concentration of $NADP^+$ are presented in **Table 24**. The initial velocity plots for assays in which the concentration of isocitrate was varied are shown in **Figure 78**. In each case, the data for Mtb CDC 1551 IDH2 acquired in **Chapter 4** are presented to aid comparison.

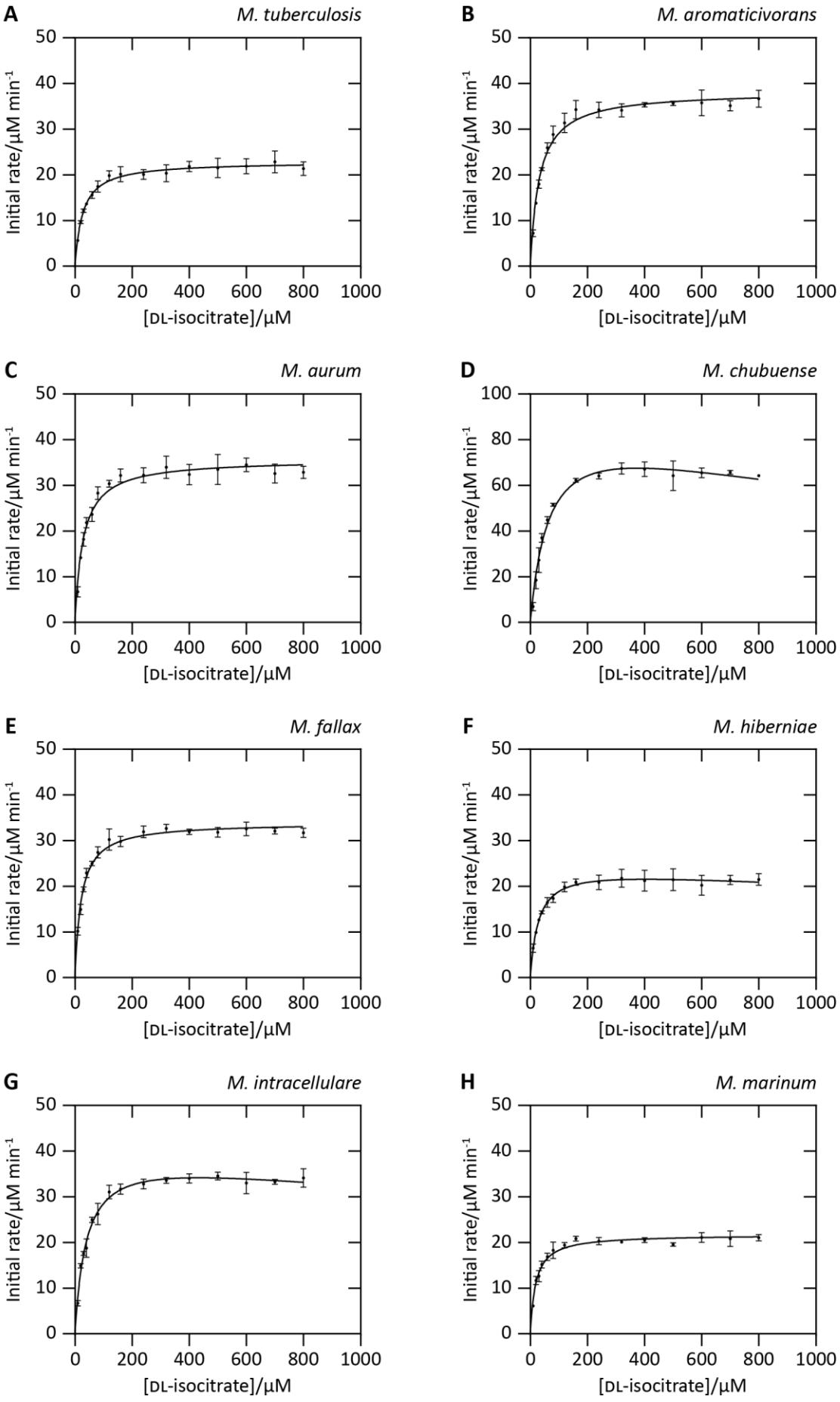
Table 23: Kinetic parameters for mycobacterial IDH2s determined by varying the concentration of DL-isocitrate.
IDH2 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation (*M. aromaticivorans*, *M. aurum*, *M. fallax*, *M. marinum*, *M. salmoniphilum*, *M. smegmatis*, and *M. terrae*) or to a substrate inhibition model (*M. chubuense*, *M. hiberniae*, and *M. intracellulare*). The kinetic parameters for Mtb CDC 1551 IDH2 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
Mtb CDC 1551	23 ± 0.7	27 ± 3.9	76 ± 2	$2.8\times10^6 \pm 4\times10^5$
M. aromaticivorans	38 ± 0.9	32 ± 3.7	130 ± 3	$4.0\times10^6 \pm 5\times10^5$
M. aurum	36 ± 1	28 ± 4.2	120 ± 4	$4.3\times10^6 \pm 7\times10^5$
M. chubuense	91 ± 10	66 ± 15	300 ± 30	$5.0\times10^6 \pm 1\times10^6$
M. fallax	34 ± 0.7	22 ± 2.3	110 ± 2	$5.2\times10^6 \pm 6\times10^5$
M. hiberniae	25 ± 2	29 ± 6.5	82 ± 6	$2.9\times10^6 \pm 7\times10^5$
M. intracellulare	40 ± 3	40 ± 6.6	130 ± 9	$3.4\times10^6 \pm 6\times10^5$
M. marinum	22 ± 0.6	19 ± 2.8	73 ± 2	$3.8\times10^6 \pm 6\times10^5$
M. salmoniphilum	29 ± 0.4	22 ± 1.8	95 ± 1	$4.3\times10^6 \pm 4\times10^5$
M. smegmatis	31 ± 0.8	33 ± 2.5	100 ± 3	$3.1\times10^6 \pm 5\times10^5$
M. terrae	30 ± 0.7	28 ± 3.2	100 ± 2	$3.6\times10^6 \pm 4\times10^5$

Table 24: Kinetic parameters for mycobacterial IDH2s determined by varying the concentration of $NADP^+$.

IDH2 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation (*M. chubuense*, *M. fallax*, *M. intracellulare*, and *M. marinum*) or to a substrate inhibition model (*M. aromaticivorans*, *M. aurum*, *M. hiberniae*, *M. salmoniphilum*, *M. smegmatis*, and *M. terrae*). The kinetic parameters for Mtb CDC 1551 IDH2 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} **(μM min^{-1})**	K_M**(NADP$^+$) (μM)**	k_{cat} **(s^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**
Mtb CDC 1551	21 ± 0.6	42 ± 6.8	71 ± 2	$1.7\times10^6 \pm 3\times10^5$
M. aromaticivorans	47 ± 2	38 ± 5.8	160 ± 8	$4.1\times10^6 \pm 7\times10^5$
M. aurum	36 ± 2	24 ± 5.4	120 ± 7	$5.0\times10^6 \pm 1\times10^6$
M. chubuense	62 ± 0.9	28 ± 3.0	210 ± 3	$7.2\times10^6 \pm 8\times10^5$
M. fallax	31 ± 0.5	58 ± 6.2	100 ± 6	$1.8\times10^6 \pm 2\times10^5$
M. hiberniae	26 ± 2	36 ± 9.0	88 ± 6	$2.4\times10^6 \pm 6\times10^5$
M. intracellulare	31 ± 0.5	36 ± 4.0	100 ± 2	$2.9\times10^6 \pm 3\times10^5$
M. marinum	22 ± 0.6	26 ± 4.4	75 ± 2	$2.9\times10^6 \pm 5\times10^5$
M. salmoniphilum	30 ± 2	35 ± 7.5	99 ± 6	$2.8\times10^6 \pm 6\times10^5$
M. smegmatis	56 ± 3	48 ± 7.1	190 ± 9	$3.9\times10^6 \pm 6\times10^5$
M. terrae	33 ± 2	38 ± 7.9	110 ± 7	$3.0\times10^6 \pm 7\times10^5$


A
M. tuberculosis
B
M. aromaticivorans
C
M. aurum
D
M. chubuense
E
M. fallax
F
M. hiberniae
G
M. intracellulare
H
M. marinum
Initial rate/μM min-1
[DL-isocitrate]/μM

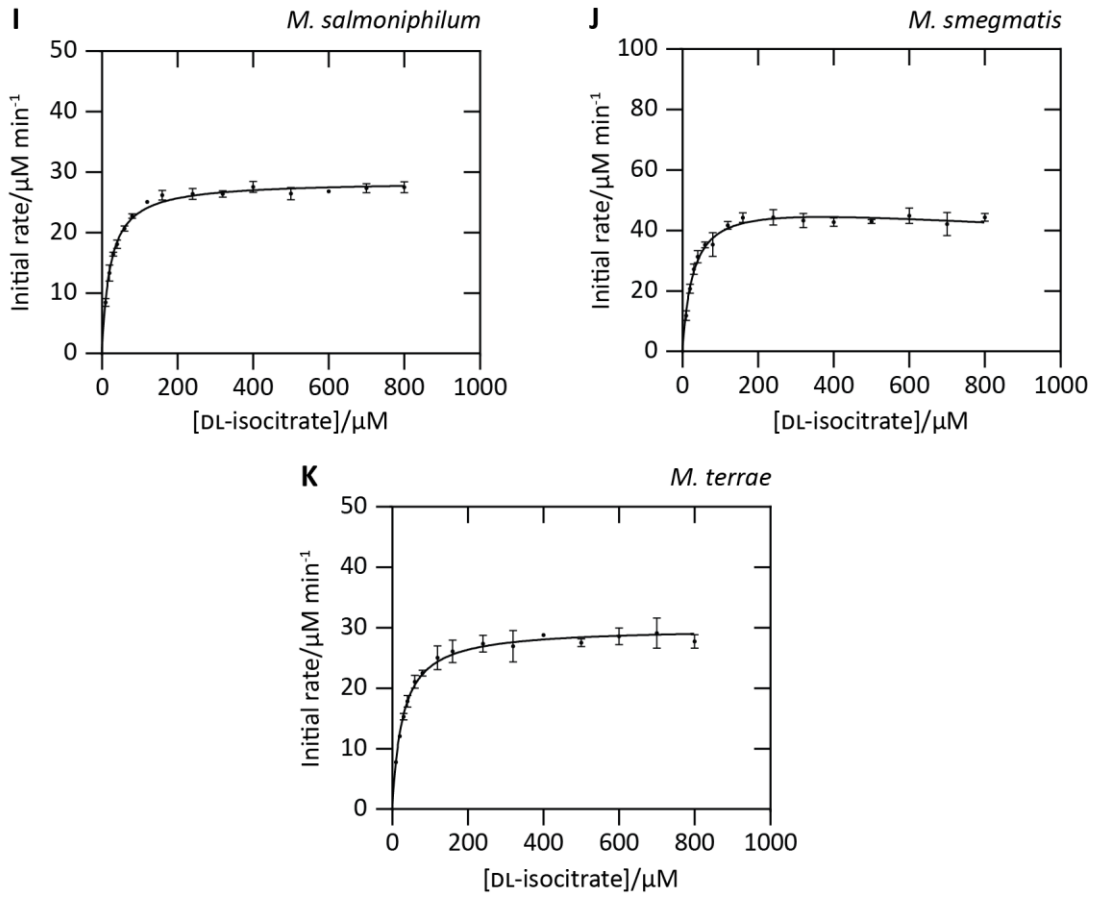


Figure 78: Steady-state kinetics of mycobacterial IDH2s.
Initial velocity data for mycobacterial IDH2s at isocitrate concentrations in the range 10—800 µM. Reactions were carried out as described above and in **Chapter 8**, and were performed in triplicate. Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation (*M. aromaticivorans*, **B**; *M. aurum*, **C**; *M. fallax*, **E**; *M. marinum*, **H**; *M. salmoniphilum*, **I**; *M. smegmatis*, **J**; and *M. terrae*, **K**) or to a substrate inhibition model (*M. chubuense*, **C**; *M. hiberniae*, **F**; and *M. intracellulare*, **G**). The resulting kinetic parameters are presented in **Table 23**. The initial velocity data for Mtb CDC 1551 IDH2 determined in **Chapter 4** are presented for comparison (**A**).

5.6.2. Comparison of mycobacterial IDH2s

Having characterised eleven IDH2s, we are now in a position to compare their kinetic parameters. K_M(DL-isocitrate) (**Table 23**) and K_M($NADP^+$) (**Table 24**) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 79A** and **Figure 79B** respectively. K_M(DL-isocitrate) falls within 0.25 $\log_{10}$ of the mean (31.2 µM) for all species except *M. chubuense*. The total variation in K_M(DL-isocitrate) falls within a 0.75 $\log_{10}$ range. For all species tested, K_M($NADP^+$) falls within 0.25 $\log_{10}$ of the mean

(37.3 µM). IDH2s therefore do not exhibit biologically significant variation in either K_M(DL-isocitrate) or K_M(NADP$^+$).

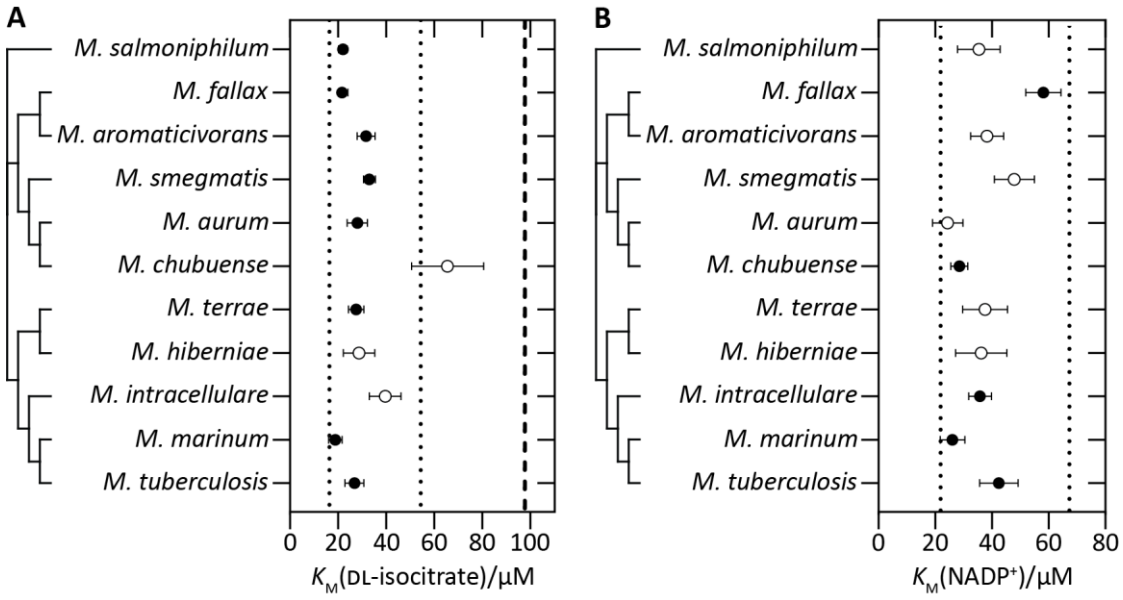


Figure 79: Comparison of K_M(DL-isocitrate) and K_M(NADP$^+$) for eleven mycobacterial IDH2s.
K_M values were obtained by fitting the initial velocity data to the Michaelis-Menten equation (black circles) or to a substrate inhibition model (open circles), as appropriate. Trees are presented in an unscaled format. There is no biologically significant variation in K_M(DL-isocitrate) or K_M(NADP$^+$) between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). **A – Comparison of K_M(DL-isocitrate).** Values for K_M(DL-isocitrate) are those obtained as described above and presented in **Table 23**. The dotted lines (at 17.5 and 55.5 µM) indicate ±0.25 $\log_{10}$ of the mean (31.2 µM). The dashed line (at 98.7 µM) marks 0.5 $\log_{10}$ above the mean. **B – Comparison of K_M(NADP$^+$).** Values for K_M(NADP$^+$) are those obtained as described above and presented in **Table 24**. The dotted lines (at 21.0 and 66.4 µM) indicate ±0.25 $\log_{10}$ of the mean (37.3 µM).

The values for k_{cat} obtained by varying the concentration of isocitrate (**Table 23**) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 80A**. As seen for K_M(DL-isocitrate), k_{cat} falls within 0.25 $\log_{10}$ of the mean for all species except *M. chubuense*. The total variation in k_{cat} falls within a 0.75 $\log_{10}$ range and is therefore not biologically significant. While *M. chubuense* IDH2 appears to have a higher activity and lower affinity for isocitrate than the other IDH2s tested, this result has yet to be confirmed in an independent experiment. If this phenotype is reproducible, investigation of the underlying differences in sequence and/or structure

relative to other IDH2s (such as Mtb IDH2) might prove mechanistically informative (IDH2s have yet to be mechanistically characterised, as noted in **Chapter 1, Section 1.7.3**).

Finally, the values for k_{cat}/K_M with respect to DL-isocitrate (**Table 23**) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 80B**. For all eleven species studied, $\log_{10}[k_{cat}/K_M]$ falls within 0.25 $\log_{10}$ of the mean (6.58), indicating that IDH2s do not exhibit biologically significant variation in their catalytic efficiency with respect to isocitrate. Overall, therefore, we did not observe any biologically significant variation in mycobacterial IDH2 kinetics.

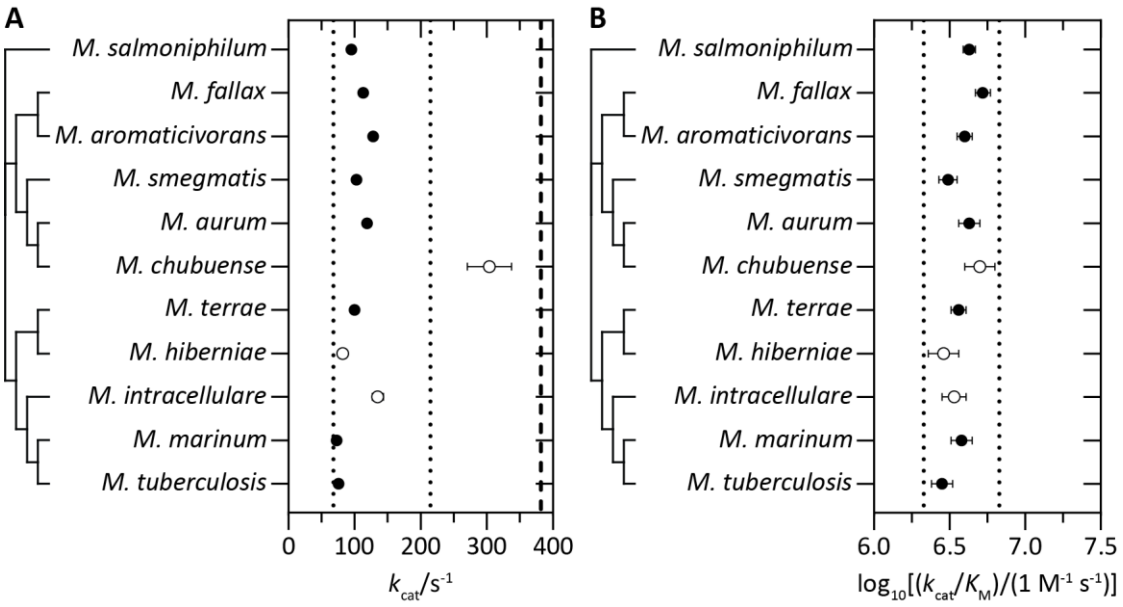


Figure 80: Comparison of k_{cat} and $\log_{10}[k_{cat}/K_M]$ for eleven mycobacterial IDH2s.
Values for k_{cat} and k_{cat}/K_M are those obtained by varying the concentration of DL-isocitrate, as described above and presented in **Table 23**. The initial velocity data were fitted to the Michaelis-Menten equation (black circles) or to a substrate inhibition model (open circles), as appropriate. Trees are presented in an unscaled format. There is no biologically significant variation in k_{cat} or $\log_{10}[k_{cat}/K_M]$ between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). **A – Comparison of k_{cat} with respect to DL-isocitrate.** The dotted lines (at 68 and 215 s^{-1}) indicate ±0.25 $\log_{10}$ of the mean (121 s^{-1}). The dashed line (at 382 s^{-1}) marks 0.5 $\log_{10}$ above the mean. **B – Comparison of $\log_{10}[k_{cat}/K_M]$ with respect to DL-isocitrate.** The dotted lines (at 6.33 and 6.83) indicate ±0.25 $\log_{10}$ of the mean (6.58).

5.7. Kinetic characterisation and comparison of IDH1s

5.7.1. Determination of kinetic parameters for IDH1s

As for the Mtb enzymes (**Chapter 4**), the assay developed for IDH2s could be used without re-optimisation for the characterisation of IDH1s. Kinetic parameters were therefore determined with respect to isocitrate and $NADP^+$, as described above (**Section 5.6**). In contrast to the IDH2s, the IDH1s did not display substrate inhibition kinetics, and the initial rates measured in each assay could be fitted to the Michaelis-Menten equation. The kinetic parameters obtained by varying the concentration of isocitrate are presented in **Table 25**; those obtained by varying the concentration of $NADP^+$ are presented in **Table 26**. The initial velocity plots for assays in which the concentration of isocitrate was varied are shown in **Figure 81**. In each case, the data for Mtb CDC 1551 IDH1 acquired in **Chapter 4** are presented to aid comparison.

Table 25: Kinetic parameters for mycobacterial IDH1s determined by varying the concentration of DL-isocitrate.
IDH1 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The kinetic parameters for Mtb CDC 1551 IDH1 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} (μM min^{-1})	K_M(DL-isocitrate) (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
Mtb CDC 1551	23 ± 1	83 ± 12	76 ± 4	$9.0\times10^5 \pm 1\times10^5$
M. intracellulare	38 ± 2	52 ± 7.3	130 ± 5	$2.5\times10^6 \pm 4\times10^5$
M. marinum	32 ± 1	42 ± 5.9	110 ± 4	$2.5\times10^6 \pm 4\times10^5$
M. paraffinicum	29 ± 1	47 ± 7.4	98 ± 4	$2.1\times10^6 \pm 4\times10^5$

Table 26: Kinetic parameters for mycobacterial IDH1s determined by varying the concentration of $NADP^+$.

IDH1 assays were performed as described above and in **Chapter 8**. Kinetic parameters were obtained by fitting the resultant initial rates to the Michaelis-Menten equation. The kinetic parameters for Mtb CDC 1551 IDH1 determined in **Chapter 4** are presented for comparison. Errors are reported as symmetrical 95% confidence intervals.

Species	v_{max} (μM min^{-1})	$K_M(NADP^+)$ (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
Mtb CDC 1551	23 ± 0.9	27 ± 6.7	76 ± 3	$2.9\times10^6 \pm 7\times10^5$
M. intracellulare	17 ± 0.7	11 ± 3.7	57 ± 2	$5.0\times10^6 \pm 2\times10^6$
M. marinum	27 ± 0.7	15 ± 2.9	90 ± 2	$6.0\times10^6 \pm 1\times10^6$
M. paraffinicum	26 ± 0.9	13 ± 3.6	86 ± 3	$7.0\times10^6 \pm 2\times10^6$

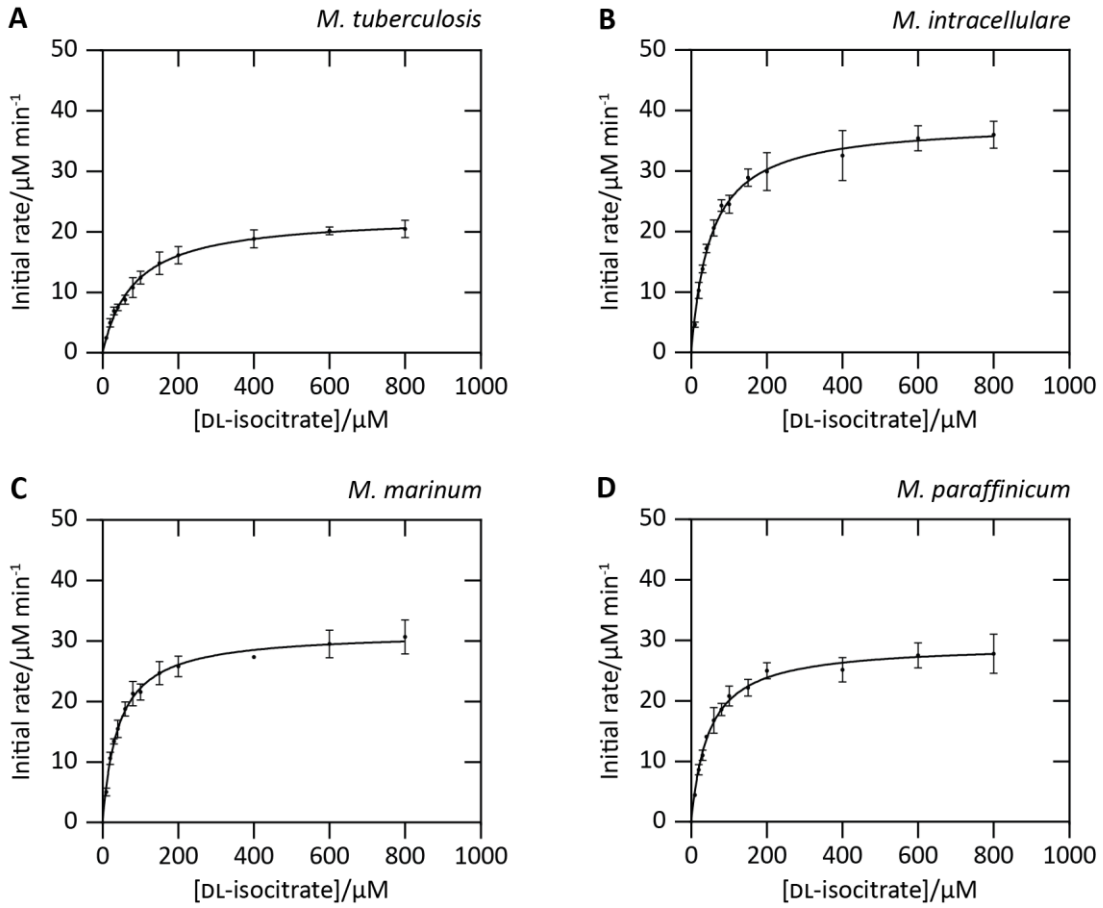


Figure 81: Steady-state kinetics of mycobacterial IDH1s.

Initial velocity data for mycobacterial IDH1s at isocitrate concentrations in the range 10—800 μM. Reactions were carried out as described above and in **Chapter 8**, and were performed in triplicate. Initial rates were extracted, converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}, and fitted to the Michaelis-Menten equation. The resulting kinetic parameters are presented in **Table 25**. The initial velocity data for Mtb CDC 1551 IDH1 determined in **Chapter 4** are presented for comparison (**A**).

5.7.2. Comparison of mycobacterial IDH1s

Having characterised four IDH1s, we are now in a position to compare their kinetic parameters. K_M(DL-isocitrate) (**Table 25**) and K_M(NADP$^+$) (**Table 26**) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 82A** and **Figure 82B** respectively. For both substrates, the variation in K_M between species falls within a 0.5 $\log_{10}$ range and is therefore not biologically significant. Notably, Mtb IDH1 has a lower affinity for both DL-isocitrate and NADP$^+$ than the other IDH1s we tested. Characterisation of additional mycobacterial IDH1s would reveal whether these affinity losses are particular to Mtb. Additionally, comparison of IDH1 sequences and/or structures would identify any differences between the enzymes that might be responsible (as noted in **Chapter 1, Section 1.7.1**, the residues involved in isocitrate binding are generally conserved between IDHs, while the residues involved in NAD(P)$^+$ binding vary).

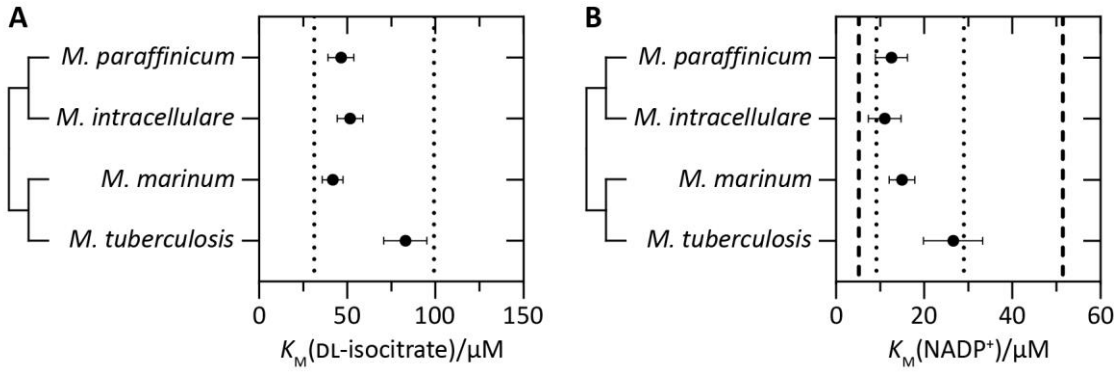


Figure 82: Comparison of K_M(DL-isocitrate) and K_M(NADP$^+$) for four mycobacterial IDH1s.
There is no biologically significant variation in in K_M(DL-isocitrate) or K_M(NADP$^+$) between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). **A – Comparison of K_M(DL-isocitrate).** Values for K_M(DL-isocitrate) are those obtained as described above and presented in **Table 25**. The dotted lines (at 31.4 and 99.2 µM) indicate ±0.25 $\log_{10}$ of the mean (55.8 µM). **B – Comparison of K_M(NADP$^+$).** Values for K_M(NADP$^+$) are those obtained as described above and presented in **Table 26. The** dotted lines (at 9.2 and 29.0 µM) indicate ±0.25 $\log_{10}$ of the mean (16.3 µM). The dashed lines (at 5.2 and 51.5 µM) indicate ±0.5 $\log_{10}$ of the mean. Trees are presented in an unscaled format.

The values of k_{cat} and k_{cat}/K_M obtained by varying the concentration of isocitrate (**Table 25**) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 83A** and **Figure 83B** respectively. Although Mtb IDH1 turns over isocitrate more slowly than the other IDH1s we characterised, k_{cat} for all species falls within 0.25 $\log_{10}$ of the mean, indicating that the variation observed is not biologically significant. As a consequence of its lower activity (**Figure 82A**) and affinity for isocitrate (**Figure 83A**), Mtb IDH1 is less catalytically efficient than the other mycobacterial IDH1s tested (**Figure 83B**). However, the total variation in k_{cat}/K_M with respect to DL-isocitrate falls within a 0.75 $\log_{10}$ range and so is not biologically significant. Overall, therefore, we did not observe any biologically significant variation in mycobacterial IDH1 kinetics.

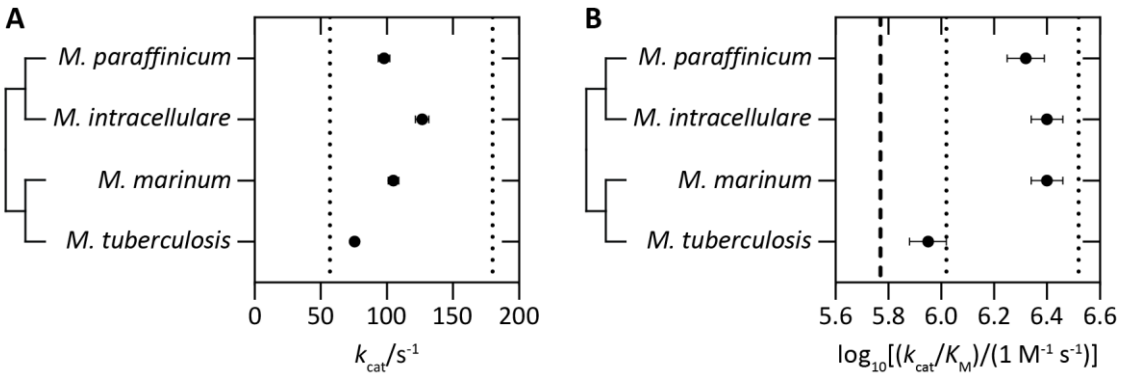


Figure 83: Comparison of k_{cat} and $\log_{10}[k_{cat}/K_M]$ for four mycobacterial IDH1s.
Values for k_{cat} and k_{cat}/K_M are those obtained by varying the concentration of DL-isocitrate, as described above and presented in **Table 25**. There is no biologically significant variation in k_{cat} or $\log_{10}[k_{cat}/K_M]$ between the enzymes tested (a $\log_{10}$ difference between species would be considered biologically significant). **A – Comparison of k_{cat} with respect to DL-isocitrate.** The dotted lines (at 57 and 180 s^{-1}) indicate ±0.25 $\log_{10}$ of the mean (101 s^{-1}). **B – Comparison of $\log_{10}[k_{cat}/K_M]$ with respect to DL-isocitrate.** The dotted lines (at 6.02 and 6.52) indicate ±0.25 $\log_{10}$ of the mean (6.27). The dashed line (at 5.77) marks 0.5 $\log_{10}$ below the mean. Trees are presented in an unscaled format.

5.8. Comparison of the kinetic parameters of the isocitrate branch point in eleven mycobacterial species

Thus far, we have completed basic kinetic characterisation of all isocitrate branch point enzymes from *M. aromaticivorans*, *M. aurum*, *M. chubuense*, *M. fallax*, *M. hiberniae*,

M. intracellulare, *M. marinum*, *M. salmoniphilum*, *M. smegmatis*, *M. terrae*, and Mtb. Having compared the homologues of ICL1 (**Section 5.4.2**), ICL2 (**Section 5.5.2**), IDH2 (**Section 5.6.2**), and IDH1 (**Section 5.7.2**), we can now compare the different isocitrate metabolising enzymes to each other to develop a kinetic picture of the branch point in eleven mycobacterial species.

The K_M(DL-isocitrate) values for 34 enzymes (20 ICLs and 14 IDHs) are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 84**. IDHs have higher affinities for isocitrate than ICLs. In all the species we characterised, the difference in K_M(DL-isocitrate) between IDH2 and ICL1 is biologically significant. On average, K_M(DL-isocitrate) is 25 times smaller for IDH2 than for ICL1 from the same species. With the exception of *M. chubuense* IDH2, IDH2s (grey squares) have higher affinities for isocitrate than IDH1s (black crosses). Each IDH1 has a lower affinity for isocitrate than the IDH2 from the same species. However, the differences in K_M(DL-isocitrate) between the two types of IDH are not biologically significant. With the exception of *M. smegmatis* ICL2, ICL2s (open circles) have lower affinities for isocitrate than ICL1s (black circles). Each ICL2 has a lower affinity for isocitrate than the ICL1 from the same species. K_M(DL-isocitrate) is 1.5—15 times larger for ICL2 than for the corresponding ICL1. The difference in K_M(DL-isocitrate) between the two ICLs is biologically significant in four species: *M. marinum*, *M. intracellulare*, *M. terrae*, and *M. aromaticivorans*. These are the four species with the highest K_M(DL-isocitrate) for ICL2.

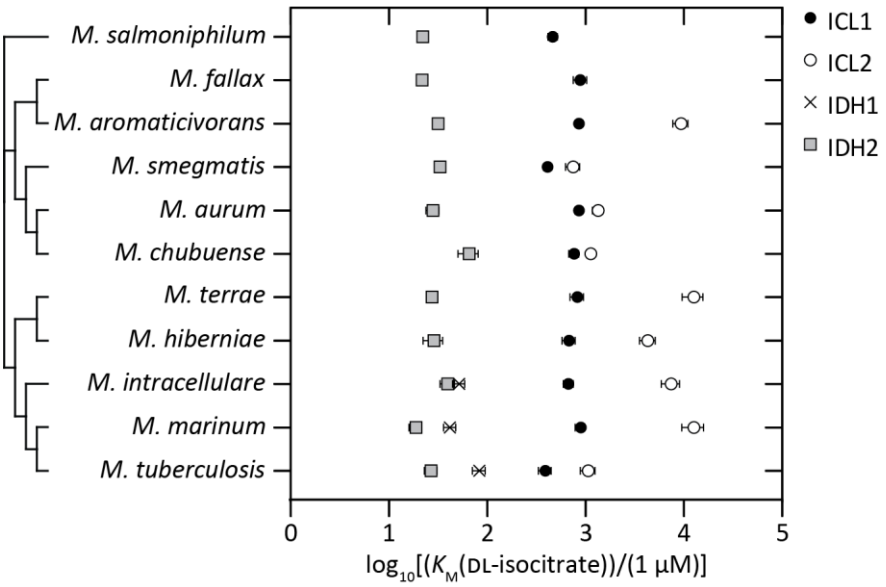


Figure 84: Comparison of K_M(DL-isocitrate) for 34 mycobacterial isocitrate metabolising enzymes. Values for K_M(DL-isocitrate) are those obtained as described above and presented in **Table 21** (ICL1s), **Table 22** (ICL2s), **Table 23** (IDH2s), and **Table 25** (IDH1s). A $\log_{10}$ scale is used to enable visualisation of K_M(DL-isocitrate) values from ICL2s (mM) alongside those from ICL1s, IDH2s, and IDH1s (µM). ICL1s, black circles; ICL2s, open circles; IDH2s, grey squares; IDH1s (crosses). The tree is presented in an unscaled format.

Values of k_{cat} obtained by varying the concentration of isocitrate are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 85**, and show that IDHs turn over isocitrate faster than ICLs. On average, an IDH2 turns over isocitrate around four times faster than the ICL1 from the same species. Depending on the species, turnover of isocitrate by IDH1 may be faster (*M. marinum*), slower (*M. intracellulare*), or comparable (Mtb) to turnover of isocitrate by IDH2. In contrast, each ICL1 turns over isocitrate faster than the ICL2 from the same species. Typically, k_{cat} for ICL1 is 1.1—3.5 times larger than k_{cat} for ICL2. However, in species with particularly slow ICL2s (*M. hiberniae*, *M. terrae*, and Mtb), k_{cat} for ICL1 is 14—20 times larger than k_{cat} for ICL2.

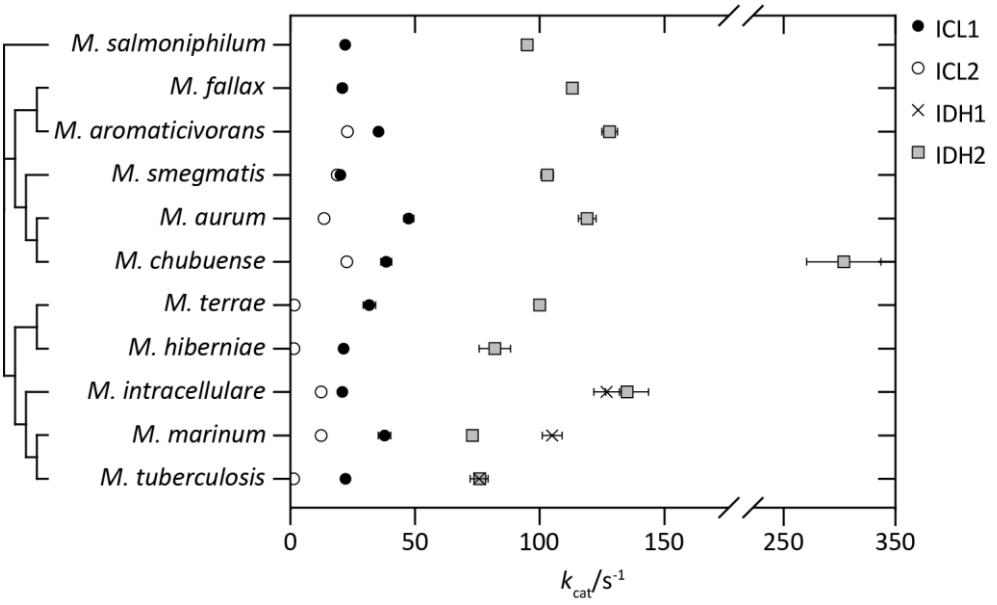


Figure 85: Comparison of k_{cat} for 34 mycobacterial isocitrate metabolising enzymes.
Values for k_{cat} are those obtained as described above and presented in **Table 21** (ICL1s), **Table 22** (ICL2s), **Table 23** (IDH2s), and **Table 25** (IDH1s). ICL1s, black circles; ICL2s, open circles; IDH2s, grey squares; IDH1s (crosses). The tree is presented in an unscaled format.

Finally, the catalytic efficiencies of these enzymes with respect to isocitrate are plotted against the phylogenetic tree of the corresponding mycobacterial species in **Figure 86**. IDHs are more catalytically efficient than ICLs. The difference in k_{cat}/K_M between IDHs and ICLs is biologically significant; in each of the species shown in **Figure 86**, IDH2 is around 100 times more efficient than ICL1. In species with both IDHs, IDH2 is more catalytically efficient than IDH1, but the difference in k_{cat}/K_M between the two enzymes is not biologically significant. In species with both ICLs, ICL1 is more catalytically efficient than ICL2. The difference in k_{cat}/K_M between the two enzymes is biologically significant except in the three species with the most efficient ICL2s (*M. aurum*, *M. chubuense*, and *M. smegmatis*).

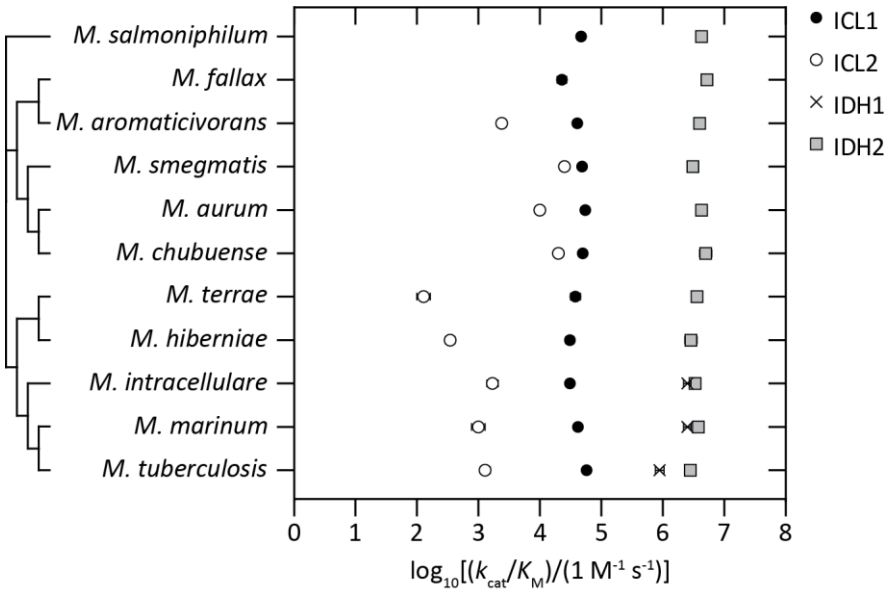


Figure 86: Comparison of $\log_{10}[k_{cat}/K_M]$ for 34 mycobacterial isocitrate metabolising enzymes.
Values for k_{cat}/K_M are those obtained as described above and presented in **Table 21** (ICL1s), **Table 22** (ICL2s), **Table 23** (IDH2s), and **Table 25** (IDH1s). As before, a $\log_{10}$ scale is used to aid the identification of biologically significant variation, considered to be a 1 $\log_{10}$ difference between species. ICL1s, black circles; ICL2s, open circles; IDH2s, grey squares; IDH1s (crosses). The tree is presented in an unscaled format.

5.9. Variation in isocitrate metabolism kinetics is not related to mycobacterial ecology

Earlier, we noted that carbon derived from lipids (thought to be the dominant carbon sources in infected macrophages) enters Mtb central carbon metabolism as Ac-CoA and so is ultimately metabolised through the glyoxylate shunt (**Chapter 1, Section 1.9**).[253,256,435] In contrast, carbon derived from sugars (the most abundant carbon sources in soil) is metabolised through glycolysis, the pentose phosphate pathway, and the TCA cycle.[442] As a result, we hypothesised that mycobacteria adapted to host and non-host niches might exhibit differences in isocitrate metabolism. Specifically, we theorised that the glyoxylate shunt might be more efficient in host-adapted mycobacterial species. In **Chapter 2**, we found that the structure of the isocitrate branch point varies between mycobacterial species, but the presence or absence of individual isocitrate metabolising enzymes does not correlate with mycobacterial ecology.

Subsequently, we questioned whether the basic kinetic parameters of isocitrate metabolising enzymes vary between mycobacterial species. Contrary to our hypothesis, kinetic characterisation of 36 mycobacterial isocitrate metabolising enzymes found no biologically significant variation between ICL1s, between ICL2s, or between IDH2s. The biologically significant variation exhibited by ICL2s was phylogenetically rather than ecologically associated.

5.9.1. The absence of biologically significant variation suggests slow evolution and may reflect metabolic essentiality

Evolutionary rates vary widely between enzymes and are influenced by a range of factors, most importantly abundance.[527] Enzymes that are highly expressed (on either the mRNA or the protein level) evolve more slowly than those that are not.[527,528] Analysing the rate of protein evolution in *E. coli* in the context of its genome-scale metabolic network shows that the evolutionary rate of an enzyme is also affected by its role in metabolism. Specifically, the metabolic flux burden and superessentiality index of an enzyme influence its rate of evolution as strongly as its mRNA expression level.[528] Enzymes that catalyse high flux reactions evolve more slowly than enzymes that catalyse low flux reactions.[528,529] Superessentiality, meanwhile, is a measure of how difficult it is to bypass the metabolic function of an enzyme.[530] Enzymes which are more difficult to bypass evolve more slowly than those whose absence does not affect bacterial viability.[528]

The lack of biologically significant variation in IDH2 and IDH1 kinetics suggests that mycobacterial species in both host and non-host niches experience evolutionary pressure to retain effective IDH catalysis. Strong purifying selection against amino acid

replacement has also been reported for the IDHs from strains of *E. coli* and *S. enterica*.[531] The observed selection pressure likely does not result from superessentiality: IDHs, like other TCA cycle enzymes, have low superessentiality indices.[530] Indeed, transposon mutagenesis studies in Mtb H37Rv indicate that *icd1* is non-essential, while *icd2* essentiality is medium-dependent.[532,533] Slow IDH evolution must therefore result from another feature of these enzymes, such as their metabolic position or flux burden.

The concept of superessentiality might, however, explain the absence of biologically significant variation in ICL1 kinetics. As previously discussed (**Chapter 1, Section 1.5.2**), glyoxylate shunt activity is required for growth on any carbon source that enters central metabolism as acetate.[261] In addition to lipids, such carbon sources include low molecular weight organic acids and various aliphatic and aromatic pollutants, carbon sources which can be found in soil (**Chapter 1, Sections 1.9.2.3 and 1.9.2.5**).[445,460] Although glyoxylate shunt substrates (carbon sources that produce acetate) are typically not the most abundant carbon sources in soil, mycobacteria can co-catabolise carbon substrates and many mycobacterial species isolated from soil have been shown to metabolise hydrocarbons and other organic pollutants (**Chapter 1, Sections 1.9.2.5 and 1.9.3**).[470,471] There may therefore be a routine requirement for ICL activity during mycobacterial growth in soil. Such a requirement could mean that, rather than evolutionary pressure to improve ICL1 kinetics being exerted during host adaptation as we originally supposed, pressure to maintain ICL1 kinetics is exerted in all niches. This scenario would slow the evolution of ICL1, resulting in the lack of kinetic variability we observed between ICL1 homologues.

Finally, we observed that there is biologically significant variation in ICL2 kinetics, but that this variation is phylogenetic. The most catalytically efficient ICL2s do not belong to host-adapted species, suggesting that mycobacterial species are not under pressure to improve the catalytic efficiency of ICL2 with respect to isocitrate during host adaptation. This result is in accordance with the view that ICL2 is less important than ICL1 in mycobacterial isocitrate metabolism (**Chapter 1, Section 1.6.3**).[276,302] It is also possible that any selection pressure on ICL2 relates to a function other than isocitrate lysis, for example its proposed role as a regulator of flux through the glyoxylate shunt.[305]

5.9.2. Understanding the contributions of individual enzymes to isocitrate metabolism will require a greater knowledge of conditions *in bacterio*

Above, we have compared kinetic parameters obtained under standardised assay conditions. When contemplating what these parameters might tell us about flux through the isocitrate branch point, it is critical to recognise that *in vitro* isocitrate metabolism kinetics will not directly reflect *in bacterio* enzyme activity. Firstly, physiological interpretation of kinetic parameters requires knowledge of the substrate concentration, and isocitrate pool sizes have yet to be quantified in mycobacteria.[534,535] Secondly, the operation of regulatory mechanisms will alter the concentrations and/or kinetic parameters of the enzymes *in bacterio*. As noted earlier (**Chapter 1, Section 1.8**), these mechanisms have yet to be fully elucidated in Mtb or *M. smegmatis*, the species in which regulation of the isocitrate branch point has been studied, and they remain uncharacterised in other mycobacterial species. Our assays were conducted in the absence of possible small molecule modulators such as Ac-CoA, and did not include post-translationally modified enzymes. The concentrations of individual isocitrate

metabolising enzymes *in bacterio* are also largely unknown. Accordingly, to connect the results presented in this chapter to *in bacterio* flux, we would need to add new layers to our comparative study of mycobacterial isocitrate metabolism. Suggested future experiments will be presented in **Chapter 6**.

5.10. Chapter conclusion

In this chapter, we selected eleven mycobacterial species for comparison to Mtb CDC 1551: *M. aromaticivorans*, *M. aurum*, *M. chubuense*, *M. fallax*, *M. hiberniae*, *M. intracellulare*, *M. marinum*, *M. paraffinicum*, *M. salmoniphilum*, *M. smegmatis*, and *M. terrae*. These species were chosen from the 2018 Mycobacterial Species Library based on their phylogenetic positions, ecological profiles, and isocitrate branch point structures. Using the methods developed for the Mtb enzymes in **Chapter 4**, we successfully cloned, expressed, purified, and kinetically characterised 32 of the 34 isocitrate metabolising enzymes that comprise the branch points in these species.

Comparison of these enzymes to each other and to the Mtb enzymes characterised in **Chapter 4** identified that there is no biologically significant variation between ICL1s, between IDH2s, or between IDH1s. There are biologically significant differences in the kinetic parameters of ICL2 homologues; these differences are phylogenetically rather than ecologically associated. Contrary to our hypothesis, therefore, mycobacterial species adapted to different ecological niches do not exhibit biologically significant differences in isocitrate metabolism kinetics. The finding that kinetic changes in isocitrate metabolism (and particularly in ICL1) are not associated with mycobacterial host adaptation may reflect the metabolic demands exerted on the isocitrate branch point by non-host niches.

Chapter 6. Conclusions and future work

6.1. Summary and conclusions

6.1.1. Research question: do mycobacteria adapted to nutritionally distinct ecological niches exhibit differences in isocitrate metabolism?

In mycobacteria, as in many other bacterial genera, isocitrate sits at a branch point in central carbon metabolism, between the TCA cycle and the glyoxylate shunt (**Chapter 1, Section 1.5**). Host and non-host niches exert different demands on this branch point as a consequence of their distinct organic carbon content (**Chapter 1, Section 1.9**). During infection, Mtb must divert flux through the glyoxylate shunt in order to process the Ac-CoA produced by the metabolism of fatty acids and cholesterol (**Chapter 1, Section 1.5**). In contrast, the glyoxylate shunt is dispensable for growth on TCA cycle substrates such as glucose, which are the dominant carbon sources in soil (**Chapter 1, Section 1.9.2**). We therefore hypothesised that changes in isocitrate metabolism might occur during mycobacterial niche adaptation. In particular, host-adapted mycobacterial species may have evolved a more efficient glyoxylate shunt as a result of the selection pressure exerted by the host niche (**Chapter 1, Section 1.9.3**).

Four distinct enzymes contribute to flux through the isocitrate branch point in Mtb: ICL1, ICL2, IDH1, and IDH2 (**Chapter 1, Sections 1.6 and 1.7**). In *M. smegmatis*, meanwhile, the isocitrate branch point consists of three enzymes: ICL1, ICL2, and IDH2. At the start of this project, little was known about the ICLs and IDHs from other mycobacterial species (**Chapter 1, Sections 1.6.4 and 1.7.4**). We therefore set out to characterise isocitrate metabolising enzymes from across the genus *Mycobacterium* to evaluate whether the structure and kinetic function of the isocitrate branch point vary between species adapted to nutritionally distinct ecological niches.

6.1.2. A genus-level comparative study of mycobacterial isocitrate metabolising enzymes: key results

In the preceding chapters, we combined bioinformatics and genus-level comparative enzymology to explore how the structure and kinetic function of the isocitrate branch point varies between mycobacterial species. First, we **established how the composition of the isocitrate branch point varies across the genus *Mycobacterium*** (**Chapter 2**). We found that mycobacterial species have between two and four enzymes for the metabolism of isocitrate, and that variation in isocitrate branch point composition is phylogenetically associated. Four mycobacterial isocitrate branch point structures were identified, each including ICL1 and at least one IDH.

Second, we **determined that the frameshift mutation in Mtb *aceA* that creates *rv1915—1916* is not restricted to the laboratory-adapted strains H37Rv and H37Ra** (**Chapter 3**). BLAST searches showed that at least 1000 Mtb strains possess a frameshifted *aceA*, indicating that the p.(Tyr296ThrfsTer73) mutation did not occur during the laboratory adaptation of H37Rv and H37Ra. To explore the evolutionary origins of frameshifted *aceA*, we performed BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment for 3825 Mtb strains from the NCBI Microbial Genomes RefSeq database. Analysis of the distribution of *aceA* genotypes across the MTBC lineages indicated that the frameshift mutation in *aceA* occurred after the establishment of the Euro-American lineage. However, spoligotype-based lineage assignment offered insufficient phylogenetic resolution to pinpoint the evolutionary origin of frameshifted *aceA*.

Third, we **established methods for the preparation and kinetic characterisation of mycobacterial isocitrate metabolising enzymes in our laboratory** (**Chapter 4**). We expressed the four isocitrate metabolising enzymes from Mtb CDC 1551 in *E. coli* BL21(DE3)Star, developed methods for their purification and implemented assays to facilitate their kinetic characterisation. Ultimately, all four enzymes – His_6-ICL1, ICL2-His_6FLAG, IDH2-His_6FLAG and IDH1-His_6FLAG – were purified in sufficient yields for steady-state kinetics, allowing their kinetic parameters to be determined. We also confirmed that ICL2 is activated by Ac-CoA and, using a truncated form of the enzyme, corroborated the crystallographic finding that this activation involves binding of Ac-CoA to the *C*-terminal domain of ICL2.[305]

Finally, we **cloned, expressed, purified and kinetically characterised the isocitrate metabolising enzymes from a representative subset of the Mycobacterial Species Library** (**Chapter 5**). Eleven additional mycobacterial species were chosen from the 2018 Mycobacterial Species Library based on their phylogenetic positions, ecological profiles, and isocitrate branch point structures. The selected species were *M. aromaticivorans*, *M. aurum*, *M. chubuense*, *M. fallax*, *M. hiberniae*, *M. intracellulare*, *M. marinum*, *M. paraffinicum*, *M. salmoniphilum*, *M. smegmatis*, and *M. terrae*. Using the methods developed in **Chapter 4**, we successfully cloned, expressed, purified, and kinetically characterised 32 of the 34 isocitrate metabolising enzymes that comprise the branch points in these species. Comparison of these enzymes to each other and to the Mtb enzymes characterised in **Chapter 4** identified that there is no biologically significant variation between ICL1s, between IDH2s, or between IDH1s. There are biologically significant differences in the kinetic parameters of ICL2s; these differences are

phylogenetically rather than ecologically associated. Mycobacterial species adapted to different ecological niches do not exhibit biologically significant differences in isocitrate metabolism kinetics.

6.1.3. Conclusion: the composition and kinetics of the mycobacterial isocitrate branch point reflect phylogeny rather than ecology

In conclusion, our results suggest that **changes in the structure and kinetic function of the isocitrate branch point do not occur as part of mycobacterial niche adaptation**. The variation in mycobacterial isocitrate branch point structures is phylogenetic, while kinetic variation is either phylogenetic (ICL2s) or not biologically significant (ICL1s, IDH2s, and IDH1s). As discussed in **Chapter 5, Section 5.9**, the finding that variation in mycobacterial isocitrate metabolism is not associated with host adaptation may reflect the metabolic demands exerted on the isocitrate branch point by non-host niches. We suggest that rather than evolutionary pressure to improve ICL1 kinetics being exerted during host adaptation as we originally supposed, pressure to maintain ICL1 kinetics is exerted in all niches, and highly host-adapted species such as Mtb exploit enzymatic activity that is available across the genus.

6.2. Questions arising and future directions

Thus far, we have offered only an evolutionary interpretation of our results. As discussed in **Chapter 5, Section 5.9.2**, a physiological interpretation will require us to develop our understanding of conditions *in bacterio* by adding microbiological experiments to our comparative study. During the time we have been working on this project, other projects in our laboratory have delivered data and methodology

necessary for the completion of such experiments.[536] We finish by briefly considering three avenues for further investigation.

6.2.1. What are the concentrations of isocitrate metabolising enzymes and relevant metabolites *in bacterio*? Do they change in response to external conditions?

As noted earlier (**Chapter 5, Section 5.9.2**), *in vitro* isocitrate metabolism kinetics will not directly reflect *in bacterio* enzyme activity. To understand how individual enzymes contribute to isocitrate metabolism, we must learn more about the conditions under which they operate *in bacterio*. A useful starting point would be the quantification of isocitrate metabolising enzymes and relevant metabolites in mycobacterial species exposed to a range of external conditions (for example different carbon sources). Determination of the concentrations of isocitrate and Ac-CoA will be particularly important. As previously explained, an understanding of isocitrate pool sizes will be required for the physiological interpretation of the kinetic parameters we have obtained (**Chapter 5, Section 5.9.2**). Quantification of Ac-CoA (and other potential regulatory metabolites) will be critical if we are to evaluate the *in bacterio* relevance of modulatory interactions identified *in vitro* (**Chapter 1, Section 1.8.3.2**).

An established regulatory mechanism for the Mtb isocitrate branch point is control of the concentration of ICL1 (**Chapter 1, Section 1.8**). Measuring the concentrations of isocitrate metabolising enzymes in different mycobacterial species (on the mRNA level using qPCR and/or on the protein level using Western blotting) would indicate whether modulation of enzyme concentrations is a general mechanism for regulation of flux through the isocitrate branch point in mycobacteria. As well as identifying regulatory

mechanisms for the isocitrate branch point, the data supplied by these experiments could be used to inform *in silico* models of mycobacterial isocitrate metabolism.

6.2.2. What is the role of ICL2 *in bacterio*?

The role of ICL2 *in bacterio* remains unclear. As discussed in **Section 6.2.1**, quantification of Ac-CoA under a range of conditions will be required to evaluate whether allosteric modulation of ICL2 activity by Ac-CoA contributes to flux control at the isocitrate branch point or whether the Ac-CoA-bound, activated state is the usual form of ICL2 *in bacterio*. It will also be necessary to determine whether other mycobacterial ICL2s are similarly activated by Ac-CoA. Ascertaining the *in bacterio* consequences of the frameshift mutation that creates *rv1915—1916* from *aceA* in some Euro-American strains of Mtb (**Chapter 3**) may also be of use in establishing the function and importance of ICL2.

6.2.3. Do mycobacteria with different ecological profiles exhibit different growth phenotypes on single carbon sources?

In **Chapter 5, Section 5.9.1**, we suggested that the absence of biologically significant variation in ICL1 kinetics could indicate a routine requirement for ICL activity during mycobacterial growth in soil. Metabolism of the glyoxylate shunt substrates available in soil, for example low molecular weight organic acids, would create such a requirement. Screening mycobacterial species for growth on a range of carbon sources relevant to host and non-host niches would highlight phenotypes that merit further investigation. Subsequent experiments could indicate whether the *in bacterio* functionality of the isocitrate branch point differs between mycobacterial species and/or identify pathways of interest for future genus-level studies of mycobacterial metabolism.

6.3. Final remarks

Herein, we have presented a genus-level comparative study of mycobacterial isocitrate metabolising enzymes. These experiments, which include the first detailed examination of isocitrate metabolism in non-tuberculous mycobacteria, illustrate our approach to elucidating the molecular mechanisms underlying mycobacterial host adaptation. By studying mycobacteria at the genus level, comparing species adapted to different ecological niches, we hope to understand why some mycobacterial species cause disease while others remain innocuous. Answering this question will enable the development of new strategies to combat existing mycobacterial pathogens and the identification of mycobacterial species at risk of becoming clinically relevant in the future, activities which will be essential for addressing the burden of current and emerging mycobacterial diseases.[157–160] We must begin by ending TB: a disease of man since antiquity and a thirty-year global health emergency.

Chapter 7. Computational Methods

7.1. Identification of mycobacterial isocitrate lyases and isocitrate dehydrogenases

The isocitrate lyases (ICLs) and isocitrate dehydrogenases (IDHs) from Mtb and *M. smegmatis* were identified by searching the MycoBrowser portal.[537] Protein sequences from these species were used as query sequences to search the UniProt Knowledgebase (UniProtKB) with the Basic Local Alignment Search Tool (BLAST).[538,539] The search parameters used are provided in **Table 27** (ICLs) and **Table 28** (IDHs).

Table 27: BLAST search parameters used to identify mycobacterial ICL sequences in UniProtKB.

	Isocitrate lyase 1 (ICL1)	**Isocitrate lyase 2 (ICL2)**
Gene identifier: *M. tuberculosis* H37Rv	*rv0467*	*rv1915—1916*
Gene identifier: *M. smegmatis* MC^{2}155	MSMEG_0911	MSMEG_3706
Query sequence	MSMEG_0911	MSMEG_3706
Program	blastp (BLASTP 2.7.1+)	blastp (BLASTP 2.7.1+)
Database	uniprotkb (Protein) Generated for BLAST on 13/07/2018	uniprotkb (Protein) Generated for BLAST on 10/08/2018
Maximum hits	250	500
Results	249 hits	500 hits

Table 28: BLAST search parameters used to identify mycobacterial IDH sequences in UniProtKB.

	Isocitrate dehydrogenase 1 (IDH1)	**Isocitrate dehydrogenase 2 (IDH2)**
Gene identifier: *M. tuberculosis* H37Rv	*rv3339c*	*rv0066c*
Gene identifier: *M. smegmatis* MC^{2}155	absent	MSMEG_1654
Query sequence	Rv3339c	MSMEG_1654
Program	blastp (BLASTP 2.7.1+)	blastp (BLASTP 2.7.1+)
Database	uniprotkb (Protein) Generated for BLAST on 10/08/2018	uniprotkb (Protein) Generated for BLAST on 10/08/2018
Maximum hits	500	500
Results	500 hits	500 hits

The results were downloaded as FASTA files. Non-mycobacterial sequences were removed manually, either before or after downloading the results. The species names were copied to the start of the sequence headers using regular expressions in Crimson Editor. Next, the results were opened in Jalview and sorted alphabetically.[540] Unidentified species (headers beginning sp.) were removed from the list. Where there was more than one sequence for a species, the number of sequences was reduced to one by applying a redundancy threshold of 100%. During the analysis of ICL2s, any sequence shorter than 700 amino acids was removed, as such sequences are too short to be full length ICL2s. The list of species containing each protein was cross-referenced with the list of 129 species used by Acely Garza-Garcia to construct our laboratory's mycobacterial phylogenetic tree (2018 edition) (**Table 29**).

Table 29: The 129 mycobacterial species in which the isocitrate branch point composition was analysed.

M. abscessus	*M. goodii*	*M. paraense*
M. algericum	*M. gordonae*	*M. paraffinicum*
M. alsense	*M. haemophilum*	*M. parafortuitum*
M. angelicum	*M. hassiacum*	*M. parascrofulaceum*
M. aromaticivorans	*M. heckeshornense*	*M. parmense*
M. arosiense	*M. heidelbergense*	*M. peregrinum*
M. arupense	*M. heraklionense*	*M. phlei*
M. asiaticum	*M. hiberniae*	*M. porcinum*
M. aurum	*M. holsaticum*	*M. rhodesiae* 44223
M. avium	*M. houstonense*	*M. rhodesiae* JS60
M. bacteremicum	*M. immunogenum*	*M. rhodesiae* NBB3
M. boenickei	*M. insubricum*	*M. riyadhense*
M. bohemicum	*M. interjectum*	*M. rufum*
M. branderi	*M. intermedium*	*M. rutilum*
M. brisbanense	*M. intracellulare*	*M. salmoniphilum*
M. brumae	*M. iranicum*	*M. saopaulense*
M. canariasense	*M. kansasii*	*M. saskatchewanense*
M. canettii	*M. koreense*	*M. scrofulaceum*
M. celatum	*M. kubicae*	*M. senegalense*
M. celeriflavum	*M. kumamotonense*	*M. senuense*
M. chelonae	*M. kyorinense*	*M. septicum*
M. chimaera	*M. lacus*	*M. setense*
M. chlorophenolicum	*M. lentiflavum*	*M. sherrisii*

M. chubuense	*M. leprae*	*M. shimoidei*
M. colombiense	*M. lepromatosis*	*M. shinjukuense*
M. conceptionense	*M. liflandii*	*M. sinense*
M. confluentis	*M. litorale*	*M. smegmatis*
M. conspicuum	*M. llatzerense*	*M. sphagni*
M. cosmeticum	*M. longobardum*	*M. szulgai*
M. diernhoferi	*M. mageritense*	*M. terrae*
M. doricum	*M. malmoense*	*M. thermoresistibile*
M. elephantis	*M. mantenii*	*M. triplex*
M. engbaekii	*M. marinum*	*M. triviale*
M. sp. EPa45	*M. monacense*	*M. tuberculosis* H37Rv
M. europaeum	*M. moriokaense*	*M. tusciae*
M. fallax	*M. mucogenicum*	*M. ulcerans*
M. farcinogenes	*M. nebraskense*	*M. vaccae*
M. flavescens	*M. neworleansense*	*M. vanbaalenii*
M. florentinum	*M. nonchromogenicum*	*M. vulneris* 45247
M. fragae	*M. noviomagense*	*M. vulneris* PRJEB5750
M. gastri	*M. novocastrense*	*M. wolinskyi*
M. genavense	*M. obuense*	*M. xenopi*
M. gilvum	*M. palustre*	*M. yongonense*

Missing species were investigated by using the protein-nucleotide 6-frame translation (tblastn) algorithm to search the National Center for Biotechnology Information (NCBI) RefSeq Representative Genomes and Nucleotide Collection (nr/nt) databases.[541] The query sequences used were the same as for the UniProtKB searches. Where a species

was identified as possessing a protein, the protein sequence was retrieved using its GenBank protein ID.[542] The individually identified sequences were added to the list of sequences obtained from UniProtKB. The completed analysis was visualised by colouring our mycobacterial phylogenetic tree.

7.2. Identification of Mtb strains encoding frameshifted and single frame *aceA* using web-based BLAST searches

The DNA sequences for H37Rv *rv1915—1916* (frameshifted) and CDC 1551 *aceA* (single frame) were obtained and translated in a single frame using ExPASy Translate.[543] The resulting protein sequences (given in **Table 30** and **Table 31**) were used to query the NCBI Mtb RefSeq Genomes database (tblastn on the NCBI web service, using the parameters provided in **Table 32**).[541] These queries returned a list of 1000 Mtb strains with frameshifted (H37Rv-like) *aceA* and 1000 Mtb strains with single frame (CDC 1551-like) *aceA*.

Table 30: Query sequence used to identify frameshifted *aceA*.

Frameshifted *aceA*	Mtb H37Rv; single frame translation of *rv1915—1916*
Accession	5’3’ Frame 1 translation of NC_000962.3:2160463—2162762

MAIAETDTEVHTPFEQDFEKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAA

GAFYERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTEDPGPDLASYPL

SQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAYDFRPFIIADAGTGHGGDPHVRNLI

RRFVEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAE

AANLIDSRADERDQPFLLGATKLDVPSYKSCFLAMVRRFTNWASRSSMVIFSMRLATASTRRPA

VGLSAKAFSAWSPTRSTRGGRTASSRSTAFSTRSSRGSWRPGRTTRA**PTERPWRTCSNSVR

ARANPLAWLPRSGGRSPRVHRCMPPGQRPRSWAPIRHGTASWPRPRRATTRSAAAYRMRSP

NRWPRHRLPTFFGWRPRPPISPTLDSSPRRSMPSSPTRCWRTTSHHRSTGTPPA*PTRRCGAS

PRSSAKWASSSTSSPMAGTRSTVSRPRNSPPRCARTACWRWLGCSARCAWSNLPIAHRKR*S

AGRAVTPHWLPPPDARRPRRQWARAPPSTSTWCKLRCRASC*RNGWPCGAVTTSSKTNCAY

SFGRSGPARRCSSSASTAKAMTSSPT*YSNRSKIAAAAPSCWYATRTRSVRNYAKSG**P*STSG

SSTASRRRRCTTSRPPTTTSTRPRR*SRMESSPRSTRRWARSSSPR*TTRASPNC*RPIGWRCGS

*SRRRR

Table 31: Query sequence used to identify single frame *aceA*.

Single-frame *aceA*	Mtb CDC 1551
Accession	WP_003409584.1
MAIAETDTEVHTPFEQDFEKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAA GAFYERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTEDPGPDLASYPL SQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAYDFRPFIIADADTGHGGDPHVRNLI RRFVEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAE AANLIDSRADERDQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDSEYAAAG GWLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGLMTYGEAVADVLEF GQSEGEPIGMAPEEWRAFAARASLHAARAKAKELGADPPWDCELAKTPEGYYQIRGGIPYAIA KSLAAAPFADILWMETKTADLADARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMR RFPEELGKMGFVFNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQTL VGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPRKLLEEWLAMWSGHYQLKDKLR VQLRPQRAGSEVLELGIHGESDDKLANVIFQPIQDRRGRTILLVRDQNTFGAELRQKRLMTLIHL WLVHRFKAQAVHYVTPTDDNLYQTSKMKSHGIFTEVNQEVGEIIVAEVNHPRIAELLTPDRVAL RKLITKEA	

Table 32: Query parameters for the web-based BLAST searches used to identify *aceA*s.

	Frameshifted *aceA*	**Single frame *aceA***
Query sequence	Frameshifted *aceA*, **Table 30**	Single frame *aceA*, **Table 31**
Program	tblastn (TBLASTN 2.8.1+)	tblastn (TBLASTN 2.8.1+)
Database	refseq_genomes	refseq_genomes
Date	04/10/2018	04/10/2018
Maximum hits	1000	1000
Number of hits	1000	1000

7.3. BLAST-based *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment of Mtb strains from the NCBI Microbial Genomes RefSeq database

7.3.1. Identification and curation of Mtb genomes

On 27/03/2020, the NCBI Microbial Genomes RefSeq database contained 6705 assemblies belonging to the Organism Group "Mycobacterium tuberculosis". 6597 of these assemblies were available for download and were downloaded as Genomic FASTA (.fna) files from the RefSeq database via NCBI Assembly (ncbi.nlm.nih.gov/assembly).

6604 assembly statistics files belonging to the Organism Group "Mycobacterium tuberculosis" were downloaded as text files from the RefSeq database via NCBI Assembly on 16/04/2020. The assembly statistics files were parsed to determine the total number of bases missing from each assembly (termed the total gap length). To maximise the size of the dataset while lowering the likelihood of errors resulting from missing bases, 500 bases was selected as the maximum allowable total gap length. 3832 assemblies had a total gap length less than or equal to 500 bases and were considered suitable for further analysis.

3825 of the suitable assemblies were successfully matched to downloaded genomes and subjected to *aceA* genotyping, *in silico* spoligotyping, and spoligotype-based predictive lineage assignment. These assemblies will be referred to as 'assemblies in the curated list' for the rest of this section.

7.3.2. *AceA* genotyping

AceA genotyping was performed in the Windows 10 PowerShell using blast 2.10.0+ and Python 3.7. For each assembly in the curated list, a BLAST database was created and tblastn queries were used to assign the *aceA* genotype. In each case, the database was first queried using the frameshifted *aceA* query sequence (**Table 30**). If the output parameters matched those given in **Table 33**, the genotype was assigned as frameshifted. If the output parameters did not match, the BLAST database was queried using the single frame *aceA* query sequence (**Table 30**). If the output parameters matched those given in **Table 33**, the genotype was assigned as single frame. If neither BLAST search gave matching output parameters, the genotype was categorised as undefined.

For each genotype, two sets of output parameters were defined: SNP-intolerant (strict) and SNP-tolerant (relaxed). When the relaxed parameters were applied, the output parameters were first checked against the strict parameters and then if they did not match, checked against the relaxed parameters before decision-making. This was done to increase the number of genomes for which the *aceA* genotype could be assigned.

Table 33: BLAST output parameters used to assign *aceA* sequences as frameshifted or single frame. Appropriate BLAST output parameters were defined by considering the results of the web-based BLAST searches described in **Section 7.2** and tested using genomes for which the *aceA* genotype was known. The strict parameters did not allow for SNPs in *aceA*. The relaxed parameters allowed for SNPs in *aceA*.

Query sequence	**Frameshifted *aceA***		**Single frame *aceA***	
SNP-tolerance	strict	relaxed	strict	relaxed
result_bitscore	1078	-	1541	-
result_evalue	0	0	0	0
result_pident	100	-	100	-
result_ppos	100	-	100	-
result_gaps	0	0	0	0
result_alignment_length	-	680	-	766
***AceA* genotype**	**frameshifted**		**single frame**	

7.3.3. *In silico* spoligotyping

In silico spoligotyping was performed in the Windows 10 PowerShell using blast 2.10.0+ and Python 3.7.

The *in silico* spoligotyping program SpoTyping.py was downloaded from GitHub: (https://raw.githubusercontent.com/xiaeryu/SpoTyping/663bce731cfdb1512dc44d73e6e6c419f8cc231a/SpoTyping-v3.0-commandLine/SpoTyping.py).[492] The parse_blast function, which constructs spoligotypes from BLAST output data, was extracted and modified slightly to suit the input files (.fna files containing one or more FASTA headers). Specifically, the variables min_strict and min_relax were both replaced with 1. The required spacer query sequences (spacer.fasta) were also downloaded from GitHub (https://github.com/xiaeryu/SpoTyping/blob/master/SpoTyping-v2.1-commandLine/ref/spacer.fasta) and are given in **Table 34**.[492]

For each assembly in the curated list, a BLAST database was created and queried using the spacer sequences in **Table 34** (blastn, using the same parameters as in SpoTyping). The parse_blast function from SpoTyping was then used to construct the octal spoligotype from the BLAST output data.[492]

Table 34: Spacer sequences used for *in silico* spoligotyping.
(https://github.com/xiaeryu/SpoTyping/blob/master/SpoTyping-v2.1-commandLine/ref/spacer.fasta)

1	ATAGAGGGTCGCCGGCTCTGGATCA	23	AGCATCGCTGATGCGGTCCAGCTCG
2	CCTCATGCTTGGGCGACAGCTTTTG	24	CCGCCTGCTGGGTGAGACGTGCTCG
3	CCGTGCTTCCAGTGATCGCCTTCTA	25	GATCAGCGACCACCGCACCCTGTCA
4	ACGTCATACGCCGACCAATCATCAG	26	CTTCAGCACCACCATCATCCGGCGC
5	TTTTCTGACCACTTGTGCGGGATTA	27	GGATTCGTGATCTCTTCCCGCGGAT
6	CGTCGTCATTTCCGGCTTCAATTTC	28	TGCCCCGGCGTTTAGCGATCACAAC
7	GAGGAGAGCGAGTACTCGGGGCTGC	29	AAATACAGGCTCCACGACACGACCA
8	CGTGAAACCGCCCCCAGCCTCGCCG	30	GGTTGCCCCGCGCCCTTTTCCAGCC
9	ACTCGGAATCCCATGTGCTGACAGC	31	TCAGACAGGTTCGCGTCGATCAAGT
10	TCGACACCCGCTCTAGTTGACTTCC	32	GACCAAATAGGTATCGGCGTGTTCA
11	GTGAGCAACGGCGGCGGCAACCTGG	33	GACATGACGGCGGTGCCGCACTTGA
12	ATATCTGCTGCCCGCCCGGGGAGAT	34	AAGTCACCTCGCCCACACCGTCGAA
13	GACCATCATTGCCATTCCCTCTCCC	35	TCCGTACGCTCGAAACGCTTCCAAC
14	GGTGTGATGCGGATGGTCGGCTCGG	36	CGAAATCCAGCACCACATCCGCAGC
15	CTTGAATAACGCGCAGTGAATTTCG	37	CGCGAACTCGTCCACAGTCCCCCTT
16	CGAGTTCCCGTCAGCGTCGTAAATC	38	CGTGGATGGCGGATGCGTTGTGCGC
17	GCGCCGGCCCGCGCGGATGACTCCG	39	GACGATGGCCAGTAAATCGGCGTGG
18	CATGGACCCGGGCGAGCTGCAGATG	40	CGCCATCTGTGCCTCATACAGGTCC
19	TAACTGGCTTGGCGCTGATCCTGGT	41	GGAGCTTTCCGGCTTCTATCAGGTA
20	TTGACCTCGCCAGGAGAGAAGATCA	42	ATGGTGGGACATGGACGAGCGCGAC
21	TCGATGTCGATGTCCCAATCGTCGA	43	CGCAGAATCGCACCGGGTGCGGGAG
22	ACCGCAGACGGCACGATTGAGACAA		

7.3.4. Spoligotype-based predictive lineage assignment

7.3.4.1. TB-Lineage queries

Major lineages were predicted from spoligotypes using the Rules and CBN functions of TB-Lineage (https://tbinsight.cs.rpi.edu/run_tb_lineage.html).[499,500] Queries were prepared as text files according to the directions provided and consisted of the NCBI RefSeq accession code (as strain identifier) and octal spoligotype (obtained in **Section 7.3.3**) for each strain. Seven batch queries were run (20/04/2020) in order to accommodate the size of the dataset. Output data (Rules, major lineage; CBN, major lineage and probability) were downloaded as Excel 2003 files.

7.3.4.2. Assignment of consensus lineages

The criteria used for consensus lineage assignment are presented in **Table 35**. In most cases, the criteria for confident assignment were that Rules and CBN returned the same lineage, and that the CBN probability was >0.995. The exceptions were the East African-Indian lineage (strains designated as unknown by Rules, but East African-Indian by CBN with a probability >0.995) and *M. caprae* (strains designated *M. bovis* by Rules, but *M. caprae* by CBN with a probability >0.995).

Table 35: Criteria used for consensus MTBC lineage assignment.[499,500]

Lineage		Rules output	CBN output	CBN probability
1	Indo-Oceanic	Indo-Oceanic	Indo-Oceanic	>0.995
2	East Asian	East Asian	East Asian	>0.995
3	East African-Indian	Unknown	East African-Indian	>0.995
4	Euro-American	Euro-American	Euro-American	>0.995
5 & 6	*M. africanum*	West African 1 West African 2 *M. africanum*	*M. africanum*	>0.995
Animal	*M. bovis*	*M. bovis*	*M. bovis*	>0.995
Animal	*M. caprae*	*M. bovis*	*M. caprae*	>0.995

7.3.4.3. SITVIT2 queries

Spoligotyping results were used to search SITVIT2, the current version of the international spoligotyping database.[496] The Excel 2003 batch query template file was downloaded from http://www.pasteur-guadeloupe.fr:8081/SITVIT2/batch.jsp and filled with the NCBI RefSeq accession code (as strain identifier) and spoligotype for each strain. Seven batch queries were run (20/04/2020) in order to accommodate the size of the dataset. Output data (Spoligotype International Type, SIT; and Lineage) were downloaded as Excel 2003 files.

7.3.4.4. Assignment of Euro-American strains to spoligotype families

The SITVIT2 spoligotypes obtained through the above queries (**Section 7.3.4.3**) were categorised as described in the literature and summarised in **Table 36**.[495,496]

Table 36: Assignment of SITVIT2 spoligotypes to Euro-American spoligotype families.
Categorisation was performed in accordance with the literature.[495,496]

Spoligotype family	SITVIT2 spoligotypes seen in this study
H	H, H1, H3, Ural-1
LAM	Cameroon, LAM, LAM1, LAM2, LAM3, LAM4, LAM5, LAM6, LAM9, LAM-RUS, Turkey
S	S
T	T, T1, T1-RUS2, T2, T3, T3-ETH, T4, T4-CEU1, T5, T5-Madrid, T-tuscany, T-H37Rv
X	X1, X2, X3
Not lineage 4	AFRI_2
Not classified	Not defined, Unknown

Chapter 8. Experimental Procedures: Biochemistry

8.1. Mycobacterial Species and Strains

The American Type Culture Collection (ATCC) and Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) strain identifiers for the mycobacterial species used in this study are provided in **Table 37**. The containment level for each species is also given. Type strains were used for all species except *M. smegmatis* and *M. tuberculosis*.

Table 37: Mycobacterial species and strains used.
List of the mycobacterial species used in this study, together with their ATCC and DSMZ strain identifiers and their containment level (CL).

Species	Strain identifier		CL
	ATCC	DSMZ	
Mycobacterium aromaticivorans	ATCC BAA-1378	DSM-45407	1
Mycobacterium aurum	ATCC 23366	DSM-43999	1
Mycobacterium chubuense	ATCC 27278	DSM-44219	1
Mycobacterium fallax	ATCC 35219	DSM-44179	1
Mycobacterium hiberniae	ATCC 49874	DSM-44241	1
Mycobacterium intracellulare	ATCC 13950	DSM-43223	2
Mycobacterium marinum	ATCC 927	DSM-44344	2
Mycobacterium paraffinicum	ATCC 12670	DSM-44181	1
Mycobacterium salmoniphilum	ATCC 13758	DSM-43276	1
Mycobacterium smegmatis MC^2155	ATCC 700084	-	1
Mycobacterium terrae	ATCC 15755	DSM-43227	1
Mycobacterium tuberculosis CDC 1551	-	-	3

8.2. Construct Preparation

8.2.1. Template DNA preparation

Genomic DNA from *M. tuberculosis* CDC 1551 was provided by Deborah Hunt; further preparation was not required. *M. smegmatis* was provided by Miha Homsak as colonies on Löwenstein-Jensen medium. The other mycobacterial species were provided by Fernanda Subtil as cell lysates or as colonies on Middlebrook 7H10 agar. Where colonies were provided, template DNA was prepared using InstaGene™ Matrix (Bio-Rad). A single colony was transferred to 1 mL sterile water and the cells were resuspended by vortexing. The suspension was centrifuged at 21,100 × *g* for 1 minute. The supernatant was discarded and 200 µL InstaGene™ Matrix was added to the cell pellet. The resulting suspension was incubated at 56 °C for 30 minutes, then vortexed at high speed for 10 seconds. After incubation at 100 °C for 8 minutes, the suspension was again vortexed at high speed for 10 seconds, then centrifuged at 21,100 × *g* for 3 minutes. The supernatant was used in PCR reactions.

After preparation, lysates were stored at −20 °C. Frozen samples were thawed and centrifuged at 21,100 × *g* for 3 minutes. The supernatant was used in PCR reactions.

8.2.2. Primer Design

M. tuberculosis and *M. smegmatis* gene sequences were obtained from Mycobrowser.[537] For the other species, gene sequences were obtained using tblastn, blasting the appropriate *M. tuberculosis* or *M. smegmatis* protein sequence against the NCBI database.[541] Primers were designed using OligoCalc to give a melting temperature (T_M (Basic)) of approximately 62 °C.[544] Appropriate vector matching sequences were then added to the primer, as shown in **Table 38**. Primers for *icl* amplification contained

matching sequences for the pNIC-NTH vector (for expression of proteins bearing an *N*-terminal His_6 tag). Primers for *aceA*, *icd1*, and *icd2* amplification contained matching sequences for the pNIC-CTHF vector (for expression of proteins bearing a *C*-terminal His_6FLAG tag). Primers were synthesised by Integrated DNA Technologies (IDT). The sequences of the primer pairs designed for each gene for which amplification was attempted in this study are given in **Table 39**.

Table 38: Vector matching sequences used for primer design.

Vector	Vector matching sequence	
	Forward primer	Reverse primer
pNIC-NTH	CGAGAACCTGTACTTCCAATCC	GATCCGTATCCACCTTTACTG
pNIC-CTHF	CCTTAAGAAGGAGATATACT	GATTGGAAGTAGAGGTTCTCTGC

Table 39: Accession codes and primer sequences for the mycobacterial isocitrate metabolism genes included in this study.
Mycobacterial isocitrate metabolism genes from the species of interest were obtained and primers for their amplification were obtained as described above. The primer sequences presented here include the vector matching sequences for assembly into pNIC-NTH or pNIC-CTHF, as appropriate. * Amplification of this gene using the primer sequences presented here was ultimately unsuccessful (for a possible explanation, see **Chapter 5, Section 5.3.1**).

M. aromaticivorans	
icl	NZ_JALN02000001.1:3374829—3376118
	CGAGAACCTGTACTTCCAATCCATGACACTGTCCAACGTTGGCCAGCC
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCG
aceA	NZ_JALN02000001.1:c5036163—5033863
	CCTTAAGAAGGAGATATACTATGTCCATCATCGAAGCGGACTCAGCAAC
	GATTGGAAGTAGAGGTTCTCTGCATCCTCCTTGGCGATAAGCCGTTTCAGC
icd2	NZ_JALN02000001.1:c1429297—1427075
	CCTTAAGAAGGAGATATACTATGAGCGCCGAGAACCCGACCATC
	GATTGGAAGTAGAGGTTCTCTGCGCTGAGCGCCTCGTTGAAGGTCTTG

M. aurum	
icl	NZ_CVQQ01000006.1:c190061—188778
	CGAGAACCTGTACTTCCAATCCATGTCGACCGTTGGCACGCCGAAG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCGA
aceA	NZ_JACKRV010000022.1:c374310—372025
	CCTTAAGAAGGAGATATACTATGGCCATCATCGAAGCTGACCTCCAG
	GATTGGAAGTAGAGGTTCTCTGCGTCCTGCTTGCTGATCAGGCGGGTC
icd2	NZ_CVQQ01000002.1:c298545—296311
	CCTTAAGAAGGAGATATACTATGAGCGCCGAGCAGCCGTCC
	GATTGGAAGTAGAGGTTCTCTGCACTCCGGGCGGCCTCCAGC
M. chubuense	
icl	NC_018027.1:597008—598297
	CGAGAACCTGTACTTCCAATCCATGTCGACCGTGGGCACGCCG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCG
aceA	NZ_JYNX01000024.1:17541—19826
	CCTTAAGAAGGAGATATACTATGGCCATCATCGAAGCCGACCTTCAG
	GATTGGAAGTAGAGGTTCTCTGCGTCCTGCTTGCTGATCAGCCGCAG
icd2	NZ_MVHO01000042.1:c32870—30636
	CCTTAAGAAGGAGATATACTATGAGCGCCGAGCAGCCCACC
	GATTGGAAGTAGAGGTTCTCTGCGCTCGGCAGCGACTCCAGCG
M. fallax	
icl	NZ_LQOJ01000039.1:128557—129852
	CGAGAACCTGTACTTCCAATCCATGGAGTCCACAGTGTCCACCGTTGG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCG
icd2	NZ_LQOJ01000021.1:47311—49551
	CCTTAAGAAGGAGATATACTATGAGCGCTGAGCAGCCGACCATC
	GATTGGAAGTAGAGGTTCTCTGCAGCCTGGGCGGAAGCCCCTTC
M. hiberniae	
icl	NZ_LQOZ01000036.1:173448—174731
	CGAGAACCTGTACTTCCAATCCATGTCGACCGTTGGCACGCCCAAG

	GATCCGTATCCACCTTTACTGTTAGTGGAACTGACCCTCTTCGGTCGAG
aceA	NZ_LQOZ01000022.1:c13203—10906
	CCTTAAGAAGGAGATATACTATGACCACGATCCAGAAGAACCCGACG
	GATTGGAAGTAGAGGTTCTCTGCGTCCGCTTTGCTGATCAGCTTGCCC
icd2	NZ_LQOZ01000021.1:c54147—51916
	CCTTAAGAAGGAGATATACTATGTGCGCCGACCAGCCGACC
	GATTGGAAGTAGAGGTTCTCTGCGTCCTTGACGTCGATCGCAGCGTTG
M. intracellulare	
icl	NC_016946.1:c4780264—4778981
	CGAGAACCTGTACTTCCAATCCATGTCTGTCGTTGGCACCCCCAAGAG
	GATCCGTATCCACCTTTACTGTTAGTGAAACTGCCCCTCTTCGGTGGAG
aceA	NC_016946.1:c2735076—2732791
	CCTTAAGAAGGAGATATACTATGGCGATCATCGACAAGGACACACAAGTC
	GATTGGAAGTAGAGGTTCTCTGCCCCCTCCTTGGTGATCAGCTTGCG
icd1	NC_016946.1:c4399863—4398625
	CCTTAAGAAGGAGATATACTATGTCGAGCCCGGGTTCCAAAGAAGC
	GATTGGAAGTAGAGGTTCTCTGCGCTAGCCAGTTCCTTTTCCAGGTTCTCC
icd2	NC_016946.1:c4402120—4399886
	CCTTAAGAAGGAGATATACTATGAGCGCCGAAAAGCCGACCGTC
	GATTGGAAGTAGAGGTTCTCTGCGCCTTGTGCGGCGGCCAGC
M. marinum	
icl	NZ_HG917972.2:895791—897074
	CGAGAACCTGTACTTCCAATCCATGTCTGTCGTTGGCACCCCGAAGAG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCG
aceA	NZ_PEDC01000044.1:6708—8993
	CCTTAAGAAGGAGATATACTATGGCTATCGCCGAAACGGACACTGAAG
	GATTGGAAGTAGAGGTTCTCTGCGTCGCCCTTGGTGATCAACTGTCGC
icd1	NZ_HG917972.2:1380983—1382206
	CCTTAAGAAGGAGATATACTATGTCTGGTGAACCCAAGATCACGGTCAG
	GATTGGAAGTAGAGGTTCTCTGCGCCCAGCTTCTTCGCCAAATTCTCGG

icd2	NZ_HG917972.2:c181470—179236
	CCTTAAGAAGGAGATATACTATGAGCGCCGAGCAGCCGACC
	GATTGGAAGTAGAGGTTCTCTGCGCCCTGTGCGGCCTCCAGC
M. paraffinicum	
icl (*)	NZ_MPNT01000008.1:177756—179039
	CGAGAACCTGTACTTCCAATCCATGTCTGCCGTTGGCACCCCGAAG
	GATCCGTATCCACCTTTACTGTTAGTGAAACTGACCCTCTTCGGTGGAGC
aceA	NZ_MPNT01000007.1:c43683—41395
	CCTTAAGAAGGAGATATACTATGGCCATCATTGACTCGGACACTGAAGTC
	GATTGGAAGTAGAGGTTCTCTGCGTCTTCCTTGGTGATCAGCTTCCGCAG
icd1	NZ_MPNT01000023.1:19082—20305
	CCTTAAGAAGGAGATATACTATGTCCACCGACAAGATCAAGGTCAAGGG
	GATTGGAAGTAGAGGTTCTCTGCGCTGACCAGCTTCTTCTCCAGGTTCTC
icd2 (*)	NZ_MPNT01000023.1:16812—19046
	CCTTAAGAAGGAGATATACTATGAGCGCCGAACAGCCGACCG
	GATTGGAAGTAGAGGTTCTCTGCCGAGCTATGCGAAGCGTCCAGCG
M. salmoniphilum	
icl	NZ_MAFE01000001.1:205592—206881
	CGAGAACCTGTACTTCCAATCCATGTCGAACGTCGGAAAGCCGCGC
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGACCCTCTTCGGTCGAG
icd2	NZ_CP024633.1:c3512218—3509984
	CCTTAAGAAGGAGATATACTATGAGTGCGCAGCAGCCGACCATC
	GATTGGAAGTAGAGGTTCTCTGCCACCTCGACGGATGCCAGAGCG
M. smegmatis	
icl	NC_008596.1:995201—996484
	CGAGAACCTGTACTTCCAATCCATGTCGACCGTTGGCACCCCGAAG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGACCCTCTTCGGTCGAG
aceA	NC_008596.1:3770113—3772419
	CCTTAAGAAGGAGATATACTATGGCAACAATCGAAGCCAACGCACAC
	GATTGGAAGTAGAGGTTCTCTGCGTCCTCCTTGCGGATCAGCCGG

icd2	NC_008596.1:1748543—1750771
	CCTTAAGAAGGAGATATACTATGAGCGCCCAGCAGCCGACC
	GATTGGAAGTAGAGGTTCTCTGCCACCGATTCGAGCGTGGTGTTGAACG
M. terrae	
icl	NZ_LQPX01000019.1:333181—334464
	CGAGAACCTGTACTTCCAATCCATGTCGACCGTTGGCACGCCGAAG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTCG
aceA	NZ_LQPX01000006.1:191951—194245
	CCTTAAGAAGGAGATATACTATGACGACCATCGAAACGACCACGGC
	GATTGGAAGTAGAGGTTCTCTGCGTCTTTGCTGATCAGCTTGCCCAGCG
icd2	NZ_LQPX01000021.1:25465—27696
	CCTTAAGAAGGAGATATACTATGTGCGCCGACCAGCCGACC
	GATTGGAAGTAGAGGTTCTCTGCGCCCTTGGCGTCGATCACAGCG
***M. tuberculosis* CDC 1551**	
icl	NC_000962.3:557527—558810
	CGAGAACCTGTACTTCCAATCCATGTCTGTCGTCGGCACCCCGAAG
	GATCCGTATCCACCTTTACTGTTAGTGGAACTGGCCCTCTTCGGTGG
aceA (full length)	NC_002755.2:2157754—2160051
	CCTTAAGAAGGAGATATACTATGGCCATCGCCGAAACGGACACC
	GATTGGAAGTAGAGGTTCTCTGCCGCCTCCTTCGTGATCAACTTCCGC
aceA (aa 1—605)	NC_002755.2:2157754—2159568
	CCTTAAGAAGGAGATATACTATGGCCATCGCCGAAACGGACACC
	GATTGGAAGTAGAGGTTCTCTGCCTCAGTTTGCACCAAGTGCTGGTGCTG

8.2.3. Polymerase Chain Reactions for amplification of mycobacterial isocitrate metabolism genes

Polymerase chain reactions (PCR) were performed using a Veriti 96 well thermal cycler (Applied Biosystems). Templates and primers were prepared as described in **Sections**

8.2.1 and **8.2.2** respectively. The PCR process was optimised for each template by varying the polymerase, annealing temperature, and DMSO concentration. Reaction mixtures and protocols were designed according to the instructions provided with the polymerases. The general PCR protocols used are provided in **Table 40** (KAPA HiFi HotStart DNA Polymerase), **Table 41** (KOD HotStart DNA Polymerase), **Table 42** (amplification of *icl* with Platinum SuperFi DNA Polymerase), and **Table 43** (amplification of *icd2* with Platinum SuperFi DNA Polymerase). Optimised PCR conditions for each of the mycobacterial genes amplified are given in **Table 44**.

PCR reaction outcomes were assessed by agarose gel electrophoresis (1.5% agarose/TAE with GelRed). Gels were visualised using a BioDoc-It^2 Imager (UVP) under UV light. The products of successful reactions were purified using a QIAquick PCR Purification Kit, eluting with 30 µL of autoclaved deionised water. DNA concentrations were measured using a NanoDrop One system. Purified inserts were stored at −20 °C until required.

Table 40: PCR protocol for KAPA HiFi HotStart DNA Polymerase.
At the end of the run, the temperature was held at 4 °C.

Step		Temperature/°C	Time	Cycles
Initial denaturation		95	3 minutes	1
Amplification	Denaturation	98	20 seconds	30
	Annealing	varied	15 seconds	
	Extension	72	90 seconds	
Final extension		72	3 minutes	1

Table 41: PCR protocol for KOD HotStart DNA Polymerase.
At the end of the run, the temperature was held at 4 °C. For amplification of *M. aurum icl*, the extension time in the amplification phase was increased to 26 seconds.

Step		Temperature/°C	Time	Cycles
Initial denaturation		95	2 minutes	1
Amplification	Denaturation	95	20 seconds	30
	Annealing	varied	10 seconds	
	Extension	70	20 seconds	
Final extension		70	3 minutes	1

Table 42: PCR protocol for the amplification of *icl* with Platinum SuperFi DNA Polymerase.
When using this polymerase, the annealing and extension steps are combined when the primer melting temperatures exceed 69 °C. The addition of GC enhancer (provided with the polymerase) did not affect the reaction outcome, despite the high GC content of the mycobacterial template DNA. At the end of the run, the temperature was held at 4 °C.

Step		Temperature/°C	Time	Cycles
Initial denaturation		98	30 seconds	1
Amplification	Denaturation	98	10 seconds	30
	Annealing/Extension	72	42 seconds	
Final extension		72	5 minutes	1

Table 43: PCR protocol for the amplification of *icd2* with Platinum SuperFi DNA Polymerase.
When using this polymerase, the annealing and extension steps are combined when the primer melting temperatures exceed 69 °C. The addition of GC enhancer (provided with the polymerase) did not affect the reaction outcome, despite the high GC content of the mycobacterial template DNA. At the end of the run, the temperature was held at 4 °C.

Step		Temperature/°C	Time	Cycles
Initial denaturation		98	30 seconds	1
Amplification	Denaturation	98	10 seconds	30
	Annealing/Extension	72	69 seconds	
Final extension		72	5 minutes	1

Table 44: Optimised PCR conditions for the amplification of mycobacterial isocitrate metabolism genes. Reaction mixtures were designed according to the instructions supplied with the polymerases. Where used, DMSO comprised 5% and GC enhancer 20% of the reaction mixture by volume. During optimisation of PCR reactions with KAPA or KOD polymerase, annealing temperatures were varied in the range 55—72 °C. **M. aurum icl* was amplified using the protocol shown in **Table 41**, but with an extension time of 26 seconds in the amplification phase. The templates used were those prepared in **Section 8.2.1**, and the primers used were those designed in **Section 8.2.2** and listed in **Table 39**. Abbreviations: KAPA, KAPA HiFi HotStart DNA Polymerase; KOD, KOD HotStart DNA Polymerase; Platinum SuperFi, Platinum SuperFi DNA Polymerase; DMSO, dimethyl sulfoxide; GC enhancer, Platinum SuperFi GC enhancer; aa, amino acids.

Species	Gene	Polymerase	Additive	Annealing temperature (°C)	Protocol
M. aromaticivorans	*icl*	KAPA	n/a	72	**Table 40**
M. aurum	*icl*	KOD	DMSO	70	**Table 41***
M. chubuense	*icl*	KOD	DMSO	70	**Table 41**
M. fallax	*icl*	KOD	DMSO	70	**Table 41**
M. hiberniae	*icl*	KOD	DMSO	60	**Table 41**
M. intracellulare	*icl*	KOD	DMSO	67	**Table 41**
M. marinum	*icl*	Platinum SuperFi	GC enhancer	72	**Table 42**
M. salmoniphilum	*icl*	KOD	DMSO	60	**Table 41**
M. smegmatis	*icl*	KOD	DMSO	65	**Table 41**
M. terrae	*icl*	Platinum SuperFi	GC enhancer	72	**Table 42**
M. tuberculosis CDC 1551	*icl*	KOD	DMSO	70	**Table 41**
M. aromaticivorans	*aceA*	KAPA	n/a	65	**Table 40**
M. aurum	*aceA*	KOD	DMSO	60	**Table 41**
M. chubuense	*aceA*	KAPA	n/a	70	**Table 40**
M. hiberniae	*aceA*	KAPA	n/a	67	**Table 40**
M. intracellulare	*aceA*	KOD	DMSO	60	**Table 41**
M. marinum	*aceA*	KAPA	n/a	65	**Table 40**
M. paraffinicum	*aceA*	KOD	DMSO	60	**Table 41**
M. smegmatis	*aceA*	KAPA	n/a	67	**Table 40**

M. terrae	*aceA*	KAPA	n/a	65	**Table 40**
M. tuberculosis CDC 1551	*aceA*, full length	KAPA	n/a	72	**Table 40**
M. tuberculosis CDC 1551	*aceA*, aa 1—605	KAPA	n/a	72	**Table 40**
M. intracellulare	*icd1*	KAPA	n/a	72	**Table 40**
M. marinum	*icd1*	KAPA	n/a	72	**Table 40**
M. paraffinicum	*icd1*	KOD	DMSO	55	**Table 41**
M. aromaticivorans	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. aurum	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. chubuense	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. fallax	*icd2*	KOD	DMSO	62	**Table 41**
M. hiberniae	*icd2*	KOD	DMSO	70	**Table 41**
M. intracellulare	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. marinum	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. salmoniphilum	*icd2*	KOD	DMSO	70	**Table 41**
M. smegmatis	*icd2*	Platinum SuperFi	GC enhancer	72	**Table 43**
M. terrae	*icd2*	KOD	DMSO	67	**Table 41**

8.2.4. Vector preparation

Glycerol stocks of pNIC-NTH and pNIC-CTHF in *E. coli* NEB5α were provided by Acely Garza-Garcia. For each vector, a flake of the glycerol stock was used to inoculate 13 mL of LB medium containing 60 µg/mL kanamycin (LB-kanamycin). The culture was incubated at 37 °C overnight, with shaking, then centrifuged at 1,811 × *g* for 10 minutes. The supernatant was discarded. Cell lysis and DNA extraction were performed using a QIAprep Spin Miniprep Kit. The extracted DNA was quantified using a NanoDrop One system.

Vectors were linearized by incubation with the appropriate restriction enzymes. Reaction progress was monitored by DNA gel electrophoresis (1.5% agarose/TAE with GelRed). Gels were visualised using a BioDoc-It^2 Imager (UVP) under UV light. Linearized vectors were purified using a QIAquick PCR purification kit, eluting with 30 µL of autoclaved deionised water. The DNA concentration was measured on a NanoDrop One system. Linearized vectors were stored at −20 °C until required.

Circular pNIC-NTH (35 µL), CutSmart buffer (5 µL), and autoclaved deionised water (9 µL) were added to a 500 µL Eppendorf tube. BsaI-HF (20 units/µL, 1 µL) was added and the mixture was incubated at 37 °C for 2.5 hours. A further 1 µL of BsaI-HF was added and the mixture was incubated at 37 °C for a further 2 hours, after which time DNA gel electrophoresis indicated that linearization was complete.

Circular pNIC-CTHF (35 µL), 3.1 buffer (5 µL), and autoclaved deionised water (7 µL) were added to a 500 µL Eppendorf tube. Bfu-AI (5 units/µL, 3 µL) was added and the mixture was incubated at 50 °C for 2.5 hours. A further 1 µL of Bfu-AI was added and

the mixture was incubated at 50 °C for a further 2 hours, after which time DNA gel electrophoresis indicated that linearization was complete.

8.2.5. Isothermal assembly

Isothermal assembly reactions were set up on ice according to the NEBuilder HiFi DNA Assembly Cloning Kit protocol, using a 2:1 insert:vector ratio. Reaction mixtures were incubated at 50 °C for 30 minutes. Assembly products were used immediately or stored at −20 °C.

8.2.6. Transformation: *E. coli* NEB5α

E. coli NEB5α cells were thawed on ice. 2 µL of construct (assembly product) was added and mixed by flicking gently. Cells were incubated on ice for 30 minutes, then heat-shocked at 42 °C for 30 seconds before being returned to the ice for a further 2 minutes. 950 µL of SOC outgrowth medium was added to each tube and the tubes were incubated at 37 °C, with shaking, for 1 hour. Cultures were centrifuged at 21,100 × *g* for 5 minutes, then 800 µL of supernatant were removed. The remaining 200 µL of supernatant were used to resuspend the pellet and then plated onto LB agar containing kanamycin (60 µg/mL) and sucrose (6% *w/v*). Plates were incubated at 37 °C overnight or at room temperature over the weekend.

Transformation success was analysed by colony PCR. A single colony was resuspended in 10 µL deionised water. 1 µL of the suspension and 15 µL of T7 master mix (1× KAPA2G Robust HotReadyMix, 0.5 µM T7 forward primer, and 0.5 µM T7 reverse primer) were added to a PCR tube. The PCR protocol is given in **Table 45**. PCR products were analysed

by agarose gel electrophoresis (1.5% agarose/TAE with GelRed). Gels were visualised on a BioDoc-It^2 Imager (UVP) under UV light.

Table 45: Protocol for colony PCR using KAPA2G Robust HotReadyMix.
At the end of the run, the temperature was held at 4 °C.

Step		Temperature/°C	Time	Cycles
Initial denaturation		95	5 minutes	1
Amplification	Denaturation	95	15 seconds	35
	Annealing	55	15 seconds	
	Extension	72	90 seconds	
Final extension		72	5 minutes	1

Where colonies tested positive for insert, 100 µL of LB-kanamycin medium were added to the remaining colony suspension. The resulting suspension was used to inoculate 13 mL of LB-kanamycin medium. Cultures were incubated at 37 °C overnight, with shaking, then centrifuged at 1,811 × *g* for 10 minutes. Cell lysis and DNA extraction were performed using a QIAprep Spin Miniprep Kit. The extracted DNA was quantified using a NanoDrop One system, then prepared for sequencing as described in **Section 8.2.7**.

8.2.7. Sequencing

Before sequencing, purified plasmids were diluted to ~50 ng/µL with autoclaved deionised water. Samples were submitted to GATC for Sanger sequencing using pBR3 and pRSET-RP primers. For constructs containing *aceA* or *icd2*, additional sequencing primers were designed to allow the full insert sequence to be obtained. Primer sequences are given in **Table 46**. The sequencing results were translated (ExPASy Translate) and compared to the desired sequences used in construct design.[543]

Table 46: Standard and custom primers used for construct sequencing.
The standard primers pBR3 and pRSET-RP were used to sequence all constructs. Additional sequencing primers were required for verification of full-length *aceA* and *icd2* sequences. The sequences of these primers are given together with the names of the mycobacterial species for which they were used.

Standard primers used to sequence all constructs	
pBR3	TCCCCATCGGTGATGTC
pRSET-RP	ATGCTAGTTATTGCTCAGC
Additional primers used to sequence *aceA*s	
M. aromaticivorans	AGCTCAACGGTCATCTGCTG
M. aurum, *M. hiberniae*, *M. paraffinicum*, *M. smegmatis*, *M. terrae*	CCGCCGTAGGTGATGAAGTTGAAG
M. chubuense, *M. tuberculosis* CDC 1551	TTCCAGCTCGACATCATGCG
M. intracellulare	TGGGCGTCAAAGACCTCAAC
M. marinum	CAAGGAACTCAACGGTCACCTG
Additional primers used to sequence *icd2*s	
M. aromaticivorans, *M. hiberniae*, *M. salmoniphilum*	ACTTCTACCACGGTGAGAAGTC
M. aurum, *M. fallax*, *M. smegmatis*, *M. terrae*	ACTTCTACCACGGCGAGAAGTC
M. chubuense	TGAAGCACGGCGACTTCTATC
M. intracellulare	ACTTCTATCACGGCGAGAAGTC
M. marinum	GCCGGTGAGAAGTCGATAAC

8.3. List of constructs

Constructs encode full length proteins except where otherwise specified. All tags used were tobacco etch virus (TEV) protease cleavable; cleavage sites are indicated by *. Protein sizes include tags. Extinction coefficients were calculated using ExPASy ProtParam.[543]

8.3.1. *Mycobacterium tuberculosis* CDC 1551, ICL1

Protein size: 446 aa/49.25 kDa	**Vector:** pNIC-NTH	**Tag:** *N*-terminal His_6
Protein sequence: MHHHHHHSSGTENLYFQ*SMSVVGTPKSAEQIQQEWDTNPRWKDVTRTYSAEDVVALQGS VVEEHTLARRGAEVLWEQLHDLEWVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANL SGHTYPDQSLYPANSVPQVVRRINNALQRADQIAKIEGDTSVENWLAPIVADGEAGFGGALNV YELQKALIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVIAR TDAEAATLITSDVDERDQPFITGERTREGFYRTKNGIEPCIARAKAYAPFADLIWMETGTPDLEA ARQFSEAVKAEYPDQMLAYNCSPSFNWKKHLDDATIAKFQKELAAMGFKFQFITLAGFHALNY SMFDLAYGYAQNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDRIATTVDPNSSTTALT GSTEEGQFH		
Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 73465		

8.3.2. *Mycobacterium tuberculosis* CDC 1551, ICL2, full length

Protein size:	**Vector:**	**Tag:**
788 aa/88.09 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIAETDTEVHTPFEQDFEKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAG

AFYERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTEDPGPDLASYPLS

QVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAYDFRPFIIADADTGHGGDPHVRNLIR

RFVEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEA

ANLIDSRADERDQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDSEYAAAGG

WLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGLMTYGEAVADVLEFG

QSEGEPIGMAPEEWRAFAARASLHAARAKAKELGADPPWDCELAKTPEGYYQIRGGIPYAIAK

SLAAAPFADILWMETKTADLADARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMRR

FPEELGKMGFVFNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQTLV

GGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPRKLLEEWLAMWSGHYQLKDKLRV

QLRPQRAGSEVLELGIHGESDDKLANVIFQPIQDRRGRTILLVRDQNTFGAELRQKRLMTLIHL

WLVHRFKAQAVHYVTPTDDNLYQTSKMKSHGIFTEVNQEVGEIIVAEVNHPRIAELLTPDRVAL

RKLITKEAAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}:	97875
Protein size (after tag cleavage):	773 aa/86.19 kDa
Extinction coefficient (280 nm)/M^{-1} cm^{-1} (after tag cleavage):	96385

8.3.3. *Mycobacterium tuberculosis* CDC 1551, ICL2, truncated (aa 1—605)

This construct contains amino acids 1—605 of Mtb CDC 1551 ICL2.

Protein size:	**Vector:**	**Tag:**
627 aa/69.42 kDa	pNIC-CTHF	*C*-terminal His_6FLAG
Protein sequence: MAIAETDTEVHTPFEQDFEKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAA GAFYERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTEDPGPDLASYPL SQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAYDFRPFIIADADTGHGGDPHVRNLI RRFVEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAE AANLIDSRADERDQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDSEYAAAG GWLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGLMTYGEAVADVLEF GQSEGEPIGMAPEEWRAFAARASLHAARAKAKELGADPPWDCELAKTPEGYYQIRGGIPYAIA KSLAAAPFADILWMETKTADLADARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMR RFPEELGKMGFVFNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQTL VGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEAENLYFQ*SHHHHHHDYKDDDDK		
Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 76905		

8.3.4. *Mycobacterium tuberculosis* CDC 1551, IDH1, codon optimised

A glycerol stock of this construct in *E. coli* NEB5α was provided by Raphael Reinbold (University of Oxford). This construct was codon-optimised for translation in *E. coli*.

Protein size:	**Vector:**	**Tag:**
431 aa/48.28 kDa	pNIC-CTHF	*C*-terminal His_6FLAG
Protein sequence: MSNAPKIKVSGPVVELDGDEMTRVIWKLIKDMLILPYLDIRLDYYDLGIEHRDATDDQVTIDAAY AIKKHGVGVKCATITPDEARVEEFNLKKMWLSPNGTIRNILGGTIFREPIVISNVPRLVPGWTKPI VIGRHAFGDQYRATNFKVDQPGTVTLTFTPADGSAPIVHEMVSIPEDGGVVLGMYNFKESIRD FARASFSYGLNAKWPVYLSTKNTILKAYDGMFKDEFERVYEEEFKAQFEAAGLTYEHRLIDDMV AACLKWEGGYVWACKNYDGDVQSDTVAQGYGSLGLMTSVLMTADGKTVEAEAAHGTVTRH YRQYQAGKPTSTNPIASIFAWTRGLQHRGKLDGTPEVIDFAHKLESVVIATVESGKMTKDLAILI GPEQDWLNSEEFLDAIADNLEKELANAENLYFQ*SHHHHHHDYKDDDDK		
Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 70945		

8.3.5. *Mycobacterium tuberculosis* CDC 1551, IDH2

A glycerol stock of this construct in *E. coli* NEB5α was provided by Raphael Reinbold (University of Oxford).

Protein size:	**Vector:**	**Tag:**
767 aa/85.32 kDa	pNIC-CTHF	*C*-terminal His_6FLAG
Protein sequence: MSAEQPTIIYTLTDEAPLLATYAFLPIVRAFAEPAGIKIEASDISVAARILAEFPDYLTEEQRVPDNL AELGRLTQLPDTNIIKLPNISASVPQLVAAIKELQDKGYAVPDYPADPKTDQEKAIKERYARCLGS AVNPVLRQGNSDRRAPKAVKEYARKHPHSMGEWSMASRTHVAHMRHGDFYAGEKSMTLD RARNVRMELLAKSGKTIVLKPEVPLDDGDVIDSMFMSKKALCDFYEEQMQDAFETGVMFSLH VKATMMKVSHPIVFGHAVRIFYKDAFAKHQELFDDLGVNVNNGLSDLYSKIESLPASQRDEIIED LHRCHEHRPELAMVDSARGISNFHSPSDVIVDASMPAMIRAGGKMYGADGKLKDTKAVNPES TFSRIYQEIINFCKTNGQFDPTTMGTVPNVGLMAQQAEEYGSHDKTFEIPEDGVANIVDVATG EVLLTENVEAGDIWRMCIVKDAPIRDWVKLAVTRARISGMPVLFWLDPYRPHENELIKKVKTYL KDHDTEGLDIQIMSQVRSMRYTCERLVRGLDTIAATGNILRDYLTDLFPILELGTSAKMLSVVPL MAGGGMYETGAGGSAPKHVKQLVEENHLRWDSLGEFLALGAGFEDIGIKTGNERAKLLGKTL DAAIGKLLDNDKSPSRKTGELDNRGSQFYLAMYWAQELAAQTDDQQLAEHFASLADVLTKNE DVIVRELTEVQGEPVDIGGYYAPDSDMTTAVMRPSKTFNAALEAVQGAENLYFQ*SHHHHHH DYKDDDDK		
Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 70625		

8.3.6. *Mycobacterium aromaticivorans*, ICL1

Protein size:	**Vector:**	**Tag:**
448 aa/49.28 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMTLSNVGQPKSPEEIQKDWDTNPRWKGITRTYTPADVVALQG

SVVEEATLARRGAEVLWNQLHDMEFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDAN

LSGHTYPDQSLYPANSVPQVVRRINNALMRADQIAKVEGDTSIENWLAPIVADGEAGFGGALN

VYELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVI

ARTDAEAATLITSDVDDRDKPFVTGERTAEGFYRVKNGLEPCIARAKAYAPYSDLIWMETGTPD

LELAAKFAEGVKSEFPDQMLAYNCSPSFNWRKHLDDATIAKFQKELGAMGFKFQFITLAGFHA

LNYSMFDLAYGYARNQMSAYVELQEREFDAEERGYTATKHQREVGAGYFDRIATTVDPTSSTT

ALSGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.7. *Mycobacterium aromaticivorans*, ICL2

Protein size:	**Vector:**	**Tag:**
789 aa/88.10 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSIIEADSATQSPHQRAVSETQQYIDSPRFEGITRLYTAAQVVEQRGTIPVDYTVAREASAAFYK

RLRELFAAKKSITTFGPYSPGQAVTMKRMGIEGIYLGGWATSAKGSTTEDPGPDLASYPLSQVP

EEAAGLVRALLTADRNQQYLRLQMTPEQRAATPAVDYRPFIIADADTGHGGDPHVRNLIRRFV

EAGVAGYHIEDQRPGTKKCGHQGGKVLVPSDEQLKRLNTARFQLDIMGVPGIIVARTDAEAAN

LIDSRGDERDQPFLLGATNLKVPPYKSCFLAMVRRFYESGVTELNGHLLYALPEGEYAVADAWL

QRQGIADLITEQAAAWRDGKQHSLDALFDKVESRFVDAWQDDASLDTYGDAVAELLAFRESE

GEPHEMSAADWREFAKTASLHAAREKARELGVDAAWDPERAKTPEGYYQVRGGIPYAIAKSL

AAAPFADLLWMETKTADLADAREFAEAVHAVYPDKMLAYNLSPSFNWDTTGMTDEEMRAFP

EELGKMGFVFNFMTYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQTLVG

GPRSDAALAASSGRTATTKSMGKGSTQHQHLVQTEVPKKLLEDWLALWSEHYELGEKLRVTLR

PRRAGGEVLELGIYGEREKGGEQELLANVLVDPIKDRHGRSILTVRDQNTFAEKLRQKRLMTLIH

LWMVHRFKADAVYYVTPTEDNIYQADKMKAHGIFSDVHKDVGEIIVADVNQERIAELLEPDRV

ALKRLIAKEDAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 103835

8.3.8. *Mycobacterium aromaticivorans*, IDH2

Protein size:	**Vector:**	**Tag:**
763 aa/84.79 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAENPTIIYTLTDEAPLLATYAFLPVLRTFVEAAGIDVKTSDISVAARILAEFSDRLTDEQKVPDN
LAELGELTQDPSANIIKLPNISASVPQLIAAIKELKSKGYDLPDFPGDPKTDEEKEIRQRYGKILGSA
VNPVLREGNSDRRAPKAVKEYAKKHPHSMGQWSQASRTHVATMKTGDFYHGEKSMTLDKD
RTVKMVLTPKTGDPIVLKPEVKLDNGDVIDSMFMSKKALIDFYEKEIEDAYKTGVMFSLHVKAT
MMKVSHPIVFGHAVKVFYKDAFAKHGKLFDELGVNVNNGLSDLYDKIESLPASQREEIIEDLHK
CHEHRPELAMVDSAKGISNFHSPSDVIVDASMPAMIRLGGKMYGADGRTKDTKAVNPESTFS
RMYQEMINFCKTHGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEIPVDGVADIVDVDTGE
VLLSWDVEEGDIWRMPIVKDAPIRDWVKLAVNRARLSGMTTVFWLDDERPHENELRKKVKAY
LKEEDTEGLDITILPQVWAMRYTLERVIRGQDTIAATGNILRDYLTDLFPILELGTSAKMLSIVPLM
AGGGLYETGAGGSAPKHVSQLVEENHLRWDSLGEFLAIGASLEDLGNKTDNAKAKVLATTLDT
AVGKLLDANKNPSRKAGELDNRGSQFYLALYWAQALAEQTEDKDLAEHFAPLATTLAENEDAI
VAELAEVQGQAVDIGGYYYPDREKTASVMRPSKTFNEALSAENLYFQ*SHHHHHHDYKDDDD
K

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 79885

8.3.9. *Mycobacterium aurum*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.25 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSTVGTPKSPEQIQHDWDHNPRWKGITRTYTPKDVVALQGH

VVEESTLARRGAEVLWEQLHDMDYVNSLGALTGNQAVQQVRAGLKAIYLSGWQVAGDANLS

GHTYPDQSLYPANSVPQVIRRINNALLRADQIAKVEGDRSVENWLAPIVADGEAGFGGALNVY

ELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVIAR

TDAEAATLITSDVDERDQPFITGERTKEGFYRVKNGLEPCIARAKSYAPYSDLIWMETGTPDLEL

AKKFAEGVKAEFPDQMLAYNCSPSFNWKQHLDDATIAKFQNELGAMGFKFQFITLAGFHALN

YSMFDLAHGYARNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDKIATTVDPTSSTTAL

AGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.10. *Mycobacterium aurum*, ICL2

Protein size:	**Vector:**	**Tag:**
784 aa/87.95 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIIEADLQPQTPFERDVADTQRYFDSPRFDGIIRLYTARQVAEQRGTIPVDYPVAREAATAFFP

HLRELYANGRSITTFGPYSPGQAVVMKRMGIEGIYLGGWATSAKGSISEDPGPDLASYPLSQVP

EEAASIVRALLTADRNQQYLRLQMNEAQLKTTPAYDYRPFIIADADTGHGGDPHVRNLIRRFVE

SGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDIMRVPGIIVARTDAEAANLI

DSRADERDQPFLLGATNLKIPSYKSCFLAMVRRFYDKGVTELNGHLLYALPDGEYATAEAWLER

NGVLALIDQTAAAWQDGTESSVDAAFDKVENQFVDAWQDDAGLNTYGEAVAELLEFREREG

EPSGMSVADWRTFAERAPLYTAREKAKELGADVAWDCERVKTPEGYYQVRGGIPYAIAKSLAA

APFADILWMETKTADLADAKQFADAIHAVYPDKMLAYNLSPSFNWDTTGMTDEEMRAFPEEI

GKMGFVFNFITYGGHQVDGVASEEFATSLRQEGMLALARLQRKMRLVESPYRTPQTLVGGPR

SDAALAASSGRTATTKSMGKGSTQHQHLVQTEVPKKLLEEWLALWSEHNKLPERLHVQLRPRR

AGSDVLDLQILGDGEEPLANVVVDPIKDRHGRSILTVRDQNTFAEQLRQKRLMDVIHLWLIHRF

KAEIVYYVTPTEDNIYQTEKMKSHGIFSDVYQEVGEIIVADVNQARIDELLAPDRAALTRLISKQD

AENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 105325

8.3.11. *Mycobacterium aurum*, IDH2

Protein size:	**Vector:**	**Tag:**
767 aa/85.33 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAEQPSIIYTLTDEAPLLATYAFLPVVQTFASAAGIDVKSTDISVAARILAEFSDRLTDEQKVPDT

LAELGELTQLRETNIIKLPNISASVPQLLAAIKELKAKGYDLPDWVGEPKNDEEKAIKERYSKVLGS

AVNPVLRQGNSDRRAPKAVKEYARKHPHSMGEWSQASRTHVATMKKGDFYHGEKSLTIDKD

RTVKMVLTTAGGETLVLKEEVKLDEGDVIDSMFMSKKALCDFYEEQIEDAYKTGVMFSLHVKAT

MMKVSHPIVFGHAVKVFYKDAFAKHQKLFDELGVNVNNGMSDLYDKIETLPASQREEIIRDIHA

CHEHRPELAMVDSARGITNFHSPSDVIVDASMPAMIRLGGKMYGADGRTKDTKAVNPESTFS

RIYQEIINFCKTNGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEIPEDGVADIVDVDTGEVLL

TQNVETGDIWRMPIVKDAAIRDWVKLAVNRARASGMTAVFWLDDERPHENELRKKVKAYLK

DEDTDGLEITILPQVWAMRYTLERVIRGQDTIAVTGNILRDYLTDLFPILELGTSAKMLSIVPLMA

GGGMYETGAGGSAPKHVSQLLEENHLRWDSLGEYLALGASFEDMGAKFDNAKATLLGKTLDS

AIGTLLENDKGPSRKAGELDNRGSQFYLAMYWAQELAKQTEDKELAEHFAPLAKALAENESTIV

SEFSAVQGKPVDIGGYYYPDRDKTTDVMRPSTTFNDALEAARSAENLYFQ*SHHHHHHDYKD

DDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 81375

8.3.12. *Mycobacterium chubuense*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.16 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSTVGTPKSPEQIQHDWDHNPRWKGVTRTYTPQDVVALQGT

VVEEHTLARRGAEVLWEQLHDMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANL

SGHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDRSVENWLAPIVADGEAGFGGALNV

YELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVAGVPTVVIA

RTDAEAATLITSDVDERDQPFITGDRTAEGFYRVKNGLEPCIARAKAYAPFSDLIWMETGTPDLE

LAKKFAEGVKSEFPDQMLAYNCSPSFNWRKHLDDATIAKFQKELGAMGFKFQFITLAGFHALN

YSMFDLAYGYARNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDRIATTVDPTSSTTAL

AGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 66475

8.3.13. *Mycobacterium chubuense*, ICL2

Protein size:	**Vector:**	**Tag:**
784 aa/88.06 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIIEADLQPETPFERDVAETQRYFDSPRFEGIIRLYTARQVAEQRGTIPADYPVAREAATAFYP

HLRELYSRGKSITTFGPYSPGQAVVMKRMGIEGIYLGGWATSAKGSISEDPGPDLASYPLSQVPE

EAAGIVRALLTADRNQQYLRLQMTEEQRAATPAFDYRPFIIADADTGHGGDPHVRNLIRRFVE

NGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDIMRVPGIIVARTDAEAANLI

DSRADERDQPFLLGVTNLKIPSYKSCFLAMVRRFYNQGVTELNGHLLYALPDGEYATAEAWLER

QGVLALIDQTAAAWQDGAEPSLDAAFDTVESRFVDAWQDDAGLDTYAGAVADLLEFRREREG

ESTTMSATEWRAFAERAPLYNAREKAQELGADVAWDCERAKTPEGYYQVRGGIPYAIAKSLAA

APFADILWMETKTADLADAKQFADAIHAVYPDKMLAYNLSPSFNWDTTGMTDEEMRAFPEEI

GKMGFVFNFITYGGHQVDGVASEEFATSLRQEGMLALARLQRKMRLVESPYRTPQTLVGGPR

SDAALAASSGRTATTKSMGKGSTQHQHLVQTEVPKKLLEEWLALWSEHYQLPEKLRVQLRPRR

AGSDVLDLQILGDGEEPLANVVVDPIKDRHGRSILTVRDQNTFAEQLRQKRLMDLIHLWLIHRF

KAEIVYYVTPTEDNIYQTEKMKSHGIFSDVYQEVGEIIVADVNQARIDELLAPDQRALLRLISKQD

AENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 106815

8.3.14. *Mycobacterium chubuense*, IDH2

Protein size:	**Vector:**	**Tag:**
767 aa/85.72 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAEQPTLFYTLTDEAPLLATYAFLPVVRAFAHAADIDVKSRDISVAARILAEFSDRLTEEQRVPD

SLTELGELTQFPETNIIKLPNISASVPQLLAAIKELKSKGYDLPDYPGEAKTDEEKKIKERYSKILGSA

VNPVLREGNSDRRAPKAVKEYARKHPHSMGKWSMASRTHVATMKHGDFYHGEKSMTLDKA

RTVKMVLETASGKTIVLKPEVKLDEGDIIDSMFMSKKALCEFYEEQMQDAFETGVMFSLHVKAT

MMKVSHPIVFGHAVKVFYKDAFAKHQALFDELGVNVNNGMSDLYDKISSLPASQREEIINDLH

ACHEHRPELAMVDSARGITNFHSPSDVIVDASMPAMIRLGGKMYGADGRTKDTKAVNPESTF

SRMYQEMINWCKTNGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEIPEDGVANIVDVDT

DEVLLSQNVEEGDIWRMPVVKDAPIRDWVKLAVTRARASGMTAVFWLDTERPHEAELRKKV

KEYLKDHDTEGLHIQIMPQVWAMRYTLERVVRGQDTIAVTGNILRDYLTDLFPILELGTSAKML

SIVPLMAGGGLYETGAGGSAPKHVHQLVEENHLRWDSLGEYLALGASFEDMGDKAGNKKASI

LGKTLDAAIGKLLENDKSPSRKTGELDNRGSQFYLTLYWAQELANQSEDQELADHFAPLATTLA

EQEETIVAELNKAQGNAVDIGGYYYPDREKTTAVMRPSKTFNDALESLPSAENLYFQ*SHHHHH

HDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 81375

8.3.15. *Mycobacterium fallax*, ICL1

Protein size:	**Vector:**	**Tag:**
450 aa/49.48 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMESTVSTVGTPKSAEQIQHDWDHNPRWKGITRTYTPADVVAL

QGTVVEEATLARRGAEVLWAQLHDMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAG

DANLSGHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDTSVENWLAPIVADGEAGFGG

ALNVYELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPT

VVIARTDAEAATLITSDVDERDRPFITGERTAEGFYHLKNGLEPCIARAKAYAPYSDLIWMETGT

PDLALAKKFAEGVKSEFPDQMLAYNCSPSFNWKQHLDDDTIAKFQRELGAMGFKFQFITLAGF

HALNYSMFDLAYGYAREQMKAYVELQEREFAAEERGYTATKHQREVGAGYFDKIATTVDPNSS

TTALSGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.16. *Mycobacterium fallax*, IDH2

Protein size:	**Vector:**	**Tag:**
769 aa/85.12 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAEQPTIIYTLTDEAPLLATYAFLPVIRTFAGAAGIDVKSSDISVAARVLAEFPEYLTEAQRVPDN

LAELGRLTQLPDTNIIKLPNISASVPQLKAAIKELQDKGYALPDFPDDPKTDAEQETRRRYTKCLG

SAVNPVLREGNSDRRAPRAVKEYARKHPHSMGRWSMASRTHVAHMTVGDFYHGEKSMTLD

RSVDVRMELVTGSGEAIVLKPKVALDEGDVIDSMYMSKTALCEFYEEQMQDAFKTGVMFSLH

VKATMMRVSHPIVFGHAVKVFYKEAFAKHQQLFDDLGVNVNNGLSDLYSKIESLPASLHEEIIE

DLHACHENRPELAMVDSARGISNFHSPSDVIVDASMPAMIRAGGKMYGADGKLKDTKAVNP

ESTFSRIYQEMINFCKTNGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEISENGLARIVDIET

GEVLLDQQVEAGDIWRLCIVKDAPIRDWVKLAVNRARQSGMPVVFWLDSERPHENELRKKVK

TYLADHDTEGLDIQIMSQVWAMRWTLERLVRGLDTISATGNILRDYLTDLFPILELGTSAKMLSI

VPLMAGGGMYETGAGGSAPKHVKQLVEENHLRWDSLGEFLALAVSLEDLGTKTGNEQANILA

RTLDAATGRLLDENKSPSRKSGELDNRGSQFYLAMYWAQELAAQTGDAELAARFAPLAAQLA

ENEDTIVAELAEVQGHPVDIGGYYKSDAALTEAVMRPSKTFNAALEGASAQAAENLYFQ*SHH

HHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 78520

8.3.17. *Mycobacterium hiberniae*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.12 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSTVGTPKSPEQIQHDWDHNPRWKGVTRTYTAEDVVALQGR

VVEEHTLARRGAEVLWEQLHDLDYINALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLS

GHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDTSVDNWLAPIVADGEAGFGGALNVY

ELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVIAR

TDAEAATLITSDVDERDQPFITGDRTAEGFYRVRNGLEPCIARAKAYAPYADLIWMETGTPDLEL

ARKFAEGVKAEFPDQMLAYNCSPSFNWRQHLDDATIAKFQNELGAMGFKFQFITLAGFHALN

YSMFDLAHGYARNQMSAYVELQEREFAAESRGYTATKHQREVGAGYFDRIATTVDPASSTTAL

AGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.18. *Mycobacterium hiberniae*, ICL2

Protein size:	**Vector:**	**Tag:**
788 aa/87.23 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MTTIQKNPTTSTPVDRGIDADIAAMQQYFDSPRFEGITRLYSARQVVEQRGTIPADYTVAREAA

AAFYARLRELFAAKKSITTFGPYSPGQAVVMKRVGIEGIYLGGWATSAKGSITEDPGPDLASYPL

SQVPDEAAGLVRALLTADRNQQYLRLQMTPEQRAATPAVDYRPFIIADADTGHGGDPHVRNLI

RRFVEAGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDLMGVPGIIVARTDA

EAANLVESRGDERDQPFILGTTNLRIPSYKACFLAMTRQLHNSGITEINGHLLYALPEGEYAAAD

AWLERAGITALITEAARGWSSDGSRSVDGIFDQIESKFVEAWQNDAGLMTFADAVAEMLEFA

EREGEAKDMTADQWREFAARAPLYTARAKAAELNADPGWDCERAKTPEGYFQVRGGIPYAIA

KSLAAAPFADLLWMETKTADLAEAREFAEAIHAQFPDQMLAYNLSPSFNWDTTGMTDDEMR

AFPAELGKMGFVFNFITYGGHQVDGVASEEFAATLLQDGMLALARLQRKMRLVESPYRTPQTL

VGGPRSDAALAASSGRTATTKAMGAGSTQHQHLVQTEVPKKLLEDWLALWAQHYQLGEKLR

VQLRPTRPGSDVLELGIYGDDDEKLANVIVDPIKDRHGRSILTVRDQNTFAEKLRQKRLMDLAH

LWLIHRFKAAAVHYVTPTQDNLYQTEKMKEHGIFSSVNQEVGEIIVADVNHARIEELLAGDRVA

LGKLISKADAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 97875

8.3.19. *Mycobacterium hiberniae*, IDH2

Protein size:	**Vector:**	**Tag:**
766 aa/84.94 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MCADQPTIFYTLTDEAPLLATYSFLPIVQAFTGPAGIDVKPSDISLAARILAEFPDELTEEQRVPDN

LAELGRRTLLPDTTIIKLPNISASVPQLIAAIKELRGKGYALPDFPEDPKTDRERDVRNRYARCLGS

AVNPVLREGNSDRRAPKAVKEYARKHPHSMGEWSQASRTHVATMRHGDFYHGEKSMTLDR

DRDVKMELKTKSGKTIVLKEKVSLLDGTVIDSMFMSVRALREFYEEQMEDARKTGVMFSLHVK

ATMMKISHPIVFGHAIKVFYKDAFAKHQQLFDDLGVNVNNGLVDLYNKIETLPASKREEVIEDL

HACHEHRPELAMVDSANGITNFHSPSDVIVDASMPAMIRAGGKMYGADGRQKDTKAVNPES

TFARIYQELINFCKTHGQFDPATMGTVPNVGLMAQKAEEYGSHDKTFEVPEDGVADIVDLATG

EVLLSQKVEAGDIWRLCTVTDAAIRDWVKLAVQRARASGMTAVFWLDTERPHEAELRKKVKA

YLADHDTEGLEIQIMPQVWAMRYTLERLIRGQDTIAVTGNILRDYLTDLFPILELGTSAKMLSIVP

LMAGGGMYETGAGGSAPKHVKQLVEENHLRWDSLGEFLALAASFDDLGAKTGNSAAKILGVT

LDAAIGKLLDANKSPSRKTGELDNRGSQFYLAKYWAEELAAQTEDAELAEHFAALANALDKNE

DAILNELNVVQGKPVDIGGYYAPDHALTSAVMRPSPTLNAAIDVKDAENLYFQ*SHHHHHHDY

KDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 71530

8.3.20. *Mycobacterium intracellulare*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.16 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSVVGTPKSAEQIQKDWDTNPRWKDVTRTYSAKDVVALQGT

VVEEATLARRGAEVLWEQLHDLEYINALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLS

GHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDTSVKNWLAPIVADGEAGFGGALNVY

ELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVCDVPTVVIAR

TDAEAATLITSDVDERDQPFITGERTKEGFYRVRNGLEPCIARAKAYAPYSDLIWMETGTPDLEL

AAKFAEGVKSEFPDQMLAYNCSPSFNWKKHLDDSTIAKFQKELGAMGFKFQFITLAGFHALNY

SMFDLAYGYARNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDRIATTVDPSSSTTALT

GSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 69580

8.3.21. *Mycobacterium intracellulare*, ICL2

Protein size:	**Vector:**	**Tag:**
784 aa/87.84 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIIDKDTQVRPSFDDEVAATQQYFDDPRFSRITRLYTARQVAEQRGTIPTDYTVARNAAAAFY

ERLRELFAQKKSITTFGPYSPGQAVAMKRMGIEGIYLGGWATSAKGSTTEDPGPDLASYPLSQV

PDDAAVLVRALLTADRNQQYQRLQMSEKQRAAAPVYDYRPFIIADADTGHGGDPHVRNLIRR

FVEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDIMKVPGIIVARTDAEAA

NLLDSRADERDQPFLLGATNLKIPSYKSCFLALVRRFYELGVKDLNGHLLYALPEGEYAEATAWLE

RQGIQGLVSDAVNAWREDGQQSIDDLFDQVESRFVAAWEDDAGLMTYGEAVADVLAFAASE

GEPADMSAEEWREFAATASLYSARAKAKELGVDPGWDCELSKTPEGYYQIRGGIPYAIAKSLAA

APFADLLWMETKTADLDDARQFADAIHAKFPDQMLAYNLSPSFNWDTTGMTDEQMKQFPE

ELGKMGFVFNFITYGGHQIDGVAAEEFATSLQQDGMLALARLQRKMRLVESPYRTPQTLVGG

PRSDAALTASSGRTATTKSMGEGSTQHQHLVQTEVPKKLLEEWLAMWSENYDLGEKLRVQLR

PRRAGSDVLELGIYGNDDEQLANVVVDPIKDRHGRSILQVRDQNTFAEKLRQKRLMTLIHLWL

VHRFKADGVIYVTPTEDNLYQTSKMKSHGIFSEVYQEVGEIIVAEVNQPRIAELLKPDRVALRKLI

TKEGAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 105325

8.3.22. *Mycobacterium intracellulare*, IDH1

Protein size:	**Vector:**	**Tag:**
435 aa/48.90 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSSPGSKEAKIKVKGRVVELDGDEMTRVIWKLIKDQLILPHLDIDLDYYDLGIEHRDRTDDQVTI

DAAYAIKKHGVGVKCATITPDEARVSEFNLKKMWLSPNGTIRNILGGTIFREPIVISNVPRLVPG

WTKPIVIGRHAFGDQYRATNFKVDKPGTVAITFTPADGSEPMVHEMVSLPDDGGVVMGMYN

FKDSVRDFARASFKYGLNAKWPVYLSTKNTILKAYDGMFKDEFQRIYDEEFKEQFEAEGLTYEHR

LIDDMVAACLKWEGGYVWACKNYDGDVQSDAIAQGYGSLGLMTSVLMTADGKTVEAEAAH

GTVTRHFRQYQAGKPTSTNPIASIFAWTRGLAHRGKLDDTPEVVEFAQTLENVVVSTVESGKM

TKDLAILIGPDQEWQQSEEFLDSIAENLEKELASAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.23. *Mycobacterium intracellulare*, IDH2

Protein size:	**Vector:**	**Tag:**
767 aa/85.57 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAEKPTVIYTLTDEAPLLATYAFLPIVRAFAEPAGIEIKTSDISVAARILAEFPEHLTEEQRVPDNL

AELGRLTKLADTNIIKLPNISASVPQLIAAVKELQGKGFKIPDFPQSPKTDEDKEIRDRYAKCLGSA

VNPVLRQGNSDRRAPKAVKEYARKHPHSMAPWSPASRTHVATMRGGDFYHGEKSMTLDRA

RNVKMELETESGETIVLKPMVSLRNADVIDSMFMSKKALVEFYEEQMQDAYETGVMFSLHVK

ATMMKVSHPIVFGHAVKIFYKEAFAKHQELFDELGVDVNNGLVDLYSKIESLPASLHEEIIRDLH

ACHEHRPELAMVDSAKGITNFHSPSDVIVDASMPAMIRAGGKMWGADGRQKDTKAVNPES

TFSRIYQEIINFCKTHGQFDPTTMGTVPNVGLMAQQAEEYGSHDKTFEIPEGGVANIVDLDTGE

VLLTQNVEAGDIWRMPIVRDEAIRDWVKLAVTRARNSGMTVVFWLDTERPHEVELRKKVKEY

LKEHDTEGLDINIMPQVWAMRYTLERAMRGQDTIAATGNILRDYLTDLFPILELGTSAKMLSIV

PLMAGGGMYETGAGGSAPKHVHQLVEENHLRWDSLGEFLALGACFEDIGIKTDNERAKLLGK

TLDAAIGKLLDNNKSPSRKTGELDNRGSQFYLALYWAQELAAQNDDEDLRERFAALADSLAEN

EDTIVGELAEVQGESVDIGGYYYPDSEKTTAVMRPSKTFNEVLAAAQGAENLYFQ*SHHHHHH

DYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 77030

8.3.24. *Mycobacterium marinum*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.45 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSVVGTPKSAEQIQHDWDHNPRWKGITRTYTPQDVVALQGH

VVEEHTLARRGSEVLWEQLHDMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANL

SGHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVENDRSVDNWLVPIVADGEAGFGGALNV

YELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVIA

RTDAEAATLITSDVDERDRPFITGERTKEGFYRIKNGLEPCIARAKAYAPFSDLIWMETGTPDLKL

AKQFAEGVKAEFPDQMLAYNCSPSFNWKKHLDDETIAKFQKELGAMGFKFQFITLAGFHALNY

SMFDLAYGYARNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDRIATTVDPTSSTTALA

GSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 66475

8.3.25. *Mycobacterium marinum*, ICL2

Protein size:	**Vector:**	**Tag:**
784 aa/88.01 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIAETDTEVRTPFEQEVADTQRYFDSSRFAGIVRLYTARQVVEQRGTIPNDYTVARTAAGAFY

QRLRELYTEKKSVTTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSTTEDPGPDLASYPLSQV

PDDAAVLVRALLTADRNQEYLRLQMTDQQRATTTAYDFRPFIIADADTGHGGDPHVRNLIRRF

VEVGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDVMRVPGIIVARTDAEAA

NLIDSRADERDQPFLLGATKLDIPSYKSCFLAMLRRFYELGVKELNGHLLYALADAEYAAANGWL

ERSGIMGLISDAVNSWREDGMQTIDDLFDQVESRFVAAWEDDAGLMTYAEAVAEVLAFGQS

EGDPVAMSPEEWRTFAARASLYAARQKAKELGVDPGWDCELAKTPEGYYQIRGGIPYAIAKSL

AAAPFADILWMETKTADLADARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMKRFP

EELGKMGFVFNFITYGGHQIDGVASEEFATALRQDGMLALARLQRKMRLVESPYRTPQTLVGG

PRSDAALQASSGRTATTKAMGKGSTQHQHLVQTEVPKKLLEEWLALWSDHYQLGEKLRVQLR

PRRAGSDVLELGIYGNDDEQLANVVFDPIKDRHGRSILTVRDQNTFAHKLRQKRLMTLIHLWLV

HRFKADAVYYVTPTEDNQYQTSKMKSHGIFSEVNQDVGEIIVAEVNRPRIEELLTADRVALRQLI

TKGDAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 105325

8.3.26. *Mycobacterium marinum*, IDH1

Protein size:	**Vector:**	**Tag:**
430 aa/48.42 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSGEPKITVSGPVVELDGDEMTRVIWKLIKDRLILPHLDVKLLYYDLGIEHRDRTDDQVTIDAAY

AIKKHGVGVKCATITPDEARVREFNLKKMWLSPNGTIRNILGGTIFREPIVISNVPRLVPGWTKPI

VIGRHAFGDQYRATNFKVDRPGTVTLTFTPADGSEPIVHEMVSIPEDGGVVMGMYNFRKSIQ

DFARASFSYGLNAKWPVYLSTKNTILKAYDGMFKDEFQRIYEEEFKERFDAEGLTYEHRLIDDMV

AACLKWEGGYVWACKNYDGDVQSDTVAQGYGSLGLMTSVLMTADGKTVEAEAAHGTVTRH

YRQYQAGQPTSTNPIASIFAWTRGLAHRGKLDNTPEVVQFAQTLEDVVIKTVESGKMTKDLAIL

IGPQQAWLNSEEFLEAIAENLAKKLGAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 69455

8.3.27. *Mycobacterium marinum*, IDH2

Protein size:	**Vector:**	**Tag:**
767 aa/85.18 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAEQPTIIYTLTDEAPLLATYAFLPIVRAFAEPAGIEIKTSDISVAARILAEFPDYLTDEQRVPDNL

AELGELTQDPDTNIIKLPNISASVPQLKAAIKELQDKGYAIPDYPADPKTDEEKALKERYARCLGS

AVNPVLRQGNSDRRAPNAVKEYARKHPHSMGEWSMASRTHVAHMRHGDFYAGEKSITLDR

AHKVKMELVTKDGQTLVLKPEVSLDDGDIIDSMFMSKKALCEFYEEQIEDAYKTGVMFSLHVKA

TMMKVSHPIVFGHAVKTFYRDAFAKHQKLFDELGVNVNNGLSDLYSKIETLPASQHDEIIDDLH

RCHEHRPELAMVDSARGITNFHSPSDVIVDASMPAMIRAGGKMYGADGKLKDTKAVNPESTF

SRIYQEIINFCKTNGQFDPTTMGTVPNVGLMAQQAEEYGSHDKTFEVPEDGVANIVDLDTGEV

LLTQDVEAGDIWRMCVVKDAPIRDWVKLAVTRARNSGMPVLFWLDPYRPHENELIKKVKTYL

KDHHTEGLDIQIMSQVRSMRYTLERLIRGLDTIAATGNILRDYLTDLFPILELGTSAKMLSIVPLM

AGGGMYETGAGGSAPKHVKQLVEENHLRWDSLGEFLALGASLEDMGIKIGNDRAKILARTLD

DAIGKLLGNNKSPSRKTGELDNRGSQFYLALYWAQELAEQTDDAELAEHFAKLAKALADGEEAI

VAELAEAQGEAVDIGGYYAPDNEKTTAVMRPSKTLNAALEAAQGAENLYFQ*SHHHHHHDYK

DDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 71990

8.3.28. *Mycobacterium paraffinicum*, ICL2

Protein size:	**Vector:**	**Tag:**
784 aa/87.80 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MAIIDSDTEVRTSFEDDIAATQRYMDDPRFTGITRLYTARQVVEQRGTIPTDYPVARNAAAAFY

DRLRELFAAKKSITTFGPYSPGQAVAMKRMGIEGIYLGGWATSAKGSTTEDPGPDLASYPLSQV

PDDAAVLVRALLTADRNQQYQRLNMTERQRATTPAYDYRPFIIADADTGHGGDPHVRNLIRRF

VEVGVAGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNAARFQLDVMKVPGIIVARTDAEAA

NLIDSRADERDQPFLLGATNLNIPSYKACFLALVRRFYELGVKDLNGHLLYALADSEYAAAGAWL

ERQGIQGLISEAVNALREDGHGSIDDLFDQVESRFVAAWEDDAGLMTYSEAVAEVLSFGETED

ESPSMTPDEWREFAARASLYAAREKAKELGVDPPWDPELSKTPEGYYQIRGGIPYAIAKSLAAA

PFADILWMETKTADLADARQFADAIHAEFPEQMLAYNLSPSFNWDTTGMSDEEMKRFPEELG

KMGFVFNFITYGGHQIDGVAAEEFATSLQQDGMLALARLQRKMRLVESPYRTPQTLVGGPRS

DAALQASSGRTATTKAMGKGSTQHQHLAQTEVPKKLLEEWLALWSENYRLGEKLRVTLRPRR

AGSDVLELGIYGNGNEQLANVVVDPIKDRHGRSILQVRDQNTFAEKLRQKRLMTLIHLWLVHR

FKADAVYYVTPTEDNLYQTEKMKSHGIFSEVYQEVGEIIVAEVNKPRIAELLTPDRVALRKLITKE

DAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 101315

8.3.29. *Mycobacterium paraffinicum*, IDH1

Protein size:	**Vector:**	**Tag:**
430 aa/48.55 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSTDKIKVKGPIVELDGDEMTRVIWKFIKDLLILPHLDIHLDYYDLGIENRDRTNDQVTIDAAYAI

KRHGVGVKCATITPDEARVSEFNLKKMWLSPNGTIRNILGGTIFREPIVISNVPRLVPGWTKPIVI

GRHAFGDQYRATNFKVDKPGTVSITFTPSDGGKPMVHEVVSIPEDGGVVMGMYNFKESVRDF

ARASLAYGLNAKWPVYLSTKNTILKAYDGMFKDEFQRIYDEEFKDKFEAEGLTYEHRLIDDMVA

ACLKWEGGYVWACKNYDGDVQSDLVAQGYGSLGLMTSVLMTADGKTVEAEAAHGTVTRHF

RQYQAGKPTSTNPIASIFAWTRGLQHRGKLDHTPEVIEFAQTLENVVVETVESGKMTKDLAILIG

PDQEWQQSEEFLNSIAENLEKKLVSAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.30. *Mycobacterium salmoniphilum*, ICL1

Protein size:	**Vector:**	**Tag:**
448 aa/48.86 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSNVGKPRTAAEIQQDWDTNPRWKGITRDYTAAQVEELQGS

VVEENTLARRGAEILWEGVTKGDGSYINALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANL

SGHTYPDQSLYPANSVPAVVRRINNALLRADEIARVEGDTSIDNWLVPIVADGEAGFGGALNVY

ELQKAMIAAGAAGTHWEDQLASEKKCGHLGGKVLIPTQQHVRTLTSARLAADVANTPTVVIAR

TDAEAATLITSDVDERDQPFVTGERTAEGFYNVQKGIEPCIARAKAYAPYADLIWMETGVPDLA

VARKFAEAVKAEFPDQLLSYNCSPSFNWKQALDDSTIAKFQKELGAMGFTFQFITLAGFHALNY

SMFDLAYGYAREGMTAYVDLQEREFAAEARGYTATKHQREVGAGYFDRIATTVDPNTSTAALK

GSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 69455

8.3.31. *Mycobacterium salmoniphilum*, IDH2

Protein size:	**Vector:**	**Tag:**
767 aa/84.68 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAQQPTIIYTLTDEAPLLATYSFLPIIRAFAGPAGIDVQTSDISVASRILSEFADYLPEDQRVPDNL

AELGRLTQLPETNIIKLPNISASVPQLLAAIKELQGKGYKLPDYPGDAKSDEEKDIVARYTKCLGSA

VNPVLREGNSDRRAPRAVKEYARKHPHSMGQWSQASRTHVATMKEGDFYHGEKSLTLDRAR

KVKMVLTTATGEATVLKAEVKLGEGDIIDSMFMSKKALCDFYEEQIEDAYKTGVMLSLHVKAT

MMRVSHPIVFGHAVKIFYKEAFAKHQELFDELGVNVNNGMSDLYSKIEALPASQHEEIIRDLHA

CHEHRPELAMVDSAKGISNFHSPSDVIVDASMPAMIRAGGKMYGADGKLKDTKAVNPESTFS

RIYQEMINFCKTHGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEITEAGVADIVDIDTGEVL

LSQNVEAGDIWRMPVVTDAAIQDWVKLAVTRARESGMTVVFWLDTERPHENELRKKVKTYL

KDHDTEGLEIQIMPQVWAIRYTMERVIRGLDTVSATGNILRDYLTDLFPILELGTSAKMLSIVPL

MAGGGLYETGAGGSAPKHVSQLVEENHLRWDSLGEFLALAVSLEDLGIKTGNPRAKILATTLDA

ATGKLLENDKGPSRKAGELDNRGSQFYLALYWASELAAQTEDAELAELFAPLAKRLVENEDAIV

AELASVQGSPADIGGYYQLDSEKTNAVMRPSQTFNAALASVEVAENLYFQ*SHHHHHHDYKD

DDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 76000

8.3.32. *Mycobacterium smegmatis*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/49.26 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSTVGTPKSPEQIQHDWDHNPRWKGIKRDYTPEDVVALQGT

VVEEHTLARRGAEVLWEQLHDMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANL

SGHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDTSVENWLAPIVADGEAGFGGALNV

YELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVADVPTVVIA

RTDAEAATLITSDVDERDQPFITGERTKEGFFRVKNGLEPCIARAKAYAPYSDLIWMETGTPDLE

LAKKFAEGVKAEFPDQMLAYNCSPSFNWKKHLDDATIAKFQKELGAMGFKFQFITLAGFHALN

YSMFDLAYGYARNQMSAYVELQEREFAAEERGYTATKHQREVGAGYFDRIATTVDPNSSTTAL

AGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 66475

8.3.33. *Mycobacterium smegmatis*, ICL2

Protein size:	**Vector:**	**Tag:**
791 aa/88.10 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MATIEANAHNLSSSAKDFDSEVAATQAYFDSPRFEGITRLYSARQVVEQRGTIATDHTVAREAA

EAFHPYLRELFAQKKSITTFGPYSPGQAVVMKRMGIGGIYLGGWATSAKGSISEDPGPDLASYP

LSQVPDEAAGLVRALLTADRNQHYLRLQMTPEQREAQPPVDYRPFIIADADTGHGGDPHVRN

LIRRFVESGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDIMRVPGIIVARTDA

EAANLIDSRADERDQPFLLGVTNLKIPSYKSCFLAMMRRFYDQGLTEINGHLLYDLPEGQYEAAE

AWLTGAGIMDEIAAAVESWTSGAASSVDALYDAVESKFVDAWQDDAGLNTYGEAVAEVIEFR

EREGESVAMSAAEWREFATRAPLYTATNKARELGVDVAWDCEPAKTPEGYYQVRGGIPYAIAK

SLAAAPFADILWMETKTADLADAKQFADAIHAVYPDQMLAYNLSPSFNWDTTGMTDDEM

RAFPEELGKMGFVFNFITYGGHQVDGVACEEFATSLRQEGMLALARLQRKMRLVESPYRTPQT

LVGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPKKLLEEWLAMWSEHYGLPEKLH

VQLRPRRAGSDVLDLQILGEGGDGTDPLANVVVDPIKDRHGRSILTVRDQNTFAEQMRKKRL

MDLIHLWLIHRFKAEIVYYVTPTEDNIYQTEKMKAHGIFSDVYQEVGEIIVADVNKPRIEELLASD

REALGRLIRKEDAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 105450

8.3.34. *Mycobacterium smegmatis*, IDH2

Protein size:	**Vector:**	**Tag:**
765 aa/85.38 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MSAQQPTIIYTLTDEAPLLATYAFLPVVRKFAEAAGIDVKTSDISVAARILAEFGDHLTEEQRVPD

NLGELGALTQDPSANIIKLPNISASVPQLLAAIKELQGKGYNVPDYPANPKTDDEKKIKDRYAKIL

GSAVNPVLREGNSDRRAPKAVKEYARKHPHSMGEWSQASRTHVATMKTGDFYHGEKSMTL

DRDRRVKMVLKTKSGEEIVLKPEVKLDAGDIIDSMYMSKKALIAFYEEQIEDAYKTGVMFSLHVK

ATMMKVSHPIVFGHAVKVFYKDAFAKHEKLFDELGVNVNNGLSDLYDKIEALPASQREEIIEDL

HKCHEHRPELAMVDSAKGISNFHSPSDVIVDASMPAMIRLGGKMYGADGRTKDTKAVNPEST

FSRMYQEMINFCKTHGQFDPTTMGTVPNVGLMAQKAEEYGSHDKTFEIPEDGVADIVDIDTG

EVLLTQNVEEGDIWRMPIVKDAPIRDWVKLAVTRARLSGMPVVFWLDTERPHEVELRKKVKEY

LKDHDTEGLKIQIMPQVWAMRYTLERVVRGKDTIAATGNILRDYLTDLFPILELGTSAKMLSIVP

LMAGGGLYETGAGGSAPKHVHQLVEENHLRWDSLGEFLALGASLEDMGNKTGNEKAKVLAK

ALDTATGKLLEENKSPSRRTGELDNRGSQFYLSLFWAQALAEQTEDAELAERFKPLAKALAEQEE

AIVSELNSVQGKTVDIGGYYYPDPEKTSEVMRPSKTFNTTLESVAENLYFQ*SHHHHHHDYKDD

DDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 75875

8.3.35. *Mycobacterium terrae*, ICL1

Protein size:	**Vector:**	**Tag:**
446 aa/48.98 kDa	pNIC-NTH	*N*-terminal His_6

Protein sequence:

MHHHHHHSSGTENLYFQ*SMSTVGTPKSPEQIQHDWDHNPRWEGVTRTYSAEDVVALQGH

VVEEHTLARRGAEVLWEQLHELDYINALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLS

GHTYPDQSLYPANSVPQVVRRINNALLRADEIAKVEGDTSVENWLAPIVADGEAGFGGALNVY

ELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQHIRTLTSARLAADVAGVPTVVIAR

TDAEAATLITSDVDERDQPFITGDRTAEGFYRVRNGLEPCIARAKAYAPYADLIWMETGTPDLEL

AAKFAEGVKAEFPDQMLAYNCSPSFNWRQHLDDATIAKFQNELGAMGFKFQFITLAGFHALN

YSMFDLAHGYARNQMSAYVELQEREFAAEARGYTATKHQREVGAGYFDRIATTVDPTSSTTAL

AGSTEEGQFH

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 67965

8.3.36. *Mycobacterium terrae*, ICL2

Protein size:	**Vector:**	**Tag:**
787 aa/87.25 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MTTIETTTAASAPVQPDVDADIVATQQYFDSPRFEGITRLYTARQVVEQRGTIPADYPVAREAA

TAFYARLRELFAAKKSITTFGPYSPGQAVVMKRVGIEGIYLGGWATSAKGSISEDPGPDLASYPLS

QVPDEAAGLVRALLTADRNQQYMRLRMTDEQRAATPAVDYRPFIIADADTGHGGDPHVRNLI

RRFVEAGVPGYHIEDQRPGTKKCGHQGGKVLVPSDEQIKRLNTARFQLDMMGVPGIIVARTD

AEAANLVESRSDERDQPFILGTTNLKIPSYKACFLAMTRRFHSSGITEINGHLLYALPEGEYAEAD

AWLERQGITALITKAAQGWRSDGRQSVDAVFDEIESAFVEAWQNDAGLMTFGDAVAEMLDF

AEREGEPKEMTAAQWREFAATAPFYTARAKAAELNADPGWDCDRAKTPEGYFQVRGGIPYAI

AKSLAAAPFADILWMETKTADLAEAREFAEAIHAVYPDQMLAYNLSPSFNWDTTGMTDDQM

RAFPAEIGKLGFVFNFITYGGHQVDGVASEEFAATLLQDGMLALARLQRKMRLVESPYRTPQTL

VGGPRSDAALAASSGRTATTKAMGAGSTQHQHLVQTEVPKKLLEDWLALWSEHYQLGEKLRA

QLRPTRPGSDVLELGIYGDGEEKLANVIVDPIKDRHGRSILTVRDQNTFSEKLRQKRLMDLTHL

WLIHRFKAAAVHYVTPTQDNLYQTEKMKEHGIFSSVNQEVGEIIVADVNQPRIDELLAPDRVAL

GKLISKDAENLYFQ*SHHHHHHDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 99365

8.3.37. *Mycobacterium terrae*, IDH2

Protein size:	**Vector:**	**Tag:**
766 aa/84.85 kDa	pNIC-CTHF	*C*-terminal His_6FLAG

Protein sequence:

MCADQPTIIYTLTDEAPLLATYSFLPIVQAFTGPADIDVKPSDISLGARILAEFPDFLTEEQRVPNN

LAELGRLTQLPDTTIIKLPNISASVPQLVAAIKELQGKGFNLPDYPEDPKTDDERDIHDRYAKCLGS

AVNPVLREGNSDRRAPKAVKEYARKHPHSMGEWSQASRTHVATMRHGDFYHGEKSMTLDR

DRDVKMELKTESGKTIVLKEKVSLLDGMVIDSMFMSVRALREFYEEQMEDAHKTGVMFSLHV

KATMMKVSHPIVFGHAIKVFYKDAFAKHQKLFDDLGVNVNNGLSDLYNKIEALPASQREEVIED

LHACHEHRPKLAMVDSANGISNFHSPSDVIVDASMPAMIRAGGKMYGADGRPRDTKAVNPE

STFARIYQEMINFCKTHGQFDPATMGTVPNVGLMAQQAEEYGSHDKTFEVPESGVADIVDLA

TGEVLLTQNVEAGDIWRLCTVTDAAIRDWVKLAVQRARASGMTAVFWLDTERPHEAELRKKV

KAYLLDHDTEGLEIQIMPQVWAMRYTLERLIRGQDTIAVTGNILRDYLTDLFPILELGTSAKMLSI

VPLMAGGGMYETGAGGSAPKHVKQLVEENHLRWDSLGEFLALAASFDDLGAKTGNPNAKIL

GSTLDAAIGKLLDANKSPSRKTGELDNRGSQFFLAKYWAEELAAQTEVPELAKQFVAVAEALEK

NQDAILTELTVVQGEPVDIGGYYAPDHDMTAAVMRPSSTLNAVIDAKGAENLYFQ*SHHHHH

HDYKDDDDK

Extinction coefficient (280 nm)/M^{-1} cm^{-1}: 70040

8.4. Protein Expression

8.4.1. Transformation: *E. coli* expressing strains

The protocol below was used to transform the following *E. coli* expressing strains: BL21(DE3)Star (prepared by Acely Garza-Garcia), NiCo21(DE3) (NEB), BL21-Gold(DE3) (Agilent), Rosetta(DE3) (Novagen), and BL21(DE3)pLysS (Novagen). **Table 47** lists the strain(s) used to express each protein in this study. Glycerol stocks were prepared for each of the resulting expressing strains.

E. coli cells (25 µL) were thawed on ice. Plasmid (1 µL, appropriate construct from **Section 8.3**) was added and mixed gently with the cells. The cells were kept on ice for 30 minutes, then heat shocked at 42 °C (30 seconds for BL21(DE3)Star and 90 seconds for the other strains listed). Cells were returned to the ice for 2 minutes prior to addition of SOC outgrowth medium (950 µL). Cultures were incubated at 37 °C, with shaking, for 1 hour. 100 µL of culture was plated onto LB agar containing 50 µg/mL kanamycin. The plates were incubated at 37 °C overnight or at room temperature over the weekend.

Table 47: *E. coli* strains used to express each protein in this study.
Mtb His_6-ICL1 was expressed in multiple strains during method development (**Chapter 4, Section 4.2.1.6**). All Mtb His_6-ICL1 used in kinetic assays was expressed in BL21(DE3)Star. Other mycobacterial enzymes were expressed in either BL21(DE3)Star or BL21-Gold(DE3). The choice of strain was determined by availability in our laboratory at the time of the experiment.

Species	Protein	Expressing strain
M. tuberculosis CDC 1551	His_6-ICL1 (during method development)	BL21(DE3)Star BL21-Gold(DE3) NiCo21(DE3) Rosetta(DE3) BL21(DE3)pLysS
	His_6-ICL1 (for kinetics)	BL21(DE3)Star
M. tuberculosis CDC 1551	ICL2-His_6FLAG	BL21(DE3)Star
M. tuberculosis CDC 1551	ICL2(T605)-His_6FLAG	BL21(DE3)Star

M. tuberculosis CDC 1551	IDH1-His_6FLAG	BL21(DE3)Star
M. tuberculosis CDC 1551	IDH2-His_6FLAG	BL21(DE3)Star
M. aromaticivorans	His_6-ICL1	BL21-Gold(DE3)
M. aurum	His_6-ICL1	BL21-Gold(DE3)
M. chubuense	His_6-ICL1	BL21-Gold(DE3)
M. fallax	His_6-ICL1	BL21(DE3)Star
M. hiberniae	His_6-ICL1	BL21(DE3)Star
M. intracellulare	His_6-ICL1	BL21-Gold(DE3)
M. marinum	His_6-ICL1	BL21(DE3)Star
M. salmoniphilum	His_6-ICL1	BL21(DE3)Star
M. smegmatis	His_6-ICL1	BL21(DE3)Star
M. terrae	His_6-ICL1	BL21(DE3)Star
M. aromaticivorans	ICL2-His_6FLAG	BL21(DE3)Star
M. aurum	ICL2-His_6FLAG	BL21(DE3)Star
M. chubuense	ICL2-His_6FLAG	BL21(DE3)Star
M. hiberniae	ICL2-His_6FLAG	BL21-Gold(DE3)
M. intracellulare	ICL2-His_6FLAG	BL21-Gold(DE3)
M. marinum	ICL2-His_6FLAG	BL21(DE3)Star
M. paraffinicum	ICL2-His_6FLAG	BL21(DE3)Star
M. smegmatis	ICL2-His_6FLAG	BL21-Gold(DE3)
M. terrae	ICL2-His_6FLAG	BL21(DE3)Star
M. intracellulare	IDH1-His_6FLAG	BL21-Gold(DE3)
M. marinum	IDH1-His_6FLAG	BL21(DE3)Star
M. paraffinicum	IDH1-His_6FLAG	BL21(DE3)Star
M. aromaticivorans	IDH2-His_6FLAG	BL21(DE3)Star
M. aurum	IDH2-His_6FLAG	BL21(DE3)Star
M. chubuense	IDH2-His_6FLAG	BL21-Gold(DE3)
M. fallax	IDH2-His_6FLAG	BL21-Gold(DE3)
M. hiberniae	IDH2-His_6FLAG	BL21-Gold(DE3)
M. intracellulare	IDH2-His_6FLAG	BL21-Gold(DE3)

M. marinum	IDH2-His_6FLAG	BL21(DE3)Star
M. salmoniphilum	IDH2-His_6FLAG	BL21-Gold(DE3)
M. smegmatis	IDH2-His_6FLAG	BL21(DE3)Star
M. terrae	IDH2-His_6FLAG	BL21(DE3)Star

8.4.2. Protein expression

Protein expression was carried out using various culture volumes depending on the desired yield. 100 mL cultures were adequate for the purification and characterisation of His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs. Where larger culture volumes were used, this was simply for convenience (for example to provide sufficient cells for multiple purification experiments). 1 L cultures were required for the purification and characterisation of IDH1-His_6FLAGs.

The target optical density at 600 nm (OD_{600}) for induction was 0.6—0.7. However, cultures could be induced at up to OD_{600} 1.0 without adverse effects. The final IPTG concentration was 0.5 mM for ICLs and 0.1 mM for IDHs. After induction, ICL cultures were incubated at 18 °C overnight, while IDH cultures were incubated at 16 °C overnight. Cell pellets for use in protein purification were weighed and stored dry at −20 °C for a maximum of one year.

8.4.2.1. 1.5 mL culture volume

1.5 mL cultures were used for protein expression tests.

For each expressing strain, a single colony or a flake from a glycerol stock was used to inoculate 10 mL of LB-kanamycin medium. Starter cultures were incubated at 37 °C

overnight, with shaking. The OD_{600} of each culture was measured and the appropriate inoculum volume was calculated (0.15 OD_{600}, to give an inoculated culture containing 0.1 OD_{600}/mL). The appropriate volume of each starter culture was then used to inoculate 1.5 mL autoclaved LB-kanamycin medium in a 12-well plate. The inoculated plate was incubated at 37 °C, with shaking. After approximately 4 hours, the plate was cooled to 4 °C and each well was induced with IPTG (final IPTG concentration 0.5 mM for ICLs and 0.1 mM for IDHs). The induced plate was incubated overnight, with shaking (18 °C for ICLs and 16 °C for IDHs).

1 mL of culture was removed from each well and transferred to a 1.5 mL Eppendorf tube for LDS-PAGE analysis. The remaining culture was discarded.

8.4.2.2. 100 mL culture volume

100 mL cultures were used for protein expression tests and to express His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs for purification.

A single colony or a flake from a glycerol stock was used to inoculate 10 mL of LB-kanamycin medium. Starter cultures were incubated at 37 °C overnight, with shaking, then centrifuged at 3,200 × *g* for 10 minutes. The supernatant was discarded and the pellet was resuspended in 10 mL of LB-kanamycin medium. 2 mL of the suspension was used to inoculate 100 mL of autoclaved LB-kanamycin medium in a 500 mL baffled flask. Cultures were incubated at 37 °C, with shaking. When the OD_{600} reached ~0.6, the cultures were cooled on ice and induced with IPTG (final IPTG concentration 0.5 mM for ICLs and 0.1 mM for IDHs). Induced cultures were incubated overnight, with shaking (18 °C for ICLs and 16 °C for IDHs).

1 mL of each culture was transferred to a 1.5 mL Eppendorf tube for LDS-PAGE analysis. Cultures were then divided equally between two 50 mL Falcon tubes and centrifuged at 3,220 × *g* for 30 minutes. The supernatant was discarded and the cell pellets were stored at −20 °C.

8.4.2.3. 200 mL culture volume

200 mL cultures were used to express His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs for purification.

A single colony or a flake from a glycerol stock was used to inoculate 10 mL of LB-kanamycin medium. Starter cultures were incubated at 37 °C overnight, with shaking, then centrifuged at 3,200 × *g* for 10 minutes. The supernatant was discarded and the pellet was resuspended in 10 mL of LB-kanamycin medium. 4 mL of the suspension was used to inoculate 200 mL of autoclaved LB-kanamycin medium in a 500 mL baffled flask. Cultures were incubated at 37 °C, with shaking. When the OD_{600} reached ~0.6, the cultures were cooled on ice and induced with IPTG (final IPTG concentration 0.5 mM for ICLs and 0.1 mM for IDHs). Induced cultures were incubated overnight, with shaking (18 °C for ICLs and 16 °C for IDHs).

1 mL of each culture was transferred to a 1.5 mL Eppendorf tube for LDS-PAGE analysis. 50 mL of culture was transferred to each of two 50 mL Falcon tubes. The tubes were centrifuged at 3,220 × *g* for 30 minutes. The supernatant was discarded and a further 45 mL of culture was added to each tube (such that each tube contained cells from 95 mL of culture). The tubes were centrifuged at 3,220 × *g* for 30 minutes. The supernatant was discarded and the cell pellets were stored at −20 °C.

8.4.2.4. 500 mL culture volume

500 mL cultures were used to express Mtb CDC 1551 His_6-ICL1, ICL2-His_6FLAG and IDH2-His_6FLAG for purification.

A single colony or a flake from a glycerol stock was used to inoculate 10 mL of LB-kanamycin medium. Starter cultures were incubated at 37 °C overnight, with shaking, then centrifuged at 3,200 × *g* for 10 minutes. The supernatant was discarded and the pellet was resuspended in 10 mL of LB-kanamycin medium. The suspension was used to inoculate 500 mL of autoclaved LB-kanamycin medium in a 2 L baffled flask. Cultures were incubated at 37 °C, with shaking, to OD_{600} ~0.6. Cultures were then cooled on ice and induced with IPTG (final IPTG concentration 0.5 mM for ICLs and 0.1 mM for IDHs). Induced cultures were incubated overnight, with shaking (18 °C for ICLs and 16 °C for IDHs).

1 mL of each culture was transferred to a 1.5 mL Eppendorf tube for LDS-PAGE analysis. Cultures were then transferred to 1 L centrifuge bottles and centrifuged at 4,544 × *g* for 30 minutes. The supernatant was discarded. Each pellet was resuspended in 25 mL PBS and distributed equally between five 50 mL Falcon tubes. These tubes were centrifuged at 3,220 × *g* for 30 minutes. The supernatant was discarded and the cell pellets were stored at −20 °C.

8.4.2.5. 1 L culture volume

1 L cultures were used to express IDH1-His_6FLAGs for purification.

Single colonies or flakes from a glycerol stock were used to inoculate LB-kanamycin medium (2 × 10 mL). Starter cultures were incubated at 37 °C overnight, with shaking,

then centrifuged at 3,220 × *g* for 10 minutes. The supernatant was discarded and each pellet was resuspended in 10 mL of LB-kanamycin medium. The resulting 20 mL suspension was used to inoculate 1 L of autoclaved LB-kanamycin medium in a 2 L baffled flask. Cultures were incubated at 37 °C, with shaking, to OD_{600} ~0.6. Cultures were then cooled on ice and induced with IPTG (final IPTG concentration 0.1 mM). Induced cultures were incubated at 16 °C overnight, with shaking.

1 mL of each culture was transferred to a 1.5 mL Eppendorf tube for LDS-PAGE analysis. Cultures were then transferred to 1 L centrifuge bottles and centrifuged at 4,544 × *g* for 30 minutes. The supernatant was discarded. Each pellet was resuspended in 35 mL PBS, transferred to a 50 mL Falcon tube, and centrifuged at 3,220 × *g* for 30 minutes. The supernatant was discarded and the cell pellets were stored at −20 °C.

8.4.3. LDS-PAGE analysis of protein expression

The 1 mL culture samples retained after protein expression were analysed by lithium dodecyl sulfate polyacrylamide gel electrophoresis (LDS-PAGE). The samples were centrifuged at 21,100 × *g* for 3 minutes. The supernatant was discarded and each pellet was resuspended in 300 µL lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8). Cells were lysed by sonication (40 × 1 second pulses with 12 second breaks between pulses). Samples were prepared for LDS-PAGE analysis as described below.

Total expression: Lysate (13 µL), TCEP (2 µL, 0.5 M, pH 8.0) and NuPAGE LDS sample buffer (5 µL) were added to a 0.5 mL Eppendorf tube and mixed by vortexing. Samples were centrifuged at 21,100 × *g* for 1 minute, then incubated at 70 °C for 10 minutes.

Samples were again centrifuged at 21,100 × *g* for 1 minute, then vortexed. 10 µL of each sample was used for LDS-PAGE.

Soluble fraction: Lysates were centrifuged at 21,100 × *g* for 3 minutes to precipitate the insoluble components. The supernatant contained the soluble proteins. Supernatant (13 µL), TCEP (2 µL, 0.5 M, pH 8.0) and NuPAGE LDS sample buffer (5 µL) were added to a 0.5 mL Eppendorf tube and mixed by vortexing. Samples were centrifuged at 21,100 × *g* for 1 minute, then incubated at 70 °C for 10 minutes. Samples were again centrifuged at 21,100 × *g* for 1 minute, then vortexed. 10 µL of each sample was used for LDS-PAGE.

LDS-PAGE was performed using NuPAGE 4—12% Bis-Tris 15 well 1 mm precast gels and NuPAGE MES running buffer. Ladder (Mark12 Unstained Standard, 5 µL) and samples (10 µL) were loaded and the gel was run according to the supplied directions. Gels were stained with Instant Blue™ Coomassie Protein Stain (Expedeon) or Quick Coomassie Stain (Neo Biotech) and photographed using a BioDoc-It^2 Imager (UVP) under UV light.

8.5. Protein Purification

8.5.1. General notes

All sample preparation and protein purification steps were performed at 4 °C. Buffer pHs were adjusted at room temperature using aqueous solutions of HCl or NaOH, as appropriate. Buffers for fast protein liquid chromatography (FPLC) were filtered and degassed under vacuum before use. Lysis buffers were not degassed. All buffers were cooled to 4 °C before use and stored at 4 °C.

Immobilised metal affinity chromatography (IMAC), anion exchange chromatography (AIEX), and size exclusion chromatography (SEC) were performed on an ÄKTA pure 25 M FPLC system equipped with an F9-C fraction collector. Samples were applied to columns using a 5 mL tubing loop or a 50 mL Superloop, as appropriate. ÄKTA system equilibration, column equilibration, and post-run cleaning were performed under manual control. FPLC methods for IMAC, AIEX, and SEC were written in Unicorn 6.3. Sample loops were equilibrated with IMAC or AIEX binding buffer, as appropriate. During IMAC and AIEX runs, the flow rate was set to 1 mL/minute and subject to pre-column pressure control. Flow rates of 1—4 mL/minute were used during pre-run equilibration and post-run cleaning. Fractions were collected in 2 mL 96-well plates. The UV absorbance at 280 nm, the UV absorbance at 220 nm, and the conductivity were monitored during each run. CV, column volume; FF, fast flow; HP, high performance.

New IMAC and AIEX columns were used for each enzyme purified in this study. Columns were equilibrated according to the supplied directions. Where an enzyme was purified on multiple occasions, the IMAC column was reused without stripping or recharging, while the AIEX column (if required) was recharged according to the supplied directions. IMAC and AIEX columns were stored in 20% ethanol at 4 °C.

8.5.2. Immobilised metal affinity chromatography (IMAC)

IMAC methods are presented in the order developed in **Chapter 4**. The optimised conditions for the purification of each enzyme prepared in **Chapter 4** and **Chapter 5** are given in **Section 8.5.8**.

8.5.2.1. Sample preparation

Cell pellets were resuspended in IMAC lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8, containing 1× c*O*mplete EDTA-free protease inhibitors) and lysed by sonication (90—180 × 2 second pulses with 5 second breaks between pulses). Lysates were centrifuged at 48,298 × *g* for 30 minutes. The supernatant was passed through a 0.45 µm cellulose acetate syringe filter before application to the column.

8.5.2.2. Buffers

IMAC buffer A – binding buffer: 50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8

IMAC buffer B – elution buffer: 50 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8

8.5.2.3. IMAC Method 1

The sample was applied to a HisTrap FF 1 mL column equilibrated with IMAC buffer A. The column was washed with IMAC buffer A (20 CV) prior to gradient elution (0—100% IMAC buffer B over 5 CV). 2 mL fractions were collected during the sample application and wash steps. 1 mL fractions were collected during elution.

This method was also used to explore the effect of detergent (2% (*v/v*) Tween 20) and reducing agent (0.5 mM TCEP) on the purification. Detergent or reducing agent were added to both IMAC buffer A and IMAC buffer B at the specified concentrations.

8.5.2.4. IMAC Method 2

The sample was applied to a HisTrap FF 1 mL column equilibrated with IMAC buffer A. The column was washed with a high salt wash buffer (50 mM Tris, 500 mM NaCl, 25 mM imidazole, pH 8, 20 CV) prior to gradient elution (0—100% IMAC buffer B over 20 CV).

2 mL fractions were collected during the sample application and wash steps. 1 mL fractions were collected during elution.

8.5.2.5. IMAC Method 3

The sample was applied to a HisTrap HP 1 mL column equilibrated with IMAC buffer A. The column was washed with IMAC buffer A (10 CV), followed by 5% IMAC buffer B (10 CV) prior to gradient elution (5—100% IMAC buffer B over 15 CV). 2 mL fractions were collected during the sample application and wash steps. 1 mL fractions were collected during elution.

8.5.2.6. IMAC Method 4

The sample was applied to a HisTrap FF 1 mL column equilibrated with IMAC buffer A. The column was washed with IMAC buffer A (20 CV) prior to gradient elution (0—100% IMAC buffer B over 30 CV). 2 mL fractions were collected during the sample application and wash steps. 1 mL fractions were collected during elution.

8.5.2.7. IMAC Method 5

The sample was applied to a HisTrap HP 1 mL column equilibrated with IMAC buffer A. The column was washed with IMAC buffer A (15 CV), 5% IMAC buffer B (15 CV) and 15% IMAC buffer B (5 CV). Gradient elution was split into two steps: 15—70% IMAC buffer B over 20 CV, followed by 70—100% IMAC buffer B over 5 CV. 2 mL fractions were collected during the sample application and during the first two wash steps (IMAC buffer A and 5% IMAC buffer B). 1 mL fractions were collected during the third wash step (15% IMAC buffer B) and the second elution gradient (70—100% IMAC buffer B). 0.5 mL fractions were collected during the first elution gradient (15—70% IMAC buffer B).

8.5.2.8. IMAC Method 6

The sample was applied to a HisTrap HP 1 mL column equilibrated with buffer A. The column was washed with buffer A (15 CV), followed by 5% buffer B (15 CV). Gradient elution was split into two steps: 5—65% B over 20 CV, followed by 65—100% B over 5 CV. 2 mL fractions were collected during the sample application and wash steps. 0.5 mL fractions were collected during the first elution gradient (5—65% B). 1 mL fractions were collected during the second elution gradient (65—100% B).

8.5.3. Anion Exchange Chromatography (AIEX)

8.5.3.1. Sample preparation

Where AIEX was performed as the first step in protein purification, cell pellets were resuspended in AIEX lysis buffer (20 mM Tris, pH 8, containing 1× c*O*mplete EDTA-free protease inhibitors) and lysed by sonication (90 × 2 second pulses with 5 second breaks between pulses). Lysates were centrifuged at 48,298 × *g* for 30 minutes. The supernatant was passed through a 0.45 µm cellulose acetate syringe filter before application to the column.

Where AIEX was performed after IMAC, the IMAC fractions containing the protein of interest were pooled and applied to a Sephadex PD10 gravity desalting column equilibrated with AIEX buffer A (20 mM Tris, pH 8). Desalted protein was eluted with 3.5 mL AIEX buffer A and used for AIEX.

8.5.3.2. Buffers

AIEX buffer A – binding buffer: 20 mM Tris, pH 8

AIEX buffer B – elution buffer: 20 mM Tris, 1 M NaCl, pH 8

8.5.3.3. AIEX Method

The sample was applied to a HiTrap Q HP 1 mL column equilibrated with AIEX buffer A. The column was washed with AIEX buffer A (10 CV) prior to gradient elution (0—50% AIEX buffer B over 20 CV). Finally, the column was washed with AIEX buffer B (5 CV). 1 mL fractions were collected throughout the run.

8.5.4. Size Exclusion Chromatography (SEC)

SEC by FPLC was only used during method development for Mtb His_6-ICL1.

SEC was performed on IMAC fractions containing the protein of interest. A HiLoad 16/60 Superdex 200 column was equilibrated with MOPS buffer (50 mM, pH 6.7). The sample was applied to the column and eluted with 1 CV MOPS (50 mM, pH 6.7). 1.5 mL fractions were collected during sample application and elution.

8.5.5. LDS-PAGE analysis

The outcomes of chromatography steps were analysed by LDS-PAGE. Appropriate volumes of FPLC fractions to use for analysis were determined by consideration of the UV absorbance trace for the run (280 nm).

FPLC fractions (2—13 µL), TCEP (2 µL, 0.5 M, pH 8.0) and NuPAGE LDS sample buffer (5 µL) were added to a 0.5 mL Eppendorf tube. IMAC lysis buffer was added to give a total sample volume of 20 µL and the samples were mixed by vortexing. Samples were centrifuged at 21,100 × *g* for 1 minute, then incubated at 70 °C for 10 minutes. Samples were again centrifuged at 21,100 × *g* for 1 minute, then vortexed. 10 µL of each sample was used for LDS-PAGE.

LDS-PAGE was performed using NuPAGE 4—12% Bis-Tris 15 well 1 mm precast gels and NuPAGE MES running buffer. Ladder (Mark12 Unstained Standard, 5 µL) and samples (10 µL) were loaded and the gel was run according to the supplied directions. Gels were stained with Instant Blue™ Coomassie Protein Stain (Expedeon) or Quick Coomassie Stain (Neo Biotech) and photographed using a BioDoc-It^2 Imager (UVP) under UV light.

8.5.6. Buffer exchange

8.5.6.1. Dialysis

Buffer exchange by dialysis was used on multiple occasions during method development. Purified proteins were dialysed against buffer containing 50 mM HEPES and 150 mM NaCl (pH 7.0, 4 °C, 2 × 1000 mL for 2 hours, then 1 × 2000 mL overnight). Protein concentration measurements were performed directly after dialysis.

8.5.6.2. Gravity size exclusion chromatography

Gravity SEC was the buffer exchange method included in the optimised conditions for the purification of each enzyme (**Section 8.5.8**) and was therefore used in all purifications leading to the generation of kinetic parameters. Protein concentration measurements were performed directly after gravity SEC.

His_6-ICL1s, ICL2-His_6FLAGs, and IDH2-His_6FLAGs: Fractions containing the protein of interest (total sample volume 2.5 mL) were pooled and applied to a Sephadex PD10 gravity desalting column equilibrated with buffer containing 50 mM HEPES and 150 mM NaCl (pH 7.0). The flow-through was discarded. The protein of interest was eluted with buffer containing 50 mM HEPES and 150 mM NaCl (pH 7.0, 3.5 mL).

IDH1-His$_6$FLAGs: Fractions containing the protein of interest (total sample volume 2.5 mL) were pooled and applied to a Sephadex PD10 gravity desalting column equilibrated with buffer containing 50 mM HEPES, 150 mM NaCl, and 5 mM $MnCl_2$ (pH 7.0). The flow-through was discarded. The protein of interest was eluted with buffer containing 50 mM HEPES, 150 mM NaCl, and 5 mM $MnCl_2$ (pH 7.0, 3.5 mL). Inclusion of $MnCl_2$ was essential for stability of IDH1-His$_6$FLAGs.

8.5.7. Protein concentration measurements

A NanoDrop One spectrophotometer was used to measure the UV absorbance at 280 nm (A_{280}) and 250 nm (A_{250}). A_{280}/A_{250} was used as an indicator of sample quality. The protein concentration was then calculated using the A_{280} reading and the extinction coefficient at 280 nm (ε_{280}, calculated for each enzyme in **Section 8.3**):

$$[\text{Enzyme}] = \frac{A_{280}}{\varepsilon_{280}}$$

8.5.8. Optimised conditions for the purification of each enzyme

8.5.8.1. His_6-ICL1s

Mycobacterial His_6-ICL1s were purified in one step using **IMAC Method 3** (**Section 8.5.2.5**). The three fractions containing the most enzyme (as assessed by LDS-PAGE) were pooled and buffer exchange was performed using gravity SEC (**Section 8.5.6.2**). Protein concentration measurements were performed directly after gravity SEC; enzyme yields are given in **Table 48**. Buffer exchanged enzymes were stored at 4 °C. All assays were completed within 36 hours of cell lysis.

Table 48: Yields of His_6-ICL1s.
n, number of times purified.

Species	Yield (mg protein/g wet cells)	n
M. tuberculosis CDC 1551	0.80—5.58	8
M. aromaticivorans	7.79	1
M. aurum	9.13	1
M. chubuense	12.70	1
M. fallax	6.88	1
M. hiberniae	5.33—5.99	2
M. intracellulare	8.23	1
M. marinum	0.69	1
M. salmoniphilum	2.52—3.89	3
M. smegmatis	8.98	1
M. terrae	1.58	1

8.5.8.2. ICL2-His_6FLAGs

Mycobacterial ICL2-His_6FLAGs were purified in one step using **IMAC Method 3** (**Section 8.5.2.5**). The three fractions containing the most enzyme (as assessed by LDS-PAGE) were pooled and buffer exchange was performed using gravity SEC (**Section 8.5.6.2**).

Protein concentration measurements were performed directly after gravity SEC; enzyme yields are given in **Table 49**. Buffer exchanged enzymes were stored at 4 °C. All assays were completed within 36 hours of cell lysis.

Table 49: Yields of ICL2-His_6FLAGs.
n, number of times purified.

Species	Yield (mg protein/g wet cells)	n
M. tuberculosis CDC 1551, full length	8.97—28.28	8
M. tuberculosis CDC 1551, aa 1—605	13.15	1
M. aromaticivorans	4.39	1
M. aurum	25.12	1
M. chubuense	6.35	1
M. hiberniae	22.73	1
M. intracellulare	26.57	1
M. marinum	16.00—23.24	2
M. paraffinicum	24.33	1
M. smegmatis	24.76	1
M. terrae	9.22	1

8.5.8.3. IDH2-His_6FLAGs

Mycobacterial IDH2-His_6FLAGs were purified in one step using **IMAC Method 3** (**Section 8.5.2.5**). The three fractions containing the most enzyme (as assessed by LDS-PAGE) were pooled and buffer exchange was performed using gravity SEC (**Section 8.5.6.2**). Protein concentration measurements were performed directly after gravity SEC; enzyme yields are given in **Table 50**. Buffer exchanged enzymes were stored at 4 °C. All assays were completed within 36 hours of cell lysis.

Table 50: Yields of IDH2-His_6FLAGs.
n, number of times purified.

Species	Yield (mg protein/g wet cells)	n
M. tuberculosis CDC 1551	10.86—21.88	5
M. aromaticivorans	8.90	1
M. aurum	1.57	1
M. chubuense	12.14	1
M. fallax	11.38	1
M. hiberniae	8.32	1
M. intracellulare	14.04	1
M. marinum	9.89	1
M. salmoniphilum	6.03	1
M. smegmatis	0.15—11.18	3
M. terrae	20.64	1

8.5.8.4. IDH1-His_6FLAGs

Mycobacterial IDH1-His_6FLAGs were captured using **IMAC Method 6** (**Section 8.5.2.8**). Appropriate fractions (to a total sample volume of 2.5 mL) were pooled, desalted and further purified using AIEX (**Section 8.5.3**). All AIEX eluate fractions containing IDH1-His_6FLAG were immediately spiked with $MnCl_2$ to stabilise the enzyme (final [$MnCl_2$] 5 mM, this step was critical). The three fractions containing the most enzyme (as assessed by LDS-PAGE) were pooled and buffer exchange was performed using gravity SEC (with buffer containing $MnCl_2$, **Section 8.5.6.2**). Protein concentration measurements were performed directly after gravity SEC; enzyme yields are given in **Table 51**. Buffer exchanged enzymes were stored at 4 °C. All assays were completed within 36 hours of cell lysis.

Table 51: Yields of IDH1-His_6FLAGs.
n, number of times purified; pI, predicted isoelectric point calculated using ExPASy ProtParam.[543]

Species	pI	Yield (mg protein/g wet cells)	n
M. tuberculosis CDC 1551	5.23	0.08	1
M. intracellulare	5.25	0.59	1
M. marinum	5.78	0.68	1
M. paraffinicum	5.61	1.28	1

8.5.9. His_6FLAG tag cleavage: preparation of tag-cleaved Mtb CDC 1551 ICL2

This protocol relates to the experiment discussed in **Chapter 4, Section 4.3.3.2**. Of note, it has not been optimised, as the first attempt yielded sufficient ICL2 for the desired experiments.

Mtb CDC 1551 ICL2-His_6FLAG was purified as described in **Section 8.5.8.2**. To the gravity SEC eluate (ICL2-His_6FLAG in buffer comprising 300 mM NaCl and 50 mM Tris at pH 7) were added TEV protease (1:100 TEV protease:ICL2-His_6FLAG; TEV protease was prepared by Daniel Groom) and dithiothreitol (DTT; stock [DTT] 100 mM, pH 8; final [DTT] 1 mM). The reaction mixture was mixed gently (rocking by hand) and left to stand at 4 °C for 18.5 hours. There was no visible precipitate. The reaction mixture was centrifuged at 3,220 × *g* for 15 minutes; there was no visible pellet.

A new HisTrap HP 1 mL column equipped with a syringe adaptor was manually equilibrated (buffer comprising 300 mM NaCl, 50 mM Tris, pH 8, 5 mL). The centrifuged reaction mixture was loaded onto the column and the flow-through was collected in ~500 µL fractions. Two column wash steps were performed (300 mM NaCl, 50 mM Tris,

pH 8, 5 mL; then 300 mM NaCl, 50 mM Tris, 20 mM imidazole, pH 8, 5 mL), collecting ~500 µL fractions. All fractions were checked for protein content by measuring the absorbance at 280 nm using a NanoDrop One spectrophotometer; fractions containing protein were analysed by LDS-PAGE.

The six fractions with the highest absorbance at 280 nm (total volume 2.5 mL) were pooled. Buffer exchange was performed using gravity SEC, as described in **Section 8.5.6.2**. The gravity SEC eluate was used for the assays presented in **Chapter 4, Section 4.3.3.2**. To enable determination of the yield, the remaining fractions containing ICL2 (total volume 5.3 mL) were pooled and the protein concentration was measured. Overall (considering all fractions that contained ICL2), ICL2 was recovered by IMAC in 82% yield.

8.6. Identification of contaminant proteins by proteomics

During IMAC method development, the contaminant proteins that co-eluted with Mtb CDC 1551 His_6-ICL1 were identified by proteomics (**Chapter 4, Section 4.2.1.7**). Mtb CDC 1551 His_6-ICL1 expressed in *E. coli* BL21(DE3)Star was purified using **IMAC Method 1** (**Section 8.5.2.3**). Following LDS-PAGE analysis, bands of interest were excised from the gel and transferred to Eppendorf tubes containing sterile deionised water. Samples were submitted to the Mass Spectrometry Proteomics Science Technology Platform at the Crick, where trypsin digestion and mass spectrometry were performed by David Frith. David then searched the data against the UniProt all species database using Mascot, and provided us with a Scaffold file containing the results. We continued the analysis by considering the hits more abundant than trypsin that were identified as *E. coli* or Mtb proteins. For each band, the mass of each hit was compared to the mass

estimated by LDS-PAGE and hits that were clearly unrealistic were discarded. Putative assignments could then be made for each band, as discussed in **Chapter 4, Section 4.2.1.7**.

8.7. Enzyme Kinetics

8.7.1. General notes

ICL activities were measured by coupling to lactate dehydrogenase (LDH). LDH converts the isocitrate cleavage product glyoxylate to glycolate. The concomitant oxidation of NADH results in a decrease in the absorbance at 340 nm. Rabbit muscle LDH was supplied as a 5 mg/mL suspension in 3.2 M ammonium sulfate (Roche). IDH activities were measured directly. Reduction of the co-substrate $NADP^+$ results in an increase in the absorbance at 340 nm.

Assay reactions were performed in 1.5 mL (semi-micro) disposable methacrylate cuvettes with a path length of 10 mm (suitable for use between 285 and 750 nm). During each reaction, the absorbance at 340 nm was recorded using a Shimadzu UV-2550 or UV-1900 spectrophotometer equipped with a 6-cell thermoelectrically temperature-controlled cell positioner. Thermal equilibration of reaction mixtures was performed in the cell positioner or in a heat block.

8.7.2. His_6-ICL1 assays

Mycobacterial His_6-ICL1s were purified as described above (**Section 8.5.8.1**) and diluted in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MgCl_2$ at pH 7.0 to the concentrations specified in **Table 52**. The diluted enzyme stocks were used to initiate assay reactions. Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and varying concentrations of DL-isocitrate

(12 concentrations in the range 25—1500 µM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with His_6-ICL1 (final [His_6-ICL1] 30 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 90 seconds after initiation. The assay was performed in triplicate or in quadruplicate, as specified in (**Table 52**). Either the sample order within each replicate was randomised or the reactions within each replicate were performed in order of increasing DL-isocitrate concentration; the choice for each His_6-ICL1 is given in **Table 52** (for His_6-ICL1s, the sample order did not affect the data quality). For each reaction, the rate was extracted and converted to concentration units using $\varepsilon_{340}(NADH) = 6220\ M^{-1}\ cm^{-1}$.

Table 52: Assay variations between His_6-ICL1s.
The concentration of the enzyme stock, the number of replicates performed, and whether or not the sample order was randomised depended on which His_6-ICL1 was being assayed. No, samples were run in order of increasing [DL-isocitrate]; Yes, the order in which assay reactions were carried out was randomised within each replicate.

Species	[His_6-ICL1 stock]/µM	Number of replicates	Samples randomised?
M. tuberculosis CDC 1551	6	4	No
M. aromaticivorans	5	3	Yes
M. aurum	5	3	Yes
M. chubuense	5	3	Yes
M. fallax	6	4	No
M. hiberniae	6	3	No
M. intracellulare	5	3	Yes
M. marinum	2.5	3	No
M. salmoniphilum	6	4	No
M. smegmatis	6	4	No
M. terrae	5	3	No

8.7.3. ICL2-His_6FLAG, ICL2, and ICL2(T605)-His_6FLAG assays

8.7.3.1. Mtb CDC 1551 ICL2-His_6FLAG

Mtb CDC 1551 ICL2-His_6FLAG was purified as described above (**Section 8.5.8.2**) and diluted to 10 µM in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MgCl_2$ at pH 7.0. The diluted enzyme stock was used to initiate assay reactions. Three experiments were performed:

a) Varying the concentration of DL-isocitrate in the absence of Ac-CoA.

Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and varying concentrations of DL-isocitrate (11 concentrations in the range 0.1—4 mM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2-His_6FLAG (final [ICL2-His_6FLAG] 200 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 100 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

b) Varying the concentration of DL-isocitrate in the presence of 25 µM Ac-CoA.

Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), Ac-CoA (25 µM), and varying concentrations of DL-isocitrate (11 concentrations in the range 0.1—4 mM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2-His_6FLAG (final [ICL2-His_6FLAG] 200 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at

340 nm was recorded for 100 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

c) Varying the concentration of Ac-CoA in the presence of 5 mM DL-isocitrate.

Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), DL-isocitrate (5 mM), and varying concentrations of Ac-CoA (8 concentrations in the range 2—150 µM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2-His_6FLAG (final [ICL2-His_6FLAG] 100 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 100 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

8.7.3.2. Mtb CDC 1551 ICL2 and ICL2(T605)-His_6FLAG

Tag-cleaved ICL2: Mtb CDC 1551 ICL2-His_6FLAG was purified as described in **Section 8.5.8.2**. The His_6FLAG tag was cleaved by overnight incubation with TEV protease, as described in **Section 8.5.9**. Tag-cleaved ICL2 was diluted to 10 µM in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MgCl_2$ at pH 7.0. The diluted enzyme stock was used to initiate assay reactions.

ICL2(T605)-His_6FLAG: Mtb CDC 1551 ICL2(T605)-His_6FLAG was purified as described above (**Section 8.5.8.2**) and diluted to 10 µM in a buffer containing 50 mM HEPES,

150 mM NaCl and 5 mM $MgCl_2$ at pH 7.0. The diluted enzyme stock was used to initiate assay reactions.

Two experiments were performed with each enzyme:

a) Varying the concentration of DL-isocitrate in the absence of Ac-CoA.

Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and varying concentrations of DL-isocitrate (11 concentrations in the range 0.1—4 mM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2 or ICL2(T605)-His_6FLAG (final [enzyme] 200 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 100 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

b) Varying the concentration of DL-isocitrate in the presence of 25 µM Ac-CoA.

Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), Ac-CoA (25 µM), and varying concentrations of DL-isocitrate (11 concentrations in the range 0.1—4 mM). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2 or ICL2(T605)-His_6FLAG (final [enzyme] 200 nM). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 100 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction,

the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

8.7.3.3. ICL2-His_6FLAGs from mycobacterial species other than Mtb

Mycobacterial ICL2-His_6FLAGs were purified as described above (**Section 8.5.8.2**) and diluted to 10 µM in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MgCl_2$ at pH 7.0. The diluted enzyme stocks were used to initiate assay reactions. Each 1 mL reaction mixture contained HEPES (50 mM, pH 7.5), $MgCl_2$ (5 mM), NADH (200 µM), LDH (110 U/mL), and varying concentrations of DL-isocitrate (10—12 concentrations in the range specified in **Table 53**). Reaction mixtures were equilibrated at 37 °C for at least 5 minutes prior to initiation with ICL2-His_6FLAG (final [ICL2-His_6FLAG], **Table 53**). HEPES (50 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 120 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADH) = 6220 M^{-1} cm^{-1}.

Table 53: Enzyme and DL-isocitrate concentrations used in ICL2-His_6FLAG assays.

Species	Final [ICL2-His_6FLAG]/nM	Range of [DL-isocitrate]/mM
M. aromaticivorans	50	1—35
M. aurum	50	0.1—35
M. chubuense	50	0.1—35
M. hiberniae	50	1—25
M. intracellulare	50	0.1—35
M. marinum	50	0.1—35
M. paraffinicum	50	0.1—35
M. smegmatis	50	0.1—35
M. terrae	200	0.1—35

8.7.4. IDH2-His_6FLAG assays

Mycobacterial IDH2-His_6FLAGs were purified as described above (**Section 8.5.8.3**) and diluted to 1 µM in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MnCl_2$ at pH 7.0. The diluted enzyme stocks were used to initiate assay reactions.

Two sets of reactions were performed with each enzyme:

a) Varying the concentration of DL-isocitrate in the range 10—800 µM, keeping the concentration of $NADP^+$ at 2 mM.

Each 1 mL reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), $NADP^+$ (2 mM), and varying concentrations of DL-isocitrate (15 concentrations in the range 10—800 µM). Reaction mixtures were equilibrated at 37 °C for at least 7 minutes (Mtb, *M. aromaticivorans*, *M. aurum*, and *M. marinum*) or for at least 10 minutes (all other IDH2-His_6FLAGs) prior to initiation with IDH2-His_6FLAG (final [IDH2-His_6FLAG] 5 nM). HEPES (100 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 30 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}.

b) Varying the concentration of $NADP^+$ in the range 10 µM—2 mM, keeping the concentration of DL-isocitrate at 2 mM.

Each 1 mL reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), DL-isocitrate (2 mM), and varying concentrations of $NADP^+$ (12 concentrations in the range 10 µM—2 mM). Reaction mixtures were equilibrated at 37 °C for at least

7 minutes (Mtb, *M. aromaticivorans*, *M. aurum*, and *M. marinum*) or for at least 10 minutes (all other IDH2-His_6FLAGs) prior to initiation with IDH2-His_6FLAG (final [IDH2-His_6FLAG] 5 nM). HEPES (100 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 30 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}.

8.7.5. IDH1-His_6FLAG assays

Mycobacterial IDH1-His_6FLAGs were purified as described above (**Section 8.5.8.4**) and diluted to 1 µM in a buffer containing 50 mM HEPES, 150 mM NaCl and 5 mM $MnCl_2$ at pH 7.0. The diluted enzyme stocks were used to initiate assay reactions.

Two sets of reactions were performed with each enzyme:

a) Varying the concentration of DL-isocitrate in the range 10—800 µM, keeping the concentration of $NADP^+$ at 2 mM.

Each 1 mL reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), $NADP^+$ (2 mM), and varying concentrations of DL-isocitrate (12 concentrations in the range 10—800 µM). Reaction mixtures were equilibrated at 37 °C for at least 10 minutes prior to initiation with IDH1-His_6FLAG (final [IDH1-His_6FLAG] 5 nM). HEPES (100 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 30 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}.

b) Varying the concentration of $NADP^+$ in the range 10 µM—2 mM, keeping the concentration of DL-isocitrate at 2 mM.

Each 1 mL reaction mixture contained HEPES (100 mM, pH 7.5), $MnCl_2$ (5 mM), DL-isocitrate (2 mM), and varying concentrations of $NADP^+$ (12 concentrations in the range 10 µM—2 mM). Reaction mixtures were equilibrated at 37 °C for at least 10 minutes prior to initiation with IDH1-His_6FLAG (final [IDH1-His_6FLAG] 5 nM). HEPES (100 mM, pH 7.5, 1 mL) was used as the blank. The absorbance at 340 nm was recorded for 30 seconds after initiation. The assay was performed in triplicate. The sample order within each replicate was randomised. For each reaction, the rate was extracted and converted to concentration units using ε_{340}(NADPH) = 6220 M^{-1} cm^{-1}.

8.7.6. Determination of kinetic parameters

8.7.6.1. Determination of v_{max} and K_M

His_6-ICL1s, ICL2-His_6FLAGs, Mtb ICL2, Mtb ICL2(T605)-His_6FLAG, and IDH1-His_6FLAGs: After conversion to concentration units, initial rates were plotted in GraphPad Prism (versions 8.8.0—9.4.1) and fitted to the Michaelis-Menten equation to obtain values for v_{max} and K_M.

IDH2-His_6FLAGs: After conversion to concentration units, initial rates were plotted in GraphPad Prism (versions 8.8.0—9.4.1) and fitted to the Michaelis Menten and substrate inhibition models to obtain values for v_{max} and K_M. The best fitting model was determined automatically using the extra sum-of-squares F test, selecting the Michaelis-Menten model unless $P<0.05$. The preferred model for each IDH2-His_6FLAG is given in **Table 54**.

Table 54: Kinetic models used to determine v_{max} and K_M values for mycobacterial IDH2-His$_6$FLAGs.
The best fitting model was determined automatically using the extra sum-of-squares F test, selecting the Michaelis-Menten model unless P<0.05.

Kinetic parameters with respect to DL-isocitrate	
Michaelis-Menten	*M. aromaticivorans*, *M. aurum*, *M. fallax*, *M. marinum*, *M. salmoniphilum*, *M. smegmatis*, *M. terrae*, Mtb CDC 1551
Substrate inhibition	*M. chubuense*, *M. hiberniae*, *M. intracellulare*
Kinetic parameters with respect to $NADP^+$	
Michaelis-Menten	*M. chubuense*, *M. fallax*, *M. intracellulare*, *M. marinum*, Mtb CDC 1551
Substrate inhibition	*M. aromaticivorans*, *M. aurum*, *M. hiberniae*, *M. salmoniphilum*, *M. smegmatis*, *M. terrae*

Michaelis-Menten equation:

$$y = \frac{V_{\max} x}{K_M + x}$$

Substrate inhibition equation:

$$y = \frac{V_{\max} x}{K_M + x(1 + x/K_i)}$$

8.7.6.2. Determination of k_{cat}

Values for k_{cat} were calculated using the enzyme concentration and the values for v_{max} determined above (**Section 8.7.6.1**):

$$k_{\mathrm{cat}} = \frac{v_{\mathrm{max}}}{[\mathrm{Enzyme}]}$$

8.7.6.3. Determination of k_{cat}/K_M

Values for k_{cat}/K_M were calculated using the values for k_{cat} calculated in **Section 8.7.6.2** and the values for K_M calculated in **Section 8.7.6.1**.

Chapter 9. References

1. Gutierrez, M. C. *et al.* Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* **1**, 0055–0061 (2005).

2. Hershkovitz, I. *et al.* Tuberculosis origin: The Neolithic scenario. *Tuberculosis* **95**, S122–S126 (2015).

3. Dutour, O. Paleopathology of Human Infections: Old Bones, Antique Books, Ancient and Modern Molecules. *Microbiol. Spectr.* **4**, 4.4.12 (2016).

4. World Health Organisation. *Global Tuberculosis Report 2019*. (2019).

5. Lyon, S. M. & Rossman, M. D. Pulmonary Tuberculosis. *Microbiol. Spectr.* **5**, (2017).

6. Elder, N. C. Extrapulmonary tuberculosis. A review. *Arch. Fam. Med.* **1**, 91–98 (1992).

7. Golden, M. P. Extrapulmonary Tuberculosis: An Overview. *Am. Fam. Physician* **72**, 1761–1768 (2005).

8. Guo, Y.-K. *et al.* Addison's disease due to adrenal tuberculosis: Contrast-enhanced CT features and clinical duration correlation. *Eur. J. Radiol.* **62**, 126–131 (2007).

9. Paulson, T. A mortal foe. *Nature* **502**, S2–S3 (2013).

10. Fieweger, R. A., Wilburn, K. M. & VanderVen, B. C. Comparing the Metabolic Capabilities of Bacteria in the *Mycobacterium tuberculosis* Complex. *Microorganisms* **7**, 177 (2019).

11. Gagneux, S. Host–pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 850–859 (2012).

12. Roberts, C. A. Old World tuberculosis: Evidence from human remains with a review of current research and future prospects. *Tuberculosis* **95**, S117–S121 (2015).

13. Müller, R., Roberts, C. A. & Brown, T. A. Biomolecular identification of ancient *Mycobacterium tuberculosis* complex DNA in human remains from Britain and continental Europe. *Am. J. Phys. Anthropol.* **153**, 178–189 (2014).

14. Daniel, T. M. The history of tuberculosis. *Respir. Med.* **100**, 1862–1870 (2006).

15. Barberis, I., Bragazzi, N. L., Galluzzo, L. & Martini, M. The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus. *J. Prev. Med. Hyg.* **58**, E9–E12 (2017).

16. Daniel, V. S. & Daniel, T. M. Old Testament Biblical References to Tuberculosis. *Clin. Infect. Dis.* **29**, 1557–1558 (1999).

17. David, R. The art of healing in ancient Egypt: a scientific reappraisal. *Lancet* **372**, 1802–1803 (2008).

18. Mackowiak, P. A., Blos, V. T., Aguilar, M. & Buikstra, J. E. On the Origin of American Tuberculosis. *Clin. Infect. Dis.* **41**, 515–518 (2005).

19. Arriaza, B. T., Salo, W., Aufderheide, A. C. & Holcomb, T. A. Pre-Columbian tuberculosis in Northern Chile: Molecular and skeletal evidence. *Am. J. Phys. Anthropol.* **98**, 37–45 (1995).

20. Donoghue, H. D. Insights gained from ancient biomolecules into past and present tuberculosis—a personal perspective. *International Journal of Infectious Diseases* **56**, 176–180 (2017).

21. Okazaki, K. *et al.* A paleopathological approach to early human adaptation for wet-rice agriculture: The first case of Neolithic spinal tuberculosis at the Yangtze River Delta of China. *Int. J. Paleopathol.* **24**, 236–244 (2019).

22. Donoghue, H. D. Human tuberculosis – an ancient disease, as elucidated by ancient microbial biomolecules. *Microbes Infect.* **11**, 1156–1162 (2009).

23. Donoghue, H. D. Insights into ancient leprosy and tuberculosis using metagenomics. *Trends in Microbiology* **21**, 448–450 (2013).

24. Nicklisch, N. *et al.* Rib lesions in skeletons from early neolithic sites in Central Germany: On the trail of tuberculosis at the onset of agriculture. *Am. J. Phys. Anthropol.* **149**, 391–404 (2012).

25. Pósa, A. *et al.* Tuberculosis in late Neolithic-Early Copper Age human skeletal remains from Hungary. *Tuberculosis* **95**, S18–S22 (2015).

26. Sparacello, V. S., Roberts, C. A., Kerudin, A. & Müller, R. A 6500-year-old Middle Neolithic child from Pollera Cave (Liguria, Italy) with probable multifocal osteoarticular tuberculosis. *Int. J. Paleopathol.* **17**, 67–74 (2017).

27. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).

28. Stone, A. C., Wilbur, A. K., Buikstra, J. E. & Roberts, C. A. Tuberculosis and leprosy in perspective. *Am. J. Phys. Anthropol.* **140**, 66–94 (2009).

29. Zink, A. R., Grabner, W. & Nerlich, A. G. Molecular identification of human tuberculosis in recent and historic bone tissue samples: The role of molecular techniques for the study of historic tuberculosis. *Am. J. Phys. Anthropol.* **126**, 32–47 (2005).

30. Lalremruata, A. *et al.* Molecular Identification of Falciparum Malaria and Human Tuberculosis Co-Infections in Mummies from the Fayum Depression (Lower Egypt). *PLoS One* **8**, e60307 (2013).

31. Laios, K., Androutsos, G. & Moschos, M. M. Aretaeus of Cappadocia and Pulmonary Tuberculosis. *Balkan Med. J.* **34**, 480–480 (2017).

32. Daniel, T. M. & Iversen, P. A. Hippocrates and tuberculosis. *Int. J. Tuberc. Lung Dis.* **19**, 373–374 (2015).

33. Karamanou, M. & Androutsos, G. The masterful description of pulmonary tuberculosis by Soranus of Ephesus (c. 98-138 A.D.). *American Journal of Respiratory and Critical Care Medicine* **186**, 571 (2012).

34. Cambau, E. & Drancourt, M. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clinical Microbiology and Infection* **20**, 196–201 (2014).

35. Riva, M. A. From milk to rifampicin and back again: history of failures and successes in the treatment for tuberculosis. *J. Antibiot. (Tokyo).* **67**, 661–665 (2014).

36. Grzybowski, S. & Allen, E. A. History and importance of scrofula. *Lancet* **346**, 1472–1474 (1995).

37. Bray, F. N., Alsaidan, M., Simmons, B. J., Falto-Aizpurua, L. A. & Nouri, K. Scrofula and the divine right of royalty: The king's touch. *JAMA Dermatology* **151**, 702 (2015).

38. Murray, J. F., Rieder, H. L. & Finley-Croswhite, A. The King's Evil and the Royal Touch: the medical history of scrofula. *Int. J. Tuberc. Lung Dis.* **20**, 713–716 (2016).

39. Abbott, E. C. Composers and tuberculosis: the effects on creativity. *Can. Med. Assoc. J.* **126**, 534–544 (1982).

40. Morens, D. M. At the deathbed of consumptive art. *Emerging Infectious Diseases* **8**, 1353–1358 (2002).

41. Glaziou, P., Floyd, K. & Raviglione, M. Trends in tuberculosis in the UK. *Thorax* **73**, 702–703 (2018).

42. Lessel, E. F. Status of the Generic Name *Mycobacterium* Lehmann and Neumann. *Int. Bull. Bacteriol. Nomencl. Taxon.* **10**, 47–54 (1960).

43. Martini, M., Gazzaniga, V., Behzadifar, M., Bragazzi, N. L. & Barberis, I. The history of tuberculosis: the social role of sanatoria for the treatment of tuberculosis in Italy between the end of the 19th century and the middle of the 20th. *J. Prev. Med. Hyg.* **59**, E323–E327 (2018).

44. Yang, H., Kruh-Garcia, N. A. & Dobos, K. M. Purified protein derivatives of tuberculin — past, present, and future. *FEMS Immunol. Med. Microbiol.* **66**, 273–280 (2012).

45. Monaghan, M. L., Doherty, M. L., Collins, J. D., Kazda, J. F. & Quinn, P. J. The tuberculin test. *Vet. Microbiol.* **40**, 111–124 (1994).

46. Knopf, S. A. Some thoughts on overcrowding and tuberculosis. *JAMA J. Am. Med. Assoc.* **XXXV**, 991 (1900).

47. Norway, N. E. Cervello's treatment of phthisis. *Br. Med. J.* **2**, 662 (1900).

48. Newsholme, A. *The Prevention of Tuberculosis (Routledge Revivals)*. (Routledge, 2015).

49. Abdallah, A. M. & Behr, M. A. Evolution and Strain Variation in BCG. in *Advances in Experimental Medicine and Biology* **1019**, 155–169 (Springer New York LLC, 2017).

50. Hawgood, B. J. Albert Calmette (1863–1933) and Camille Guérin (1872–1961): the C and G of BCG vaccine. *J. Med. Biogr.* **15**, 139–146 (2007).

51. Lugosi, L. Theoretical and methodological aspects of BCG vaccine from the discovery of Calmette and Guérin to molecular biology. A review. *Tubercle and Lung Disease* **73**, 252–261 (1992).

52. Bentley, R. Mycophenolic Acid: A One Hundred Year Odyssey from Antibiotic to Immunosuppressant. *Chem. Rev.* **100**, 3801–3826 (2000).

53. Mohr, K. I. History of antibiotics research. *Curr. Top. Microbiol. Immunol.* **398**, 237–272 (2016).

54. Bosch, F. & Rosich, L. The Contributions of Paul Ehrlich to Pharmacology: A Tribute on the Occasion of the Centenary of His Nobel Prize. *Pharmacology* **82**, 171–179 (2008).

55. Zaffiri, L., Gardner, J. & Toledo-Pereyra, L. H. History of Antibiotics. From Salvarsan to Cephalosporins. *J. Investig. Surg.* **25**, 67–77 (2012).

56. Aminov, R. I. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front. Microbiol.* **1**, (2010).

57. Houbraken, J., Frisvad, J. C. & Samson, R. A. Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*. *IMA Fungus* **2**, 87–95 (2011).

58. Alharbi, S. A. *et al.* What if Fleming had not discovered penicillin? *Saudi J. Biol. Sci.* **21**, 289–293 (2014).

59. Moberg, C. L. Penicillin's Forgotten Man: Norman Heatley. *Science* **253**, 734–735 (1991).

60. Kerantzas, C. A. & Jacobs, W. R. Origins of Combination Therapy for Tuberculosis: Lessons for Future Antimicrobial Development and Application. *mBio* **8**, (2017).

61. Waksman, S. A. & Woodruff, H. B. Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. *Exp. Biol. Med.* **45**, 609–614 (1940).

62. Schatz, A. & Waksman, S. A. Effect of Streptomycin and Other Antibiotic Substances upon *Mycobacterium tuberculosis* and Related Organisms.,. *Exp. Biol. Med.* **57**, 244–248 (1944).

63. Markoe, R. C. L. The use of *para*-amino salicylic acid in the treatment of tuberculosis. *J. Natl. Med. Assoc.* **42**, 364–6 (1950).

64. *Para*-aminosalicylic acid. *Tuberculosis* **88**, 137–138 (2008).

65. Isoniazid. *Tuberculosis* **88**, 112–116 (2008).

66. Gomez, J. E. & McKinney, J. D. *M. tuberculosis* persistence, latency, and drug tolerance. *Tuberculosis* **84**, 29–44 (2004).

67. Murray, J. F., Schraufnagel, D. E. & Hopewell, P. C. Treatment of tuberculosis: A historical perspective. *Annals of the American Thoracic Society* **12**, 1749–1759 (2015).

68. Street, E. W. The Chemotherapy of Tuberculosis—A Personal Overview. *Scott. Med. J.* **22**, 279–285 (1977).

69. Drugs for tuberculosis. *Br. Med. J.* **3**, 664–7 (1968).

70. Iseman, M. D. Tuberculosis therapy: past, present and future. *Eur. Respir. J.* **20**, Suppl. 36, 87s-94s (2002).

71. Hobby, G. L. & Lenert, T. F. The antimycobacterial activity of rifampin. *Am. Rev. Respir. Dis.* **97**, 713–714 (1968).

72. Steele, M. A. & Des Prez, R. M. The Role of Pyrazinamide in Tuberculosis Chemotherapy. *Chest* **94**, 845–850 (1988).

73. Lawn, S. D. & Zumla, A. I. Tuberculosis. *Lancet* **378**, 57–72 (2011).

74. Furin, J., Cox, H. & Pai, M. Tuberculosis. *Lancet* **393**, 1642–1656 (2019).

75. Fogel, N. Tuberculosis: A disease without boundaries. *Tuberculosis* **95**, 527–531 (2015).

76. World Health Organisation. *The work of WHO, 1975: annual report of the Director-General to the World Health Assembly and to the United Nations*. (1976).

77. World Health Organisation. *Who Expert Committee on Tuberculosis: ninth report*. (1973).

78. World Health Organisation. *Tuberculosis control in the world - situation analysis (WHA36.30)*. (1983).

79. World Health Organisation. *TB: A global emergency, WHO report on the TB epidemic, 1994 (WHO/TB/94.177)*. (1994).

80. World Health Organisation. *Global medium-term programme: programme 13.8, tuberculosis (TUB/MTP/87.1)*. (1987).

81. World Health Organisation. *Tuberculosis programme: report by the Director-General (EB91/16)*. (1992).

82. World Health Organisation. *Tuberculosis control programme: progress and evaluation report (EB87/4)*. (1990).

83. World Health Organisation. *Global Tuberculosis Report 2020*. (2020).

84. World Health Organisation. *Global Tuberculosis Report 2021*. (2021).

85. World Health Organisation. *Global Tuberculosis Report 2022*. (2022).

86. United Nations Department of Economic and Social Affairs Population Division. *World Population Prospects 2019: Highlights (ST/ESA/SER.A/423)*. (2019).

87. United Nations Department of Economic and Social Affairs Population Division. *2019 Revision of World Population Prospects*. (2019).

88. Hertz, D. & Schneider, B. Sex differences in tuberculosis. *Semin. Immunopathol.* **41**, 225–237 (2019).

89. Cardona, P.-J., Català, M. & Prats, C. Origin of tuberculosis in the Paleolithic predicts unprecedented population growth and female resistance. *Sci. Rep.* **10**, 42 (2020).

90. World Health Organisation. *WHO consolidated guidelines on tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update*. (2022).

91. Conradie, F. *et al.* Bedaquiline–Pretomanid–Linezolid Regimens for Drug-Resistant Tuberculosis. *N. Engl. J. Med.* **387**, 810–823 (2022).

92. Conradie, F. *et al.* Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N. Engl. J. Med.* **382**, 893–902 (2020).

93. Berry, C. *et al.* TB-PRACTECAL: study protocol for a randomised, controlled, open-label, phase II–III trial to evaluate the safety and efficacy of regimens containing bedaquiline and pretomanid for the treatment of adult patients with pulmonary multidrug-resistant tuberculosis. *Trials* **23**, 484 (2022).

94. Zumla, A. I. *et al.* New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *Lancet. Infect. Dis.* **14**, 327–40 (2014).

95. Alsayed, S. S. R. & Gunosewoyo, H. Tuberculosis: Pathogenesis, Current Treatment Regimens and New Drug Targets. *Int. J. Mol. Sci.* **24**, 5202 (2023).

96. Koehler, N. *et al.* Pretomanid-resistant tuberculosis. *J. Infect.* **0**, 1–5 (2023).

97. The Review on Antimicrobial Resistance. *Tackling Drug-Resistant Infections Globally: final report and recommendations*. (2016).

98. World Health Organisation. *The end TB strategy (WHO/HTM/TB/2015.19)*. (2015).

99. World Health Organisation. *Global strategy and targets for tuberculosis prevention, care and control after 2015: Report by the Secretariat (EB134/12)*. (2014).

100. Kyu, H. H. *et al.* Global, regional, and national burden of tuberculosis, 1990–2016: results from the Global Burden of Diseases, Injuries, and Risk Factors 2016 Study. *Lancet Infect. Dis.* **18**, 1329–1349 (2018).

101. Parrish, N. M., Dick, J. D. & Bishai, W. R. Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends Microbiol.* **6**, 107–112 (1998).

102. Russell, D. G. *et al. Mycobacterium tuberculosis* Wears What It Eats. *Cell Host Microbe* **8**, 68–76 (2010).

103. Matteelli, A. *et al.* Tuberculosis elimination and the challenge of latent tuberculosis. *Presse Med.* **46**, e13–e21 (2017).

104. Salgame, P., Geadas, C., Collins, L., Jones-López, E. & Ellner, J. J. Latent tuberculosis infection – Revisiting and revising concepts. *Tuberculosis* **95**, 373–384 (2015).

105. Sparacello, V. S., Roberts, C. A., Canci, A., Moggi-Cecchi, J. & Marchi, D. Insights on the paleoepidemiology of ancient tuberculosis from the structural analysis of postcranial remains from the Ligurian Neolithic (northwestern Italy). *Int. J. Paleopathol.* **15**, 50–64 (2016).

106. Chee, C. B. E., Reves, R., Zhang, Y. & Belknap, R. Latent tuberculosis infection: Opportunities and challenges. *Respirology* **23**, 893–900 (2018).

107. Donoghue, H. D. Insights gained from palaeomicrobiology into ancient and modern tuberculosis. *Clin. Microbiol. Infect.* **17**, 821–829 (2011).

108. Wirth, T. *et al.* Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. *PLoS Pathog.* **4**, e1000160 (2008).

109. Saelens, J. W., Viswanathan, G. & Tobin, D. M. Mycobacterial Evolution Intersects With Host Tolerance. *Front. Immunol.* **10**, 528 (2019).

110. Nouioui, I. *et al.* Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Front. Microbiol.* **9**, 2007 (2018).

111. Verma, M. *et al.* Phylogenetic analyses of phylum Actinobacteria based on whole genome sequences. *Res. Microbiol.* **164**, 718–728 (2013).

112. Lewin, G. R. *et al.* Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu. Rev. Microbiol.* **70**, 235–254 (2016).

113. Battistuzzi, F. U., Feijao, A. & Hedges, S. B. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4**, 44 (2004).

114. Salam, N., Jiao, J.-Y., Zhang, X.-T. & Li, W.-J. Update on the classification of higher ranks in the phylum Actinobacteria. *Int. J. Syst. Evol. Microbiol.* **70**, 1331–1355 (2020).

115. Jurado, V. *et al. Hoyosella altamirensis* gen. nov., sp. nov., a new member of the order Actinomycetales isolated from a cave biofilm. *Int. J. Syst. Evol. Microbiol.* **59**, 3105–3110 (2009).

116. Kazda, J., Pavlik, I., Falkinham III, J. O. & Hruska, K. *The Ecology of Mycobacteria: Impact on Animal's and Human's Health*. (Springer Netherlands, 2009). doi:10.1007/978-1-4020-9413-2

117. Hall, L. & Roberts, G. D. Non-molecular identification of nontuberculous mycobacteria in the clinical microbiology laboratory: What's the real deal? *Clin. Microbiol. Newsl.* **28**, 73–80 (2006).

118. Turenne, C. Y. Nontuberculous mycobacteria: Insights on taxonomy and evolution. *Infect. Genet. Evol.* **72**, 159–168 (2019).

119. Tortoli, E. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin. Microbiol. Rev.* **16**, 319–54 (2003).

120. Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C. & Göker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607–5612 (2020).

121. Pavlik, I., Ulmann, V., Hubelova, D. & Weston, R. T. Nontuberculous Mycobacteria as Sapronoses: A Review. *Microorganisms* **10**, 1345 (2022).

122. Magee, J. G. & Ward, A. C. *Mycobacterium*. in *Bergey's Manual of Systematics of Archaea and Bacteria* 1–84 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118960608.gbm00029

123. Runyon, E. H. Anonymous Mycobacteria in Pulmonary Disease. *Med. Clin. North Am.* **43**, 273–290 (1959).

124. Shinnick, T. M. & Good, R. C. Mycobacterial taxonomy. *Eur. J. Clin. Microbiol. Infect. Dis.* **13**, 884–901 (1994).

125. van Ingen, J. Mycobacteria. in *Infectious Diseases* 1645-1659.e2 (Elsevier, 2017). doi:10.1016/B978-0-7020-6285-8.00185-4

126. García-Agudo, L. & García-Martos, P. Clinical significance and antimicrobial susceptibility of rapidly growing mycobacteria. in *Science against Microbial Pathogens: Communicating Current Research and Technological Advances* 363–377 (2011).

127. Koh, W. J., Kwon, O. J. & Lee, K. S. Nontuberculous Mycobacterial Pulmonary Diseases in Immunocompetent Patients. *Korean Journal of Radiology* **3**, 145–157 (2002).

128. Butler, W. R. & Guthertz, L. S. Mycolic Acid Analysis by High-Performance Liquid Chromatography for Identification of *Mycobacterium* Species. *Clin. Microbiol. Rev.* **14**, 704–726 (2001).

129. Kieser, K. J. & Rubin, E. J. How sisters grow apart: mycobacterial growth and division. *Nat. Rev. Microbiol.* **12**, 550–562 (2014).

130. Tran, T., Bonham, A. J., Chan, E. D. & Honda, J. R. A paucity of knowledge regarding nontuberculous mycobacterial lipids compared to the tubercle bacillus. *Tuberculosis* **115**, 96–107 (2019).

131. Nguyen, L. & Thompson, C. J. Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm. *Trends Microbiol.* **14**, 304–312 (2006).

132. Jarher, V. & Nlkaldo, H. Mycobacterial cell wall: Structure and role in natural resistance to antibiotics. *FEMS Microbiol. Lett.* **123**, 11–18 (1994).

133. Primm, T. P., Lucero, C. A. & Falkinham, J. O. Health Impacts of Environmental Mycobacteria. *Clin. Microbiol. Rev.* **17**, 98–106 (2004).

134. Bayot, M. L., Mirza, T. M. & Sharma, S. *Acid Fast Bacteria*. *StatPearls* (StatPearls Publishing, 2022).

135. Marrakchi, H., Lanéelle, M. A. & Daffé, M. Mycolic Acids: Structures, Biosynthesis, and Beyond. *Chem. Biol.* **21**, 67–85 (2014).

136. Tortoli, E. Microbiological features and clinical relevance of new species of the genus *Mycobacterium*. *Clin. Microbiol. Rev.* **27**, 727–52 (2014).

137. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S Ribosomal DNA Amplification for Phylogenetic Study. *J. Bacteriol.* **173**, 697–703 (1991).

138. Fox, G. E., Pechman, K. R. & Woese, C. R. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. *Int. J. Syst. Bacteriol.* **27**, 44–57 (1977).

139. Rosselló-Mora, R. & Amann, R. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25**, 39–67 (2001).

140. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol. 2014 129* **12**, 635–645 (2014).

141. Yarza, P. *et al.* The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* **31**, 241–250 (2008).

142. Rogall, T., Wolters, J., Flohr, T. & Bottger, E. C. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *Int. J. Syst. Bacteriol.* **40**, 323–330 (1990).

143. Ludwig, W. *et al.* Release LTP_12_2020, featuring a new ARB alignment and improved 16S rRNA tree for prokaryotic type strains. *Syst. Appl. Microbiol.* **44**, 126218 (2021).

144. Coimbra, N. D. R., Goes-Neto, A., Azevedo, V. & Ouangraoua, A. Reconstructing the Phylogeny of Corynebacteriales while Accounting for Horizontal Gene Transfer. *Genome Biol. Evol.* **12**, 381–395 (2020).

145. Tortoli, E. Phylogeny of the genus *Mycobacterium*: Many doubts, few certainties. *Infect. Genet. Evol.* **12**, 827–831 (2012).

146. Beye, M., Fahsi, N., Raoult, D. & Fournier, P.-E. Careful use of 16S rRNA gene sequence similarity values for the identification of *Mycobacterium* species. *New Microbes New Infect.* **22**, 24–29 (2018).

147. Mignard, S. & Flandrois, J. P. A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. *Int. J. Syst. Evol. Microbiol.* **58**, 1432–1441 (2008).

148. Devulder, G., de Montclos, M. P. & Flandrois, J. P. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int. J. Syst. Evol. Microbiol.* **55**, 293–302 (2005).

149. Prasanna, A. N. & Mehra, S. Comprehensive Phylogenetic Analysis of Mycobacteria. *IFAC Proc. Vol.* **46**, 101–106 (2013).

150. Gupta, R. S., Lo, B. & Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Front. Microbiol.* **9**, 67 (2018).

151. Fedrizzi, T. *et al.* Genomic characterization of Nontuberculous Mycobacteria. *Sci. Rep.* **7**, 1–14 (2017).

152. Tortoli, E. *et al.* The new phylogeny of the genus *Mycobacterium*: The old and the news. *Infect. Genet. Evol.* **56**, 19–25 (2017).

153. Bachmann, N. L. *et al.* Key Transitions in the Evolution of Rapid and Slow Growing Mycobacteria Identified by Comparative Genomics. *Front. Microbiol.* **10**, 1–12 (2020).

154. Tortoli, E. *et al.* Same meat, different gravy: ignore the new names of mycobacteria. *Eur. Respir. J.* **54**, 1900795 (2019).

155. Tsukamura, M. Identification of mycobacteria. *Tubercle* **48**, 311–338 (1967).

156. Meehan, C. J., Barco, R. A., Loh, Y.-H. E., Cogneau, S. & Rigouts, L. Reconstituting the genus *Mycobacterium*. *Int. J. Syst. Evol. Microbiol.* **71**, 004922 (2021).

157. Weerdenburg, E. M. *et al.* Genome-Wide Transposon Mutagenesis Indicates that *Mycobacterium marinum* Customizes Its Virulence Mechanisms for Survival and Replication in Different Hosts. *Infect. Immun.* **83**, 1778–1788 (2015).

158. Allen, A. R. One bacillus to rule them all? – Investigating broad range host adaptation in *Mycobacterium bovis*. *Infect. Genet. Evol.* **53**, 68–76 (2017).

159. Pereira, A. C., Ramos, B., Reis, A. C. & Cunha, M. V. Non-Tuberculous Mycobacteria: Molecular and Physiological Bases of Virulence and Adaptation to Ecological Niches. *Microorganisms* **8**, 1380 (2020).

160. Bottai, D., Stinear, T. P., Supply, P. & Brosch, R. Mycobacterial Pathogenomics and Evolution. *Microbiol. Spectr.* **2**, (2014).

161. Sapriel, G. & Brosch, R. Shared Pathogenomic Patterns Characterize a New Phylotype, Revealing Transition toward Host-Adaptation Long before Speciation of *Mycobacterium tuberculosis*. *Genome Biol. Evol.* **11**, 2420–2438 (2019).

162. Pai, H. H., Chen, W. C. & Peng, C. F. Isolation of non-tuberculous mycobacteria from hospital cockroaches (*Periplaneta americana*). *J. Hosp. Infect.* **53**, 224–228 (2003).

163. Hruska, K. & Kaevska, M. Mycobacteria in water, soil, plants and air: a review. *Vet. Med. (Praha).* **57**, 623–679 (2012).

164. Kawamura, Y. *et al.* Bacterial population in Russian space station 'Mir'. *Microbiol. Immunol.* **45**, 819–828 (2001).

165. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).

166. Walsh, C. M., Gebert, M. J., Delgado-Baquerizo, M., Maestre, F. T. & Fierer, N. A Global Survey of Mycobacterial Diversity in Soil. *Appl. Environ. Microbiol.* **85**, (2019).

167. Falkinham, J. O. & III. Ecology of Nontuberculous Mycobacteria. *Microorganisms* **9**, (2021).

168. King, H. C. *et al.* Environmental reservoirs of pathogenic mycobacteria across the Ethiopian biogeographical landscape. *PLoS One* **12**, e0173811 (2017).

169. Thorel, M.-F., Falkinham, J. O. & Moreau, R. G. Environmental mycobacteria from alpine and subalpine habitats. *FEMS Microbiol. Ecol.* **49**, 343–347 (2004).

170. Kazda, J. *et al. Mycobacterium hiberniae* sp. nov. *Int. J. Syst. Bacteriol.* **43**, 352–357 (1993).

171. Galazka, A. *et al.* Genetic and functional diversity of bacterial microbiome in soils with long term impacts of petroleum hydrocarbons. *Front. Microbiol.* **9**, 1923 (2018).

172. Asmar, S., Sassi, M., Phelippeau, M. & Drancourt, M. Inverse correlation between salt tolerance and host-adaptation in mycobacteria. *BMC Res. Notes* **9**, 249 (2016).

173. Honda, J. R., Virdi, R. & Chan, E. D. Global Environmental Nontuberculous Mycobacteria and Their Contemporaneous Man-Made and Natural Niches. *Front. Microbiol.* **9**, (2018).

174. Gebert, M. J. *et al.* Ecological Analyses of Mycobacteria in Showerhead Biofilms and Their Relevance to Human Health. *mBio* **9**, (2018).

175. Howard, S. T. & Byrd, T. F. The rapidly growing mycobacteria: saprophytes and parasites. *Microbes Infect.* **2**, 1845–1853 (2000).

176. Webster, T. M. *et al.* Structure and Functional Attributes of Bacterial Communities in Premise Plumbing across the United States. *Environ. Sci. Technol.* **55**, 14105–14114 (2021).

177. Ma, X. *et al.* Revealing the changes of bacterial community from water source to consumers tap: A full-scale investigation in eastern city of China. *J. Environ. Sci.* **87**, 331–340 (2020).

178. Van Ingen, J., Blaak, H., De Beer, J., De Roda Husman, A. M. & Van Soolingen, D. Rapidly growing nontuberculous mycobacteria cultured from home tap and shower water. *Appl. Environ. Microbiol.* **76**, 6017–6019 (2010).

179. Zhang, X., Xia, S., Ye, Y. & Wang, H. Opportunistic pathogens exhibit distinct growth dynamics in rainwater and tap water storage systems. *Water Res.* **204**, 117581 (2021).

180. Makovcova, J., Slany, M., Babak, V., Slana, I. & Kralik, P. The water environment as a source of potentially pathogenic mycobacteria. *J. Water Health* **12**, 254–263 (2014).

181. Halstrom, S., Price, P. & Thomson, R. Review: Environmental mycobacteria as a cause of human infection. *Int. J. Mycobacteriology* **4**, 81–91 (2015).

182. Forbes, B. A. *et al.* Practical Guidance for Clinical Microbiology Laboratories: Mycobacteria. *Clin. Microbiol. Rev.* **31**, e00038-17 (2018).

183. Falkinham, III, J. O. Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *J. Appl. Microbiol.* **107**, 356–367 (2009).

184. Guo, F. *et al.* Mycobacterial species and their contribution to cholesterol degradation in wastewater treatment plants. *Sci. Rep.* **9**, 836 (2019).

185. Lamrabet, O., Merhej, V., Pontarotti, P., Raoult, D. & Drancourt, M. The Genealogic Tree of Mycobacteria Reveals a Long-Standing Sympatric Life into Free-Living Protozoa. *PLoS One* **7**, e34754 (2012).

186. Drancourt, M. Looking in amoebae as a source of mycobacteria. *Microb. Pathog.* **77**, 119–124 (2014).

187. Salah, I. B., Ghigo, E. & Drancourt, M. Free-living amoebae, a training field for macrophage resistance of mycobacteria. *Clin. Microbiol. Infect.* **15**, 894–905 (2009).

188. Greub, G. & Raoult, D. Microorganisms Resistant to Free-Living Amoebae. *Clinical Microbiology Reviews* **17**, 413–433 (2004).

189. Delafont, V. *et al.* First Evidence of Amoebae–Mycobacteria Association in Drinking Water Network. *Environ. Sci. Technol.* **48**, 11872–11882 (2014).

190. Chavarro-Portillo, B., Soto, C. Y. & Guerrero, M. I. *Mycobacterium leprae*'s evolution and environmental adaptation. *Acta Trop.* **197**, 105041 (2019).

191. Cosma, C. L., Sherman, D. R. & Ramakrishnan, L. The Secret Lives of the Pathogenic Mycobacteria. *Annu. Rev. Microbiol.* **57**, 641–676 (2003).

192. Rindi, L. & Garzelli, C. Genetic diversity and phylogeny of *Mycobacterium avium*. *Infection, Genetics and Evolution* **21**, 375–383 (2014).

193. Barrios-Payán, J. *et al.* Extrapulmonary locations of *Mycobacterium tuberculosis* DNA during latent infection. *J. Infect. Dis.* **206**, 1194–1205 (2012).

194. Beigier-Bompadre, M. *et al. Mycobacterium tuberculosis* infection modulates adipose tissue biology. *PLOS Pathog.* **13**, e1006676 (2017).

195. Guirado, E., Schlesinger, L. S. & Kaplan, G. Macrophages in Tuberculosis: Friend or Foe. *Semin. Immunopathol.* **35**, 563 (2013).

196. Cook, G. M. *et al.* Physiology of Mycobacteria. *Adv. Microb. Physiol.* **55**, 81–319 (2009).

197. Lerner, T. R. *et al.* Lymphatic endothelial cells are a replicative niche for *Mycobacterium tuberculosis*. *J. Clin. Invest.* **126**, 1093–1108 (2016).

198. Corner, L. A. L., Murphy, D. & Gormley, E. *Mycobacterium bovis* Infection in the Eurasian Badger (*Meles meles*): the Disease, Pathogenesis, Epidemiology and Control. *J. Comp. Pathol.* **144**, 1–24 (2011).

199. van Tonder, A. J. *et al.* Inferring *Mycobacterium bovis* transmission between cattle and badgers using isolates from the Randomised Badger Culling Trial. *PLOS Pathog.* **17**, e1010075 (2021).

200. Zingue, D., Bouam, A., Tian, R. B. D. & Drancourt, M. Buruli Ulcer, a Prototype for Ecosystem-Related Infection, Caused by *Mycobacterium ulcerans*. *Clin. Microbiol. Rev.* **31**, (2018).

201. Johansen, M. D., Herrmann, J.-L. & Kremer, L. Non-tuberculous mycobacteria and the rise of *Mycobacterium abscessus*. *Nat. Rev. Microbiol.* **18**, 392–407 (2020).

202. Rastogi, N., Legrand, E. & Sola, C. The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev. Sci. Tech.* **20**, 21–54 (2001).

203. Bouam, A., Armstrong, N., Levasseur, A. & Drancourt, M. *Mycobacterium terramassiliense*, *Mycobacterium rhizamassiliense* and *Mycobacterium numidiamassiliense* sp. nov., three new *Mycobacterium simiae* complex species cultured from plant roots. *Sci. Rep.* **8**, 9309 (2018).

204. Robinson, R. T. & Huppler, A. R. The Goldilocks model of immune symbiosis with Mycobacteria and Candida colonizers. *Cytokine* **97**, 49–65 (2017).

205. Amoroso, M., Böttcher, A., Lowry, C. A., Langgartner, D. & Reber, S. O. Subcutaneous *Mycobacterium vaccae* promotes resilience in a mouse model of chronic psychosocial stress when administered prior to or during psychosocial stress. *Brain. Behav. Immun.* **87**, 309–317 (2020).

206. Callahan, G. N. Eating dirt. *Emerg. Infect. Dis.* **9**, 1016–21 (2003).

207. Selway, C. A. *et al.* Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure. *Environ. Int.* **145**, 106084 (2020).

208. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).

209. Fellag, M., Loukil, A. & Drancourt, M. The puzzle of the evolutionary natural history of tuberculosis. *New Microbes New Infect.* **41**, 100712 (2021).

210. Cambier, C. J., Falkow, S. & Ramakrishnan, L. Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. *Cell* **159**, 1497–1509 (2014).

211. Qvist, T. *et al.* Chronic pulmonary disease with *Mycobacterium abscessus* complex is a biofilm infection. *Eur. Respir. J.* **46**, 1823–1826 (2015).

212. Middleton, A. M. *et al.* Investigation of mycobacterial colonisation and invasion of the respiratory mucosa. *Thorax* **58**, 246–251 (2003).

213. Silva, M. T. Classical Labeling of Bacterial Pathogens According to Their Lifestyle in the Host: Inconsistencies and Alternatives. *Front. Microbiol.* **3**, 71 (2012).

214. Akinola, R. O., Mazandu, G. K. & Mulder, N. J. A Quantitative Approach to Analyzing Genome Reductive Evolution Using Protein–Protein Interaction Networks: A Case Study of *Mycobacterium leprae*. *Front. Genet.* **7**, 39 (2016).

215. Ventura, M. *et al.* Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiol. Mol. Biol. Rev.* **71**, 495–548 (2007).

216. Moran, N. A. Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).

217. Gómez-Valero, L., Rocha, E. P. C., Latorre, A. & Silva, F. J. Reconstructing the ancestor of *Mycobacterium leprae*: The dynamics of gene loss and genome reduction. *Genome Res.* **17**, 1178–1185 (2007).

218. Batut, B., Knibbe, C., Marais, G. & Daubin, V. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* **12**, 841–850 (2014).

219. Rahman, S. A. *et al.* Comparative Analyses of Nonpathogenic, Opportunistic, and Totally Pathogenic Mycobacteria Reveal Genomic and Biochemical Variabilities and Highlight the Survival Attributes of *Mycobacterium tuberculosis*. *mBio* **5**, e02020 (2014).

220. Malhotra, S., Vedithi, S. C. & Blundell, T. L. Decoding the similarities and differences among mycobacterial species. *PLoS Negl. Trop. Dis.* **11**, e0005883 (2017).

221. Bäumler, A. J., Tsolis, R. M., Ficht, T. A. & Adams, L. G. Evolution of host adaptation in *Salmonella enterica*. *Infection and Immunity* **66**, 4579–4587 (1998).

222. Baumler, A. & Fang, F. C. Host Specificity of Bacterial Pathogens. *Cold Spring Harb. Perspect. Med.* **3**, a010041 (2013).

223. Reinink, P. *et al.* Discovery of *Salmonella* trehalose phospholipids reveals functional convergence with mycobacteria. *J. Exp. Med.* **216**, 757–771 (2019).

224. Lovell-Read, S. R. & de Carvalho, L. P. S. "*Salmonella*, meet mycobacteria." *J. Exp. Med.* **216**, 721–722 (2019).

225. Muthusamy, S. *et al.* Comparative proteomics reveals signature metabolisms of exponentially growing and stationary phase marine bacteria. *Environ. Microbiol.* **19**, 2301–2319 (2017).

226. Teeling, H. *et al.* Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* **336**, 608–611 (2012).

227. Eldholm, V. & Balloux, F. Antimicrobial Resistance in *Mycobacterium tuberculosis*: The Odd One Out. *Trends Microbiol.* **24**, 637–648 (2016).

228. Segerman, B. The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* **2**, 116 (2012).

229. van Tonder, A. J. *et al.* Defining the Estimated Core Genome of Bacterial Populations Using a Bayesian Decision Model. *PLoS Comput. Biol.* **10**, e1003788 (2014).

230. Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).

231. Yang, T. *et al.* Pan-genomic study of *Mycobacterium tuberculosis* reflecting the primary/secondary genes, generality/individuality, and the interconversion through copy number variations. *Front. Microbiol.* **9**, 1886 (2018).

232. Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **15**, 141–161 (2015).

233. Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**, 123–140 (2004).

234. Koli, P., Sudan, S., Fitzgerald, D., Adhya, S. & Kar, S. Conversion of commensal *Escherichia coli* K-12 to an invasive form via expression of a mutant histone-like protein. *mBio* **2**, 12 (2011).

235. Behr, M. A. Comparative Genomics of Mycobacteria: Some Answers, Yet More New Questions. *Cold Spring Harb. Perspect. Med.* **5**, a021204 (2015).

236. Joao, I., Cristovao, P., Antunes, L., Nunes, B. & Jordao, L. Identification of nontuberculous mycobacteria by partial gene sequencing and public databases. *Int. J. Mycobacteriology* **3**, 144–151 (2014).

237. Altaf, M., Miller, C. H., Bellows, D. S. & O'Toole, R. Evaluation of the *Mycobacterium smegmatis* and BCG models for the discovery of *Mycobacterium tuberculosis* inhibitors. *Tuberculosis* **90**, 333–337 (2010).

238. Namouchi, A., Cimino, M., Favre-Rochex, S., Charles, P. & Gicquel, B. Phenotypic and genomic comparison of *Mycobacterium aurum* and surrogate model species to *Mycobacterium tuberculosis*: implications for drug discovery. *BMC Genomics* **18**, 530 (2017).

239. T, J. A. S., J, R., Rajan, A. & Shankar, V. Features of the biochemistry of *Mycobacterium smegmatis*, as a possible model for *Mycobacterium tuberculosis*. *J. Infect. Public Health* **13**, 1255–1264 (2020).

240. Shiloh, M. U. & DiGiuseppe Champion, P. A. To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Curr. Opin. Microbiol.* **13**, 86–92 (2010).

241. Jordao, L., Bleck, C. K. E., Mayorga, L., Griffiths, G. & Anes, E. On the killing of mycobacteria by macrophages. *Cell. Microbiol.* **10**, 529–48 (2008).

242. Anes, E. *et al.* Dynamic life and death interactions between *Mycobacterium smegmatis* and J774 macrophages. *Cell. Microbiol.* **8**, 939–60 (2006).

243. Bange, F. C., Collins, F. M. & Jacobs, W. R. Survival of mice infected with *Mycobacterium smegmatis* containing large DNA fragments from *Mycobacterium tuberculosis*. *Tuber. Lung Dis.* **79**, 171–80 (1999).

244. Bouz, G. & Al Hasawi, N. The zebrafish model of tuberculosis - no lungs needed. *Crit. Rev. Microbiol.* **44**, 779–792 (2018).

245. Reyrat, J.-M. & Kahn, D. *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends Microbiol.* **9**, 472–473 (2001).

246. Barry, C. E. *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends Microbiol.* **9**, 473–474 (2001).

247. McGarvey, J. A., Wagner, D. & Bermudez, L. E. Differential gene expression in mononuclear phagocytes infected with pathogenic and non-pathogenic mycobacteria. *Clin. Exp. Immunol.* **136**, 490–500 (2004).

248. Schnappinger, D., Schoolnik, G. K. & Ehrt, S. Expression profiling of host pathogen interactions: how *Mycobacterium tuberculosis* and the macrophage adapt to one another. *Microbes Infect.* **8**, 1132–1140 (2006).

249. Guinn, K. M. & Rubin, E. J. Tuberculosis: Just the FAQs. *mBio* **8**, e01910-17 (2017).

250. Wee, W. Y., Dutta, A. & Choo, S. W. Comparative genome analyses of mycobacteria give better insights into their evolution. *PLoS One* **12**, e0172831 (2017).

251. Ahamad, N., Gupta, S. & Parashar, D. Using Omics to Study Leprosy, Tuberculosis, and Other Mycobacterial Diseases. *Front. Cell. Infect. Microbiol.* **12**, 792617 (2022).

252. Fondi, M. & Liò, P. Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiol. Res.* **171**, 52–64 (2015).

253. Rohde, K. H., Abramovitch, R. B. & Russell, D. G. *Mycobacterium tuberculosis* Invasion of Macrophages: Linking Bacterial Gene Expression to Environmental Cues. *Cell Host Microbe* **2**, 352–364 (2007).

254. Homolka, S., Niemann, S., Russell, D. G. & Rohde, K. H. Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: Delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathog.* **6**, 1–17 (2010).

255. Russell, D. G. *et al.* The Sculpting of the *Mycobacterium tuberculosis* Genome by Host Cell–Derived Pressures. *Microbiol. Spectr.* **2**, (2014).

256. Schnappinger, D. *et al.* Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages. *J. Exp. Med.* **198**, 693–704 (2003).

257. Rhee, K. Y. *et al.* Central carbon metabolism in *Mycobacterium tuberculosis*: an unexpected frontier. *Trends Microbiol.* **19**, 307–314 (2011).

258. Hards, K. *et al.* Two for the price of one: Attacking the energetic-metabolic hub of mycobacteria to produce new chemotherapeutic agents. *Prog. Biophys. Mol. Biol.* **152**, 35–44 (2020).

259. McKinney, J. D. *et al.* Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735–738 (2000).

260. Krebs, H. A. & Holzach, O. The conversion of citrate into *cis*-aconitate and isocitrate in the presence of aconitase. *Biochem. J.* **52**, 527–528 (1952).

261. Dolan, S. K. & Welch, M. The Glyoxylate Shunt, 60 Years On. *Annu. Rev. Microbiol.* **72**, 309–330 (2018).

262. Walsh, K. & Koshland, D. E. Determination of Flux through the Branch Point of Two Metabolic Cycles. *J. Biol. Chem.* **259**, 9646–9654 (1984).

263. Vanni, P., Giachetti, E., Pinzauti, G. & McFadden, B. A. Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **95**, 431–458 (1990).

264. Kornberg, H. L. Anaplerotic Sequences in Microbial Metabolism. *Angew. Chemie Int. Ed. English* **4**, 558–565 (1965).

265. Dunn, M. F., Ramírez-Trujillo, J. A. & Hernández-Lucas, I. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology* **155**, 3166–3175 (2009).

266. Nakazawa, M. *et al.* Molecular characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, in *Euglena gracilis*. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **141**, 445–452 (2005).

267. Murima, P. *et al.* A rheostat mechanism governs the bifurcation of carbon flux in mycobacteria. *Nat. Commun.* **7**, 12527 (2016).

268. Chung, T., Klumpp, D. J. & LaPorte, D. C. Glyoxylate bypass operon of *Escherichia coli*: cloning and determination of the functional map. *J. Bacteriol.* **170**, 386–392 (1988).

269. Zhang, C. & Anderson, A. J. The gluconeogenic pathway in a soil mycobacterium isolate with bioremediation ability. *Curr. Microbiol.* **66**, 122–131 (2013).

270. Ahn, S., Jung, J., Jang, I.-A., Madsen, E. L. & Park, W. Role of Glyoxylate Shunt in Oxidative Stress Response. *J. Biol. Chem.* **291**, 11928–11938 (2016).

271. Bhusal, R. P., Bashiri, G., Kwai, B. X. C., Sperry, J. & Leung, I. K. H. Targeting isocitrate lyase for the treatment of latent tuberculosis. *Drug Discov. Today* **22**, 1008–1016 (2017).

272. Upton, A. M. & McKinney, J. D. Role of the methylcitrate cycle in propionate metabolism and detoxification in *Mycobacterium smegmatis*. *Microbiology* **153**, 3973–3982 (2007).

273. Wipperman, M. F., Sampson, N. S. & Thomas, S. T. Pathogen roid rage: Cholesterol utilization by *Mycobacterium tuberculosis*. *Crit. Rev. Biochem. Mol. Biol.* **49**, 269–293 (2014).

274. Kunze, M., Pracharoenwattana, I., Smith, S. M. & Hartig, A. A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta - Mol. Cell Res.* **1763**, 1441–1452 (2006).

275. Eprintsev, A. T., Salnikov, A. V., Haba, A. M. & Zaichikova, M. V. Isocitrate lyase isozymes and their role in organisms with different levels of organization. *Biol. Bull. Rev. 2014 44* **4**, 323–334 (2014).

276. Terwilliger, T. . *et al.* The TB structural genomics consortium: a resource for *Mycobacterium tuberculosis* biology. *Tuberculosis* **83**, 223–249 (2003).

277. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–44 (1998).

278. Höner Zu Bentrup, K., Miczak, A., Swenson, D. L. & Russell, D. G. Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J. Bacteriol.* **181**, 7161–7167 (1999).

279. Muñoz-Elías, E. J. & McKinney, J. D. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat. Med.* **11**, 638–644 (2005).

280. Shukla, H. *et al.* Alternate pathway to ascorbate induced inhibition of *Mycobacterium tuberculosis*. *Tuberculosis* **111**, 161–169 (2018).

281. Michelucci, A. *et al.* Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7820–7825 (2013).

282. Eoh, H. & Rhee, K. Y. Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of *Mycobacterium tuberculosis* on fatty acids. *Proc. Natl. Acad. Sci.* **111**, 4976–4981 (2014).

283. Akao, T., Kusaka, T. & Kobashi, K. Two esterases released from *Mycobacterium smegmatis* from the hydrolysis of long chain acyl-CoAs and Tween. *J. Biochem.* **90**, 1661–9 (1981).

284. Gould, T. A., van de Langemheen, H., Munoz-Elias, E. J., McKinney, J. D. & Sacchettini, J. C. Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **61**, 940–947 (2006).

285. Munoz-Elias, E. J., Upton, A. M., Cherian, J. & McKinney, J. D. Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. *Mol. Microbiol.* **60**, 1109–1122 (2006).

286. Rohde, K. H., Veiga, D. F. T., Caldwell, S., Balázsi, G. & Russell, D. G. Linking the Transcriptional Profiles and the Physiological States of *Mycobacterium tuberculosis* during an Extended Intracellular Infection. *PLoS Pathog.* **8**, e1002769 (2012).

287. Dubnau, E., Fontán, P., Manganelli, R., Soares-Appel, S. & Smith, I. *Mycobacterium tuberculosis* Genes Induced during Infection of Human Macrophages. *Infect. Immun.* **70**, 2787–2795 (2002).

288. Timm, J. *et al.* Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14321–6 (2003).

289. Talaat, A. M. *et al.* Mycobacterial Bacilli Are Metabolically Active during Chronic Tuberculosis in Murine Lungs: Insights from Genome-Wide Transcriptional Profiling. *J. Bacteriol.* **189**, 4265–4274 (2007).

290. Bentrup, K. H. zu & Russell, D. G. Mycobacterial persistence: adaptation to a changing environment. *Trends Microbiol.* **9**, 597–605 (2001).

291. Beste, D. J. V *et al.* 13C Metabolic Flux Analysis Identifies an Unusual Route for Pyruvate Dissimilation in Mycobacteria which Requires Isocitrate Lyase and Carbon Dioxide Fixation. *PLoS Pathog.* **7**, e1002091 (2011).

292. Ehrt, S., Rhee, K. & Schnappinger, D. Mycobacterial genes essential for the pathogen's survival in the host. *Immunol. Rev.* **264**, 319–326 (2015).

293. Nandakumar, M., Nathan, C. & Rhee, K. Y. Isocitrate lyase mediates broad antibiotic tolerance in *Mycobacterium tuberculosis*. *Nat. Commun.* **5**, 4306 (2014).

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

295. Kumar, R. & Bhakuni, V. *Mycobacterium tuberculosis* isocitrate lyase (MtbIcl): Role of divalent cations in modulation of functional and structural properties. *Proteins Struct. Funct. Bioinforma.* **72**, 892–900 (2008).

296. Park, Y., Cho, Y., Lee, Y.-H., Lee, Y.-W. & Rhee, S. Crystal structure and functional analysis of isocitrate lyases from *Magnaporthe oryzae* and *Fusarium graminearum*. *J. Struct. Biol.* **194**, 395–403 (2016).

297. Jongkon, N., Chotpatiwetchkul, W. & Gleeson, M. P. Probing the Catalytic Mechanism Involved in the Isocitrate Lyase Superfamily: Hybrid Quantum Mechanical/Molecular Mechanical Calculations on 2,3-Dimethylmalate Lyase. *J. Phys. Chem. B* **119**, 11473–11484 (2015).

298. Müller, S. *et al.* Gene acquisition, duplication and metabolic specification: the evolution of fungal methylisocitrate lyases. *Environ. Microbiol.* **13**, 1534–1548 (2011).

299. Britton, K. *et al.* The crystal structure and active site location of isocitrate lyase from the fungus *Aspergillus nidulans*. *Structure* **8**, 349–362 (2000).

300. Britton, K. L. *et al.* The structure and domain organization of *Escherichia coli* isocitrate lyase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1209–1218 (2001).

301. Sharma, V. *et al.* Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nat. Struct. Biol.* **7**, 663–8 (2000).

302. Smith, C. V, Sharma, V. & Sacchettini, J. C. TB drug discovery: addressing issues of persistence and resistance. *Tuberculosis* **84**, 45–55 (2004).

303. Antil, M. & Gupta, V. Rv1915 and Rv1916 from *Mycobacterium tuberculosis* H37Rv form in vitro protein-protein complex. *Biochim. Biophys. Acta - Gen. Subj.* **1866**, 130130 (2022).

304. Antil, M. *et al.* Structure-function insights into elusive *Mycobacterium tuberculosis* protein Rv1916. *Int. J. Biol. Macromol.* **141**, 927–936 (2019).

305. Bhusal, R. P. *et al.* Acetyl-CoA-mediated activation of *Mycobacterium tuberculosis* isocitrate lyase 2. *Nat. Commun.* **10**, 4639 (2019).

306. Singh, V. & Ghosh, I. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets. *Theor. Biol. Med. Model.* **3**, 27 (2006).

307. Lee, Y.-V., Wahab, H. A. & Choong, Y. S. Potential Inhibitors for Isocitrate Lyase of *Mycobacterium tuberculosis* and Non-*M. tuberculosis* : A Summary. *Biomed Res. Int.* **2015**, 1–20 (2015).

308. Shukla, H., Shukla, R., Sonkar, A. & Tripathi, T. Alterations in conformational topology and interaction dynamics caused by L418A mutation leads to activity loss of *Mycobacterium tuberculosis* isocitrate lyase. *Biochem. Biophys. Res. Commun.* **490**, 276–282 (2017).

309. Shukla, H., Shukla, R., Sonkar, A., Pandey, T. & Tripathi, T. Distant Phe345 mutation compromises the stability and activity of *Mycobacterium tuberculosis* isocitrate lyase by modulating its structural flexibility. *Sci. Rep.* **7**, 1058 (2017).

310. Chaudhuri, B. N. *et al.* The Crystal Structure of the First Enzyme in the Pantothenate Biosynthetic Pathway, Ketopantoate Hydroxymethyltransferase, from *M. tuberculosis*. *Structure* **11**, 753–764 (2003).

311. Shukla, H. *et al.* Insight into the structural flexibility and function of *Mycobacterium tuberculosis* isocitrate lyase. *Biochimie* **110**, 73–80 (2015).

312. Lee, Y. V., Choi, S. B., Wahab, H. A. & Choong, Y. S. Active Site Flexibility of *Mycobacterium tuberculosis* Isocitrate Lyase in Dimer Form. *J. Chem. Inf. Model.* **57**, 2351–2357 (2017).

313. Shukla, R., Shukla, H. & Tripathi, T. Activity loss by H46A mutation in *Mycobacterium tuberculosis* isocitrate lyase is due to decrease in structural plasticity and collective motions of the active site. *Tuberculosis (Edinb).* **108**, 143–150 (2018).

314. Krátký, M., Novotná, E., Stolaříková, J., Švarcová, M. & Vinšová, J. Substituted *N*-phenylitaconamides as inhibitors of mycobacteria and mycobacterial isocitrate lyase. *Eur. J. Pharm. Sci.* **176**, 106252 (2022).

315. Kwon, S., Chun, H. L., Ha, H. J., Lee, S. Y. & Park, H. H. Heterogeneous multimeric structure of isocitrate lyase in complex with succinate and itaconate provides novel insights into its inhibitory mechanism. *PLoS One* **16**, e0251067 (2021).

316. Pham, T. V. *et al.* Covalent Inactivation of *Mycobacterium tuberculosis* Isocitrate Lyase by *cis*-2,3-Epoxy-Succinic Acid. *ACS Chem. Biol.* **16**, 463–470 (2021).

317. Pham, T. V *et al.* Mechanism-based inactivator of isocitrate lyases 1 and 2 from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* **114**, 7617–7622 (2017).

318. Moynihan, M. M. & Murkin, A. S. Cysteine Is the General Base That Serves in Catalysis by Isocitrate Lyase and in Mechanism-Based Inhibition by 3-Nitropropionate. *Biochemistry* **53**, 178–187 (2014).

319. Quartararo, C. E., Hadi, T., Cahill, S. M. & Blanchard, J. S. Solvent Isotope-Induced Equilibrium Perturbation for Isocitrate Lyase. *Biochemistry* **52**, 9286–9293 (2013).

320. Ibeji, C. U. *et al.* Demystifying the catalytic pathway of *Mycobacterium tuberculosis* isocitrate lyase. *Sci. Rep.* **10**, 18925 (2020).

321. Shukla, H. *et al.* Isocitrate lyase of *Mycobacterium tuberculosis* is inhibited by quercetin through binding at *N*-terminus. *Int. J. Biol. Macromol.* **78**, 137–141 (2015).

322. Shingnapurkar, D. *et al.* Synthesis and characterization of pyruvate–isoniazid analogs and their copper complexes as potential ICL inhibitors. *Bioorg. Med. Chem. Lett.* **22**, 3172–3176 (2012).

323. Ji, L., Long, Q., Yang, D. & Xie, J. Identification of Mannich Base as a Novel Inhibitor of *Mycobacterium tuberculosis* Isocitrate by High-Throughput Screening. *Int. J. Biol. Sci.* **7**, 376–382 (2011).

324. Sriram, D. *et al.* Synthesis and Antimycobacterial Evaluation of Novel Phthalazin-4-ylacetamides Against log- and Starved Phase Cultures. *Chem. Biol. Drug Des.* **75**, 381–391 (2010).

325. Sriram, D., Yogeeswari, P., Senthilkumar, P., Naidu, G. & Bhat, P. 5-Nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides: synthesis, *in vitro* antimycobacterial activity, cytotoxicity, and isocitrate lyase inhibition studies. *J. Enzyme Inhib. Med. Chem.* **25**, 765–772 (2010).

326. Sriram, D. *et al.* 5-Nitro-2-furoic acid hydrazones: Design, synthesis and *in vitro* antimycobacterial evaluation against log and starved phase cultures. *Bioorg. Med. Chem. Lett.* **20**, 4313–4316 (2010).

327. Sriram, D. *et al.* Novel phthalazinyl derivatives: synthesis, antimycobacterial activities, and inhibition of *Mycobacterium tuberculosis* isocitrate lyase enzyme. *Med. Chem.* **5**, 422–422 (2009).

328. Sriram, D. *et al.* Synthesis of various 3-nitropropionamides as *Mycobacterium tuberculosis* isocitrate lyase inhibitor. *Bioorg. Med. Chem. Lett.* **21**, 5149–5154 (2011).

329. Bhusal, R. P. *et al.* Development of NMR and thermal shift assays for the evaluation of *Mycobacterium tuberculosis* isocitrate lyase inhibitors. *Medchemcomm* **8**, 2155–2163 (2017).

330. Mdluli, K. & Spigelman, M. Novel targets for tuberculosis drug discovery. *Curr. Opin. Pharmacol.* **6**, 459–67 (2006).

331. Mellott, D. M. *et al.* Mechanism-Based Inactivation of *Mycobacterium tuberculosis* Isocitrate Lyase 1 by (2R,3S)-2-Hydroxy-3-(nitromethyl)succinic acid. *J. Am. Chem. Soc.* **143**, 17666–17676 (2021).

332. Kwai, B. X. C. *et al.* Itaconate is a covalent inhibitor of the *Mycobacterium tuberculosis* isocitrate lyase. *RSC Med. Chem.* **12**, 57–61 (2021).

333. Ray, S., Kreitler, D. F., Gulick, A. M. & Murkin, A. S. The Nitro Group as a Masked Electrophile in Covalent Enzyme Inhibition. *ACS Chem. Biol.* **13**, 1470–1473 (2018).

334. Liu, Y. *et al.* Identification of a novel inhibitor of isocitrate lyase as a potent antitubercular agent against both active and non-replicating *Mycobacterium tuberculosis*. *Tuberculosis* **97**, 38–46 (2016).

335. Duan, C. *et al.* Discovery of a Novel Inhibitor Structure of *Mycobacterium tuberculosis* Isocitrate Lyase. *Molecules* **27**, 2447 (2022).

336. Lee, Y.-V., Choi, S. B., Wahab, H. A., Lim, T. S. & Choong, Y. S. Applications of Ensemble Docking in Potential Inhibitor Screening for *Mycobacterium tuberculosis* Isocitrate Lyase Using a Local Plant Database. *J. Chem. Inf. Model.* **59**, 2487–2495 (2019).

337. Tiwari, A., Kumar, A., Srivastava, G. & Sharma, A. Screening of Anti-mycobacterial Phytochemical Compounds for Potential Inhibitors against *Mycobacterium tuberculosis* Isocitrate Lyase. *Curr. Top. Med. Chem.* **19**, 600–608 (2019).

338. Shukla, R., Shukla, H., Sonkar, A., Pandey, T. & Tripathi, T. Structure-based screening and molecular dynamics simulations offer novel natural compounds as potential inhibitors of *Mycobacterium tuberculosis* isocitrate lyase. *J. Biomol. Struct. Dyn.* **36**, 2045–2057 (2018).

339. Liu, X., Zang, Y., Sun, B. & Yin, Y. Optimization of phage heptapeptide library-screening process for developing inhibitors of the isocitrate lyase homologue from *Mycobacterium tuberculosis*. *Med. Chem. Res.* **23**, 2543–2553 (2014).

340. Alston, T. A., Mela, L. & Bright, H. J. 3-Nitropropionate, the toxic substance of *Indigofera*, is a suicide inactivator of succinate dehydrogenase. *Proc. Natl. Acad. Sci.* **74**, 3767–71 (1977).

341. Xiong, X. *et al.* Comparative Genomics of a Bovine *Mycobacterium tuberculosis* Isolate and Other Strains Reveals Its Potential Mechanism of Bovine Adaptation. *Front. Microbiol.* **8**, 2500 (2017).

342. Pradheep, S. Arvindh; Narayanan, Sujatha; Vasan, S. K.; Narayanan, P. R. Cloning and expression of *aceA* gene encoding isocitrate lyase from *Mycobacterium tuberculosis*. *Curr. Sci.* **79**, 1585–1588 (2000).

343. Antil, M., Gouin, S. G. & Gupta, V. Truncation of *C*-Terminal Intrinsically Disordered Region of Mycobacterial Rv1915 Facilitates Production of "Difficult-to-Purify" Recombinant Drug Target. *Front. Bioeng. Biotechnol.* **8**, 522 (2020).

344. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.* **43**, 717–731 (2002).

345. Kelkar, D. S. *et al.* Proteogenomic Analysis of *Mycobacterium tuberculosis* By High Resolution Mass Spectrometry. *Mol. Cell. Proteomics* **10**, M111.011445 (2011).

346. Takamura, Y. & Nomura, G. Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of *Escherichia coli* K12. *J. Gen. Microbiol.* **134**, 2249–2253 (1988).

347. Bharadwaj, V. P. *et al.* Metabolic studies on mycobacteria. IV. Assay of isocitrate lyase and malate synthase activity in *M. leprae*. *Indian J. Lepr.* **59**, 158–162 (1987).

348. Kannan, K. B. *et al.* Metabolic studies on mycobacteria. II. Glyoxylate by-pass (TCA cycle) enzymes of slow and fast growing mycobacteria. *Indian J. Lepr.* **57**, 542–548 (1985).

349. Wheeler, R. P. & Ratledge, C. Use of carbon sources for lipid biosynthesis in *Mycobacterium leprae*: a comparison with other pathogenic mycobacteria. *J. Gen. Microbiol.* **134**, 2111–2121 (1988).

350. Ko, E.-M. *et al.* Regulation of the *icl1* Gene Encoding the Major Isocitrate Lyase in *Mycobacterium smegmatis*. *J. Bacteriol.* **203**, e00402-21 (2021).

351. Xu, Y., You, D. & Ye, B.-C. RegX3 Controls Glyoxylate Shunt and Mycobacteria Survival by Directly Regulating the Transcription of Isocitrate Lyase Gene in *Mycobacterium smegmatis*. *ACS Infect. Dis.* **7**, 927–936 (2021).

352. Qi, N., She, G.-L., Du, W. & Ye, B.-C. *Mycobacterium smegmatis* GlnR Regulates the Glyoxylate Cycle and the Methylcitrate Cycle on Fatty Acid Metabolism by Repressing *icl* Transcription. *Front. Microbiol.* **12**, 603835 (2021).

353. Kwofie, S. *et al. In Silico* Screening of Isocitrate Lyase for Novel Anti-Buruli Ulcer Natural Products Originating from Africa. *Molecules* **23**, 1550 (2018).

354. Kozic, J. *et al.* Synthesis and *in vitro* antimycobacterial and isocitrate lyase inhibition properties of novel 2-methoxy-2'-hydroxybenzanilides, their thioxo analogues and benzoxazoles. *Eur. J. Med. Chem.* **56**, 108–119 (2012).

355. Wang, P., Lv, C. & Zhu, G. Novel Type II and Monomeric NAD^+ Specific Isocitrate Dehydrogenases: Phylogenetic Affinity, Enzymatic Characterization and Evolutionary Implication. *Sci. Rep.* **5**, 9150 (2015).

356. Meléndez-Hevia, E., Waddell, T. G. & Cascante, M. The puzzle of the Krebs citric acid cycle: Assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *J. Mol. Evol.* **43**, 293–303 (1996).

357. Yasutake, Y. *et al.* Crystal Structure of the Monomeric Isocitrate Dehydrogenase in the Presence of $NADP^+$. *J. Biol. Chem.* **278**, 36897–36904 (2003).

358. Imabayashi, F., Aich, S., Prasad, L. & Delbaere, L. T. J. Substrate-free structure of a monomeric NADP isocitrate dehydrogenase: An open conformation phylogenetic relationship of isocitrate dehydrogenase. *Proteins Struct. Funct. Bioinforma.* **63**, 100–112 (2006).

359. Tang, W., Song, P., Cao, Z., Wang, P. & Zhu, G. A unique homodimeric NAD^+-linked isocitrate dehydrogenase from the smallest autotrophic eukaryote *Ostreococcus tauri*. *FASEB J.* **29**, 2462–72 (2015).

360. Huang, S. *et al.* Biochemical and Phylogenetic Characterization of a Novel $NADP^+$-Specific Isocitrate Dehydrogenase From the Marine Microalga *Phaeodactylum tricornutum*. *Front. Mol. Biosci.* **8**, 702083 (2021).

361. Zhu, G., Golding, G. B. & Dean, A. M. The selective cause of an ancient adaptation. *Science* **307**, 1279–82 (2005).

362. Aktas, D. F. & Cook, P. F. A lysine-tyrosine pair carries out acid-base chemistry in the metal ion-dependent pyridine dinucleotide-linked beta-hydroxyacid oxidative decarboxylases. *Biochemistry* **48**, 3565–77 (2009).

363. Quartararo, C. E., Hazra, S., Hadi, T. & Blanchard, J. S. Structural, Kinetic and Chemical Mechanism of Isocitrate Dehydrogenase-1 from *Mycobacterium tuberculosis*. *Biochemistry* **52**, 1765–1775 (2013).

364. Gonçalves, S., Miller, S. P., Carrondo, M. A., Dean, A. M. & Matias, P. M. Induced Fit and the Catalytic Mechanism of Isocitrate Dehydrogenase. *Biochemistry* **51**, 7098 (2012).

365. Wang, P., Wang, Y., Guo, X., Huang, S. & Zhu, G. Biochemical and phylogenetic characterization of a monomeric isocitrate dehydrogenase from a marine methanogenic archaeon *Methanococcoides methylutens*. *Extremophiles* **24**, 319–328 (2020).

366. Tang, W. *et al.* Crystal structures of NAD^+-linked isocitrate dehydrogenase from the green alga *Ostreococcus tauri* and its evolutionary relationship with eukaryotic $NADP^+$-linked homologs. *Arch. Biochem. Biophys.* **708**, 108898 (2021).

367. Clark, O., Yen, K. & Mellinghoff, I. K. Molecular Pathways: Isocitrate Dehydrogenase Mutations in Cancer. *Clin. Cancer Res.* **22**, 1837–42 (2016).

368. Johnson, B. E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).

369. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–66 (2009).

370. Tommasini-Ghelfi, S. *et al.* Cancer-associated mutation and beyond: The emerging biology of isocitrate dehydrogenases in human disease. *Sci. Adv.* **5**, 4543–4565 (2019).

371. Reitman, Z. J. & Yan, H. Isocitrate Dehydrogenase 1 and 2 Mutations in Cancer: Alterations at a Crossroads of Cellular Metabolism. *JNCI J. Natl. Cancer Inst.* **102**, 932–941 (2010).

372. Wang, P. *et al.* Crystal Structure of the Isocitrate Dehydrogenase 2 from *Acinetobacter baumannii* (AbIDH2) Reveals a Novel Dimeric Structure with Two Monomeric-IDH-Like Subunits. *Int. J. Mol. Sci.* **19**, 1131 (2018).

373. Zhang, B. *et al.* Enzymatic characterization of a monomeric isocitrate dehydrogenase from *Streptomyces lividans* TK54. *Biochimie* **91**, 1405–10 (2009).

374. Huang, S.-P., Cheng, H.-M., Wang, P. & Zhu, G.-P. Biochemical Characterization and Complete Conversion of Coenzyme Specificity of Isocitrate Dehydrogenase from *Bifidobacterium longum*. *Int. J. Mol. Sci.* **17**, 296 (2016).

375. Dean, A. M. & Golding, G. B. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3104–9 (1997).

376. Knoll, A. H. The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627 (1992).

377. Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470–477 (1996).

378. Armingol, E., Tobar, E. & Cabrera, R. Understanding the impact of the cofactor swapping of isocitrate dehydrogenase over the growth phenotype of *Escherichia coli* on acetate by using constraint-based modeling. *PLoS One* **13**, e0196182 (2018).

379. Banerjee, S., Nandyala, A., Podili, R., Katoch, V. M. & Hasnain, S. E. Comparison of *Mycobacterium tuberculosis* isocitrate dehydrogenases (ICD-1 and ICD-2) reveals differences in coenzyme affinity, oligomeric state, pH tolerance and phylogenetic affiliation. *BMC Biochem.* **6**, 20 (2005).

380. Cheng, Y. S., Sacchettini, J. C. & TB Structural Genomics Consortium. *RCSB PDB 5KVU: Crystal structure of isocitrate dehydrogenase-2 in complex with NADP(+) from Mycobacterium tuberculosis*. (2017). doi:10.2210/pdb5kvu/pdb

381. Hatzopoulos, G. N., Kefala, G. & Mueller-Dieckmann, J. Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of isocitrate dehydrogenase 2 (Rv0066c) from *Mycobacterium tuberculosis*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64**, 1139–1142 (2008).

382. Wheeler, P. R. Oxidation of carbon sources through the tricarboxylic acid cycle in *Mycobacterium leprae* grown in armadillo liver. *J. Gen. Microbiol.* **130**, 381–389 (1984).

383. Katoch, V. M. *et al.* Metabolic studies on mycobacteria. III. Demonstration of key enzymes of TCA cycle in *M. leprae*. *Indian J. Lepr.* **59**, 152–7 (1987).

384. Sharma, V. D. *et al.* Metabolic studies on mycobacteria-I. Demonstration of key enzymes of glycolysis and tricarboxylic acid cycle on polyacrylamide gels. *Indian J. Lepr.* **57**, 534–41 (1985).

385. Ohman, R. & Ridell, M. Purification and characterisation of isocitrate dehydrogenase and malate dehydrogenase from *Mycobacterium tuberculosis* and evaluation of their potential as suitable antigens for the serodiagnosis of tuberculosis. *Tuber. Lung Dis.* **77**, 454–61 (1996).

386. Florio, W. *et al.* Identification, Molecular Cloning, and Evaluation of Potential Use of Isocitrate Dehydrogenase II of *Mycobacterium bovis* BCG in Serodiagnosis of Tuberculosis. *Clin. Vaccine Immunol.* **9**, 846–851 (2002).

387. Pojer, F., Murima, P. & McKinney, J. D. *RCSB PDB - 4ZDA: Crystal structure of isocitrate dehydrogenase in complex with isocitrate and Mn from M. smegmatis*. (2015). doi:10.2210/pdb4zda/pdb

388. Dhariwal, K. R. & Venkitasubramanian, T. A. NADP-specific isocitrate dehydrogenase of *Mycobacterium phlei* ATCC 354: purification and characterization. *J. Gen. Microbiol.* **133**, 2457–60 (1987).

389. Warner, D. F. *Mycobacterium tuberculosis* Metabolism. *Cold Spring Harb. Perspect. Med.* **5**, a021121 (2015).

390. Micklinghoff, J. C. *et al.* Role of the transcriptional regulator RamB (Rv0465c) in the control of the glyoxylate cycle in *Mycobacterium tuberculosis*. *J. Bacteriol.* **191**, 7260–9 (2009).

391. Datta, P., Shi, L., Bibi, N., Balázsi, G. & Gennaro, M. L. Regulation of central metabolism genes of *Mycobacterium tuberculosis* by parallel feed-forward loops controlled by sigma factor E (σ(E)). *J. Bacteriol.* **193**, 1154–60 (2011).

392. Gerstmeir, R., Cramer, A., Dangel, P., Schaffer, S. & Eikmanns, B. J. RamB, a novel transcriptional regulator of genes involved in acetate metabolism of *Corynebacterium glutamicum*. *J. Bacteriol.* **186**, 2798–809 (2004).

393. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–50 (2016).

394. Sachdeva, P., Misra, R., Tyagi, A. K. & Singh, Y. The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J.* **277**, 605–26 (2010).

395. Fontán, P. A. *et al.* The *Mycobacterium tuberculosis* sigma factor sigmaB is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for in vivo growth. *J. Bacteriol.* **191**, 5628–33 (2009).

396. Manganelli, R., Voskuil, M. I., Schoolnik, G. K. & Smith, I. The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol. Microbiol.* **41**, 423–37 (2001).

397. Ascensao, J. A. *et al.* Non-monotonic Response to Monotonic Stimulus: Regulation of Glyoxylate Shunt Gene-Expression Dynamics in *Mycobacterium tuberculosis*. *PLOS Comput. Biol.* **12**, e1004741 (2016).

398. Masiewicz, P., Brzostek, A., Wolański, M., Dziadek, J. & Zakrzewska-Czerwińska, J. A Novel Role of the PrpR as a Transcription Factor Involved in the Regulation of Methylcitrate Pathway in *Mycobacterium tuberculosis*. *PLoS One* **7**, e43651 (2012).

399. Liu, W.-B., Liu, X.-X., Shen, M.-J., She, G.-L. & Ye, B.-C. The Nitrogen Regulator GlnR Directly Controls Transcription of the *prpDBC* Operon Involved in Methylcitrate Cycle in *Mycobacterium smegmatis*. *J. Bacteriol.* **201**, e00099-19 (2019).

400. Dean, A. M., Lee, M. H. I. & Koshland, D. E. Phosphorylation Inactivates *Escherichia coli* Isocitrate Dehydrogenase by Preventing Isocitrate Binding. *J. Biol. Chem.* **264**, 20482–20486 (1989).

401. Crousilles, A., Dolan, S. K., Brear, P., Chirgadze, D. Y. & Welch, M. Gluconeogenic precursor availability regulates flux through the glyoxylate shunt in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **293**, 14260–14269 (2018).

402. Cozzone, A. J. & El-Mansi, M. Control of Isocitrate Dehydrogenase Catalytic Activity by Protein Phosphorylation in *Escherichia coli*. *Microb. Physiol.* **9**, 132–146 (2005).

403. Nimmo, G. A. & Nimmo, H. G. The regulatory properties of isocitrate dehydrogenase kinase and isocitrate dehydrogenase phosphatase from *Escherichia coli* ML308 and the roles of these activities in the control of isocitrate dehydrogenase. *Eur. J. Biochem.* **141**, 409–14 (1984).

404. Soundarajan, B. T., Santanu, D. & Ghosh, I. Method. (2004).

405. Vinekar, R. & Ghosh, I. Determination of phosphorylation sites for NADP-specific isocitrate dehydrogenase from *Mycobacterium tuberculosis*. *J. Biomol. Struct. Dyn.* **26**, 741–754 (2009).

406. Gil, M. *et al.* New substrates and interactors of the mycobacterial Serine/Threonine protein kinase PknG identified by a tailored interactomic approach. *J. Proteomics* **192**, 321–333 (2019).

407. Nakedi, K. C. *et al.* Identification of Novel Physiological Substrates of *Mycobacterium bovis* BCG Protein Kinase G (PknG) by Label-free Quantitative Phosphoproteomics. *Mol. Cell. Proteomics* **17**, 1365–1377 (2018).

408. Wu, F.-L. *et al.* The Ser/Thr Protein Kinase Protein-Protein Interaction Map of *M. tuberculosis*. *Mol. Cell. Proteomics* **16**, 1491–1506 (2017).

409. Baros, S. S., Blackburn, J. M. & Soares, N. C. Phosphoproteomic Approaches to Discover Novel Substrates of Mycobacterial Ser/Thr Protein Kinases. *Mol. Cell. Proteomics* **19**, 233–244 (2020).

410. Venkat, S. *et al.* Characterizing Lysine Acetylation of Isocitrate Dehydrogenase in *Escherichia coli*. *J. Mol. Biol.* **430**, 1901–1911 (2018).

411. Wang, B. *et al.* SIRT2-dependent IDH1 deacetylation inhibits colorectal cancer and liver metastases. *EMBO Rep.* **21**, e48183 (2020).

412. Savoia, M. *et al.* P300/CBP-associated factor regulates transcription and function of isocitrate dehydrogenase 2 during muscle differentiation. *FASEB J.* **33**, 4107–4123 (2019).

413. Liao, X. *et al.* Poplar acetylome profiling reveals lysine acetylation dynamics in seasonal bud dormancy release. *Plant. Cell Environ.* **44**, 1830–1845 (2021).

414. Bi, J. *et al.* Modulation of Central Carbon Metabolism by Acetylation of Isocitrate Lyase in *Mycobacterium tuberculosis*. *Sci. Rep.* **7**, 44826 (2017).

415. Lee, W., VanderVen, B. C., Walker, S. & Russell, D. G. Novel protein acetyltransferase, Rv2170, modulates carbon and energy metabolism in *Mycobacterium tuberculosis*. *Sci. Rep.* **7**, 72 (2017).

416. Xie, L. *et al.* Proteome-wide lysine acetylation profiling of the human pathogen *Mycobacterium tuberculosis*. *Int. J. Biochem. Cell Biol.* **59**, 193–202 (2015).

417. Imber, M., Pietrzyk-Brzezinska, A. J. & Antelmann, H. Redox regulation by reversible protein *S*-thiolation in Gram-positive bacteria. *Redox Biol.* **20**, 130–145 (2019).

418. Hillion, M. *et al.* Monitoring global protein thiol-oxidation and protein *S*-mycothiolation in *Mycobacterium smegmatis* under hypochlorite stress. *Sci. Rep.* **7**, 1195 (2017).

419. Zhou, M., Xie, L., Yang, Z., Zhou, J. & Xie, J. Lysine succinylation of *Mycobacterium tuberculosis* isocitrate lyase (ICL) fine-tunes the microbial resistance to antibiotics. *J. Biomol. Struct. Dyn.* **35**, 1030–1041 (2017).

420. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).

421. Laval, T., Chaumont, L. & Demangel, C. Not too fat to fight: The emerging role of macrophage fatty acid metabolism in immunity to *Mycobacterium tuberculosis*. *Immunol. Rev.* **301**, 84–97 (2021).

422. Shi, L., Jiang, Q., Bushkin, Y., Subbian, S. & Tyagi, S. Biphasic Dynamics of Macrophage Immunometabolism during *Mycobacterium tuberculosis* Infection. *mBio* **10**, e02550-18 (2019).

423. Galván-Peña, S. & O'Neill, L. A. J. Metabolic reprograming in macrophage polarization. *Front. Immunol.* **5**, 420 (2014).

424. Rodríguez-Prados, J.-C. *et al.* Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* **185**, 605–14 (2010).

425. Warburg, O. On the Origin of Cancer Cells. *Science* **123**, 309–314 (1956).

426. Vrieling, F. *et al.* Analyzing the impact of *Mycobacterium tuberculosis* infection on primary human macrophages by combined exploratory and targeted metabolomics. *Sci. Rep.* **10**, 7085 (2020).

427. Russell, D. G., Cardona, P.-J., Kim, M.-J., Allain, S. & Altare, F. Foamy macrophages and the progression of the human tuberculosis granuloma. *Nat. Immunol.* **10**, 943–8 (2009).

428. Pagán, A. J. & Ramakrishnan, L. The Formation and Function of Granulomas. *Annu. Rev. Immunol.* **36**, 639–665 (2018).

429. Peyron, P. *et al.* Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for *M. tuberculosis* persistence. *PLoS Pathog.* **4**, e1000204 (2008).

430. Sholeye, A. R. *et al.* Tuberculous Granuloma: Emerging Insights From Proteomics and Metabolomics. *Front. Neurol.* **13**, 804838 (2022).

431. Kim, M.-J. *et al.* Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO Mol. Med.* **2**, 258–74 (2010).

432. Somashekar, B. S. *et al.* Metabolic profiling of lung granuloma in *Mycobacterium tuberculosis* infected guinea pigs: ex vivo ^{1}H magic angle spinning NMR studies. *J. Proteome Res.* **10**, 4186–4195 (2011).

433. Somashekar, B. S. *et al.* Metabolomic signatures in guinea pigs infected with epidemic-associated W-Beijing strains of *Mycobacterium tuberculosis*. *J. Proteome Res.* **11**, 4873–4884 (2012).

434. Shin, J. H. *et al.* ^{1}H NMR-based metabolomic profiling in mice infected with *Mycobacterium tuberculosis*. *J. Proteome Res.* **10**, 2238–2247 (2011).

435. Beste, D. J. V. *et al.* ^{13}C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chem. Biol.* **20**, 1012–1021 (2013).

436. Lv, J., Huang, Z., Luo, L., Zhang, S. & Wang, Y. Advances in Molecular and Microscale Characterization of Soil Organic Matter: Current Limitations and Future Prospects. *Environ. Sci. Technol.* **56**, 12793–12810 (2022).

437. Aksoy, E., Yigini, Y. & Montanarella, L. Combining Soil Databases for Topsoil Organic Carbon Mapping in Europe. *PLoS One* **11**, e0152098 (2016).

438. Rial, M., Martínez Cortizas, A. & Rodríguez-Lado, L. Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. *Sci. Total Environ.* **609**, 1411–1422 (2017).

439. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).

440. Weiss, M. S. *et al.* Molecular Architecture and Electrostatic Properties of a Bacterial Porin. *Science* **254**, 1627–1630 (1991).

441. Hedges, J. . & Oades, J. . Comparative organic geochemistries of soils and marine sediments. *Org. Geochem.* **27**, 319–361 (1997).

442. Gunina, A. & Kuzyakov, Y. Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. *Soil Biol. Biochem.* **90**, 87–100 (2015).

443. IUSS Working Group WRB. *World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106.* (2015).

444. Tian, L., Dell, E. & Shi, W. Chemical composition of dissolved organic matter in agroecosystems: Correlations with soil enzyme activity and carbon and nitrogen mineralization. *Appl. Soil Ecol.* **46**, 426–435 (2010).

445. Vranova, V., Rejsek, K. & Formanek, P. Aliphatic, Cyclic, and Aromatic Organic Acids, Vitamins, and Carbohydrates in Soil: A Review. *Sci. World J.* **2013**, 1–15 (2013).

446. Fischer, H., Meyer, A., Fischer, K. & Kuzyakov, Y. Carbohydrate and amino acid composition of dissolved organic matter leached from soil. *Soil Biol. Biochem.* **39**, 2926–2935 (2007).

447. Klemm, D., Heublein, B., Fink, H. P. & Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angew. Chemie Int. Ed.* **44**, 3358–3393 (2005).

448. Fujii, K., Hayakawa, C., Van Hees, P. A. W., Funakawa, S. & Kosaki, T. Biodegradation of low molecular weight organic compounds and their contribution to heterotrophic soil respiration in three Japanese forest soils. *Plant Soil* **334**, 475–489 (2010).

449. van Hees, P. A. W., Dahlen, J., Lundstrom U. S., Boren, H. & Allard, B. Determination of low molecular weight organic acids in soil solution by HPLC. *Talanta* **48**, 173–179 (1999).

450. Van Hees, P. A. W., Jones, D. L., Finlay, R., Godbold, D. L. & Lundström, U. S. The carbon we do not see—the impact of low molecular weight compounds on carbon dynamics and respiration in forest soils: a review. *Soil Biol. Biochem.* **37**, 1–13 (2005).

451. Jones, D. L. Organic acids in the rhizosphere-a critical review. *Plant Soil* **205**, 25–44 (1998).

452. Vogel, C. S. & Dawson, J. O. In vitro growth of five *Frankia* isolates in the presence of four phenolic acids and juglone. *Soil Biol. Biochem.* **18**, 227–231 (1986).

453. Serafini, A. *et al. Mycobacterium tuberculosis* requires glyoxylate shunt and reverse methylcitrate cycle for lactate and pyruvate metabolism. *Mol. Microbiol.* **112**, 1284–1307 (2019).

454. Hori, T., Noll, M., Igarashi, Y., Friedrich, M. W. & Conrad, R. Identification of acetate-assimilating microorganisms under methanogenic conditions in anoxic rice field soil by comparative stable isotope probing of RNA. *Appl. Environ. Microbiol.* **73**, 101–109 (2007).

455. St. James, A. R., Yavitt, J. B., Zinder, S. H. & Richardson, R. E. Linking microbial Sphagnum degradation and acetate mineralization in acidic peat bogs: from global insights to a genome-centric case study. *ISME J.* **15**, 293–303 (2021).

456. Zhu, X. *et al.* Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion ecosystem. *Microbiome* **8**, 22 (2020).

457. Jones, D. L. *et al.* Rapid intrinsic rates of amino acid biodegradation in soils are unaffected by agricultural management strategy. *Soil Biol. Biochem.* **37**, 1267–1275 (2005).

458. Hobbie, J. E. & Hobbie, E. A. Microbes in nature are limited by carbon and energy: the starving-survival lifestyle in soil and consequences for estimating microbial rates. *Front. Microbiol.* **4**, 324 (2013).

459. Auti, A. M., Narwade, N. P., Deshpande, N. M. & Dhotre, D. P. Microbiome and imputed metagenome study of crude and refined petroleum-oil-contaminated soils: Potential for hydrocarbon degradation and plant-growth promotion. *J. Biosci.* **44**, 114 (2019).

460. Haritash, A. K. & Kaushik, C. P. Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. *J. Hazard. Mater.* **169**, 1–15 (2009).

461. Kotoky, R., Rajkumari, J. & Pandey, P. The rhizosphere microbiome: Significance in rhizoremediation of polyaromatic hydrocarbon contaminated soil. *J. Environ. Manage.* **217**, 858–870 (2018).

462. Ahmad, M. *et al.* Impacts of bio-stimulants on pyrene degradation, prokaryotic community compositions, and functions. *Environ. Pollut.* **289**, 117863 (2021).

463. Rehmann, K., Noll, H. P., Steinberg, C. E. W. & Kettrup, A. A. Pyrene degradation by *Mycobacterium* sp. strain KR2. *Chemosphere* **36**, 2977–2992 (1998).

464. Khan, A. A., Kim, S.-J., Paine, D. D. & Cerniglia, C. E. Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. *Int. J. Syst. Evol. Microbiol.* **52**, 1997–2002 (2002).

465. McLellan, S. L., Warshawsky, D. & Shann, J. R. The effect of polycyclic aromatic hydrocarbons on the degradation of benzo[a]pyrene by *Mycobacterium* sp. strain RJGII-135. *Environ. Toxicol. Chem.* **21**, 253–259 (2002).

466. Bogan, B. ., Lahner, L. M., Sullivan, W. R. & Paterek, J. R. Degradation of straight-chain aliphatic and high-molecular-weight polycyclic aromatic hydrocarbons by a strain of *Mycobacterium austroafricanum*. *J. Appl. Microbiol.* **94**, 230–239 (2003).

467. Lee, S.-E., Seo, J.-S., Keum, Y.-S., Lee, K.-J. & Li, Q. X. Fluoranthene metabolism and associated proteins in *Mycobacterium* sp. JS14. *Proteomics* **7**, 2059–2069 (2007).

468. Fredslund, L. *et al.* Surface motility of polycyclic aromatic hydrocarbon (PAH)-degrading mycobacteria. *Res. Microbiol.* **159**, 255–262 (2008).

469. Kweon, O. *et al.* Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus *Mycobacterium*. *BMC Evol. Biol.* **15**, 21 (2015).

470. Hennessee, C. T., Seo, J.-S., Alvarez, A. M. & Li, Q. X. Polycyclic aromatic hydrocarbon-degrading species isolated from Hawaiian soils: *Mycobacterium crocinum* sp. nov., *Mycobacterium pallens* sp. nov., *Mycobacterium rutilum* sp. nov., *Mycobacterium rufum* sp. nov. and *Mycobacterium aromaticivorans* sp. nov. *Int. J. Syst. Evol. Microbiol.* **59**, 378–387 (2009).

471. de Carvalho, L. P. S. *et al.* Metabolomics of *Mycobacterium tuberculosis* Reveals Compartmentalized Co-Catabolism of Carbon Substrates. *Chem. Biol.* **17**, 1122–1131 (2010).

472. Borah, K. *et al.* Metabolic fluxes for nutritional flexibility of *Mycobacterium tuberculosis*. *Mol. Syst. Biol.* **17**, e10280 (2021).

473. Lovewell, R. R., Sassetti, C. M. & VanderVen, B. C. Chewing the fat: lipid metabolism and homeostasis during *M. tuberculosis* infection. *Curr. Opin. Microbiol.* **29**, 30–36 (2016).

474. Pandey, A. K. & Sassetti, C. M. Mycobacterial persistence requires the utilization of host cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4376–80 (2008).

475. Aguilar-Ayala, D. A. *et al.* The transcriptome of *Mycobacterium tuberculosis* in a lipid-rich dormancy model through RNAseq analysis. *Sci. Rep.* **7**, 1–13 (2017).

476. Bloch, H. & Segal, W. Biochemical differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. *J. Bacteriol.* **72**, 132–41 (1956).

477. Rücker, N. *et al.* Acetate dissimilation and assimilation in *Mycobacterium tuberculosis* depend on carbon availability. *J. Bacteriol.* **197**, 3182–3190 (2015).

478. Matsumoto, Y. *et al.* Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles. *Emerg. Microbes Infect.* **8**, 1043–1053 (2019).

479. Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **16**, 202–213 (2018).

480. Hershberg, R. *et al.* High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* **6**, 2658–2671 (2008).

481. Bañuls, A.-L., Sanou, A., Van Anh, N. T. & Godreuil, S. *Mycobacterium tuberculosis*: ecology and evolution of a human bacterium. *J. Med. Microbiol.* **64**, 1261–1269 (2015).

482. Gagneux, S. *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2869–2873 (2006).

483. Firdessa, R. *et al.* Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg. Infect. Dis.* **19**, 460–463 (2013).

484. Wiens, K. E. *et al.* Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Med.* **16**, 196 (2018).

485. Gagneux, S. & Small, P. M. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**, 328–337 (2007).

486. Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R. & Gordon, S. V. Myths and misconceptions: The origin and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **7**, 537–544 (2009).

487. Couvin, D., Zozio, T. & Rastogi, N. SpolSimilaritySearch – A web tool to compare and search similarities between spoligotypes of *Mycobacterium tuberculosis* complex. *Tuberculosis* **105**, 49–52 (2017).

488. Thain, N. *et al.* Towards better prediction of *Mycobacterium tuberculosis* lineages from MIRU-VNTR data. *Infect. Genet. Evol.* **72**, 59–66 (2019).

489. Lipworth, S. *et al.* SNP-IT Tool for Identifying Subspecies and Associated Lineages of *Mycobacterium tuberculosis* Complex. *Emerg. Infect. Dis.* **25**, 482 (2019).

490. Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**, 907 (1997).

491. Vitol, I., Driscoll, J., Kreiswirth, B., Kurepina, N. & Bennett, K. P. Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. *Infect. Genet. Evol.* **6**, 491–504 (2006).

492. Xia, E., Teo, Y. Y. & Ong, R. T. H. SpoTyping: Fast and accurate *in silico Mycobacterium* spoligotyping from sequence reads. *Genome Med.* **8**, 19 (2016).

493. Genestet, C. *et al.* Consistency of *Mycobacterium tuberculosis* Complex Spoligotyping between the Membrane-Based Method and *In Silico* Approach. *Microbiol. Spectr.* **10**, e00223-22 (2022).

494. Filliol, I. *et al.* Global Distribution of *Mycobacterium tuberculosis* Spoligotypes. *Emerg. Infect. Dis.* **8**, 1347–1349 (2002).

495. Demay, C. *et al.* SITVITWEB - A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect. Genet. Evol.* **12**, 755–766 (2012).

496. Couvin, D., David, A., Zozio, T. & Rastogi, N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infect. Genet. Evol.* **72**, 31–43 (2019).

497. Brudey, K. *et al. Mycobacterium tuberculosis* complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**, 1–17 (2006).

498. Jones, R. *et al.* Resolving a clinical tuberculosis outbreak using palaeogenomic genome reconstruction methodologies. *Tuberculosis* **119**, 101865 (2019).

499. Aminian, M., Shabbeer, A. & Bennett, K. P. A conformal Bayesian network for classification of *Mycobacterium tuberculosis* complex lineages. *BMC Bioinformatics* **11**, 1–14 (2010).

500. Shabbeer, A. *et al.* TB-Lineage: An online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infect. Genet. Evol.* **12**, 789–797 (2012).

501. Napier, G. *et al.* Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med.* **12**, 114 (2020).

502. Brown, T., Nikolayevskyy, V., Velji, P. & Drobniewski, F. Associations between *Mycobacterium tuberculosis* strains and phenotypes. *Emerg. Infect. Dis.* **16**, 272–80 (2010).

503. Kato-Maeda, M. *et al.* Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. *Int. J. Tuberc. Lung Dis.* **15**, 131–3 (2011).

504. Ramos, D. F., Tavares, L., Silva, P. E. A. da & Dellagostin, O. A. Molecular typing of *Mycobacterium bovis* isolates: a review. *Brazilian J. Microbiol.* **45**, 365–372 (2014).

505. Reyes, J. F., Chan, C. H. S. & Tanaka, M. M. Impact of homoplasy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* **12**, 811–818 (2012).

506. Comas, I., Homolka, S., Niemann, S. & Gagneux, S. Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in *Mycobacterium tuberculosis* Highlights the Limitations of Current Methodologies. *PLoS One* **4**, e7815 (2009).

507. Rindi, L., Lari, N. & Garzelli, C. Large Sequence Polymorphisms of the Euro-American lineage of *Mycobacterium tuberculosis*: A phylogenetic reconstruction and evidence for convergent evolution in the DR locus. *Infect. Genet. Evol.* **12**, 1551–1557 (2012).

508. Benavente, E. D. *et al.* PhyTB: Phylogenetic tree visualisation and sample positioning for *M. tuberculosis*. *BMC Bioinformatics* **16**, 1–5 (2015).

509. Homolka, S. *et al.* High Resolution Discrimination of Clinical *Mycobacterium tuberculosis* Complex Strains Based on Single Nucleotide Polymorphisms. *PLoS One* **7**, e39855 (2012).

510. Bornhorst, J. A. & Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* **326**, 245–54 (2000).

511. Cytiva. *Affinity Chromatography, Vol. 2: Tagged Proteins*. (2021).

512. Robichon, C., Luo, J., Causey, T. B., Benner, J. S. & Samuelson, J. C. Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* **77**, 4634–46 (2011).

513. Bolanos-Garcia, V. M. & Davies, O. R. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 1304–1313 (2006).

514. Cytiva. *Strategies for protein purification*. (2021).

515. Cytiva. *Ion exchange chromatography*. (2021).

516. Andersen, K. R., Leksa, N. C. & Schwartz, T. U. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**, 1857 (2013).

517. Lui, N. S. T. & Roels, O. A. An improved method for determining glyoxylic acid. *Anal. Biochem.* **38**, 202–209 (1970).

518. Giachetti, E., Pinzauti, G. & Vanni, P. A new continuous optical assay for isocitrate lyase. *Exp. 1984 402* **40**, 227–228 (1984).

519. Bisswanger, H. Enzyme assays. *Perspect. Sci.* **1**, 41–55 (2014).

520. Van Ingen, J., Turenne, C. Y., Tortoli, E., Wallace, R. J. & Brown-Elliott, B. A. A definition of the *Mycobacterium avium* complex for taxonomical and clinical purposes, a review. *Int. J. Syst. Evol. Microbiol.* **68**, 3666–3677 (2018).

521. Toney, N., Adekambi, T., Toney, S., Yakrus, M. & Butler, W. R. Revival and emended description of '*Mycobacterium paraffinicum*' Davis, Chase and Raymond 1956 as *Mycobacterium paraffinicum* sp. nov., nom. rev. *Int. J. Syst. Evol. Microbiol.* **60**, 2307–2313 (2010).

522. Barretto, A. R., Monteiro, J. T. C., Lopes, M. L. & da Costa, A. R. F. A case of pulmonary infection due to *Mycobacterium paraffinicum* from the Amazon Region. *IDCases* **10**, 1–3 (2017).

523. Das, S. *et al.* Characterization of Three *Mycobacterium* spp. with Potential Use in Bioremediation by Genome Sequencing and Comparative Genomics. *Genome Biol. Evol.* **7**, 1871–1886 (2015).

524. Le, N. B. & Coleman, N. V. Biodegradation of vinyl chloride, *cis*-dichloroethene and 1,2-dichloroethane in the alkene/alkane-oxidising *Mycobacterium* strain NBB4. *Biodegradation* **22**, 1095–1108 (2011).

525. Simmon, K. E., Low, Y. Y., Brown-Elliott, B. A., Wallace, R. J. & Petti, C. A. Phylogenetic analysis of *Mycobacterium aurum* and *Mycobacterium neoaurum* with redescription of *M. aurum* culture collection strains. *Int. J. Syst. Evol. Microbiol.* **59**, 1371–1375 (2009).

526. Hartmans, S. & De Bont, J. A. M. Aerobic vinyl chloride metabolism in *Mycobacterium aurum* L1. *Appl. Environ. Microbiol.* **58**, 1220–1226 (1992).

527. Zhang, J. & Yang, J. R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015).

528. Aguilar-Rodríguez, J. & Wagner, A. Metabolic determinants of enzyme evolution in a genome-scale bacterial metabolic network. *Genome Biol. Evol.* **10**, 3076–3088 (2018).

529. Vitkup, D., Kharchenko, P. & Wagner, A. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol.* **7**, R39 (2006).

530. Barve, A., Rodrigues, J. F. M. & Wagner, A. Superessential reactions in metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1121 (2012).

531. Wang, F. S., Whittam, T. S. & Selander, R. K. Evolutionary genetics of the isocitrate dehydrogenase gene (*icd*) in *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* **179**, 6551–6559 (1997).

532. Minato, Y. *et al.* Genomewide Assessment of *Mycobacterium tuberculosis* Conditionally Essential Metabolic Pathways. *mSystems* **4**, e00070-19 (2019).

533. DeJesus, M. A. *et al.* Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio* **8**, e02133-16 (2017).

534. Eisenthal, R., Danson, M. J. & Hough, D. W. Catalytic efficiency and kcat/KM: a useful comparator? *Trends Biotechnol.* **25**, 247–249 (2007).

535. Carrillo, N., Ceccarelli, E. A. & Roveri, O. A. Usefulness of Kinetic Enzyme Parameters in Biotechnological Practice. *Biotechnol. Genet. Eng. Rev.* **27**, 367–382 (2010).

536. Subtil, F. T. Discovery of novel mechanisms of antibiotic resistance in mycobacteria. (King's College London, 2022).

537. Kapopoulou, A., Lew, J. M. & Cole, S. T. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis* **91**, 8–13 (2011).

538. Pundir, S., Martin, M. J. & O'Donovan, C. UniProt Protein Knowledgebase. in *Protein Bioinformatics. Methods in Molecular Biology, vol 1558.* 41–55 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-6783-4_2

539. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

540. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

541. Agarwala, R. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).

542. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).

543. Gasteiger, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).

544. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).

545. Goodfellow, M. & Jones, A. L. *Corynebacteriales ord. nov.* in *Bergey's Manual of Systematics of Archaea and Bacteria* 1–14 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118960608.obm00009

546. Tindall, B. J. On the status of the names Corynebacteriales Goodfellow and Jones 2015, Mycobacteriales Janke 1924 (Approved Lists 1980) and Mycobacteriales Cavalier-Smith 2002. *Int. J. Syst. Evol. Microbiol.* **69**, 3310–3312 (2019).

547. Parker, C. T., Tindall, B. J. & Garrity, G. M. International code of nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* **69**, S1 (2019).

548. Gupta, R. S. Commentary: Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Front. Microbiol.* **10**, 206 (2019).

549. Kim, C.-J. *et al.* Differentiating rapid- and slow-growing mycobacteria by difference in time to growth detection in liquid media. *Diagn. Microbiol. Infect. Dis.* **75**, 73–76 (2013).

550. Bojalil, L. F., Cerbon, J. & Trujillo, A. Adansonian classification of mycobacteria. *J. gen. Microbiol* **28**, 333–346 (1962).

551. Tsukamura, M. Adansonian Classification of Mycobacteria. *J. gen. Microbiol* **45**, 253–273 (1966).

552. Tsukamura, M. A Review of the Methods of Identification and Differentiation of Mycobacteria. *Clin. Infect. Dis.* **3**, 841–861 (1981).

553. Saifuddin, M., Bhatnagar, J. M., Segrè, D. & Finzi, A. C. Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat. Commun. 2019 101* **10**, 1–10 (2019).

554. Geyer, K. M., Kyker-Snowman, E., Grandy, A. S. & Frey, S. D. Microbial carbon use efficiency: accounting for population, community, and ecosystem-scale controls over the fate of metabolized organic matter. *Biogeochemistry* **127**, 173–188 (2016).

555. Schmid, F. Biological Macromolecules: UV-visible Spectrophotometry. in *eLS* (Wiley, 2001). doi:10.1038/npg.els.0003142

556. Raynal, B., Lenormand, P., Baron, B., Hoos, S. & England, P. Quality assessment and optimization of purified protein samples: why and how? *Microb. Cell Fact.* **13**, 180 (2014).

557. Rodger, A. & Wormell, P. Absorption Spectroscopy: Practical Aspects. in *Encyclopedia of Biophysics* 33–35 (Springer Berlin Heidelberg, 2013). doi:10.1007/978-3-642-16712-6_779

558. Porterfield, J. Z. & Zlotnick, A. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology* **407**, 281–288 (2010).

559. Mach, H., Middaugh, C. R. & Denslow, N. Determining the Identity and Purity of Recombinant Proteins by UV Absorption Spectroscopy. *Curr. Protoc. Protein Sci.* **1**, 7.2.1-7.2.21 (1995).

560. Motulsky, H. J. GraphPad Prism 9 Curve Fitting Guide - Replicates test. Available at: https://www.graphpad.com/guides/prism/latest/curve-fitting/reg_diagnostics_tab_5_3.htm. (Accessed: 28th April 2023)

561. Nichols, C. M. *et al.* Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **90**, 14484–14492 (2018).

562. Paglia, G., Smith, A. J. & Astarita, G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrom. Rev.* **41**, 722–765 (2022).