Structural and Functional Studies on Acyl-CoA Carboxylases of *Mycobacterium tuberculosis*

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Madhankumar Anandhakrishnan

2013

Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
B.Tech., Madhankumar Anandhakrishnan
born in Madurai, India

Oral examination:

Structural and Functional Studies on Acyl-CoA Carboxylases of *Mycobacterium tuberculosis*

Referees:

Dr. Rob Meijers

Prof. Dr. Robert B. Russell

Table of Contents

Summary

Mycobacterium tuberculosis, one of the deadliest human pathogens, causes several million new infections and about two million fatalities annually. The cell wall of *M. tuberculosis* is endowed with a highly impermeable, complex array of diverse lipids such as mycolic acids, which bestow the bacterium with not only virulence, but also resistance to host immunity and antibiotics. Mycobacterial lipid metabolism has thus emerged as an attractive target for the design and development of novel anti-mycobacterial therapeutics. The first committed step in the biosynthesis of long-chain fatty acids – the carboxylation of acetyl-CoA to malonyl-CoA - is catalysed by the multi-functional multi-subunit acetyl-CoA carboxylase (ACC) enzyme. Some ACC complexes, especially from actinobacteria, are active on diverse substrates and are generally referred to as acyl-CoA carboxylases (YCCs). Typically, YCCs are composed of biotin carboxylase, biotin carboxyl carrier protein (collectively known as alpha) and carboxyltransferase (beta) subunits. Interestingly, the genomes of most mycobacteria code for three alpha subunits (AccA1 - AccA3), six beta subunits (AccD1 - AccD6), and a unique epsilon subunit (AccE5) while most other forms of life possess not more than two YCCs. Despite the significant roles of YCCs in mycobacterial fatty and mycolic acid biosyntheses and hence in cell wall integrity and antibiotic resistance, a comprehensive understanding of their properties and functions is lacking.

This dissertation is focused on the structural and functional characterisation of the essential components (AccA3, AccD4 - AccD6, and AccE5) of *M. tuberculosis* YCCs implicated or known to be involved in fatty acid metabolism. X-ray crystallography and complementary biophysical and biochemical approaches have been employed in an attempt to address questions concerning interactions between YCC components and differences in β-subunit substrate specificity.

Multiple co-expression and co-purification strategies yielded YCC complexes (the propionyl-CoA carboxylase AccA3-AccD5 and the putative long-chain acyl-CoA carboxylase AccA3-AccD4) that were catalytically active but did not assemble into stable forms amenable to structural analyses; possible explanations for these

observations have been discussed in detail. Significant effort was invested in the production of the ε-subunit AccE5, but its association with its interacting partners (AccA3 and AccD5) could not be investigated due to technical impediments stemming, presumably, from the intrinsic disordered nature of the protein. Biophysical studies of AccD4 and AccD6 (the β-subunit of ACC) revealed unexpected structural diversity in the *M. tuberculosis* YCC β-subunit subfamily. Unlike all other actinobacterial homohexameric β-subunits characterised to date, AccD4 and AccD6 were found to function as lower oligomers, highlighting that hexameric assembly is not a requisite for carboxyltransferase function. Endeavours to crystallize AccD4, in apo-form or in complex with substrate/cofactor analogs, were unsuccessful. The high-resolution crystal structure of AccD6, on the other hand, was determined by the method of molecular replacement. The structure of AccD6 has elucidated the molecular basis of homodimeric arrangement, besides throwing light on the conserved and non-conserved features of the active site, and the putative determinants of substrate specificity.

Taken together, the findings of this study have added to the existing knowledge of the *M. tuberculosis* structural proteome and have furthered our understanding of the biophysical attributes and functions of YCC β-subunits, validated anti-mycobacterial drug targets. Interesting insights into the likely molecular evolution of YCC β-subunits have been acquired.

Zusammenfassung

Mycobacterium tuberculosis ist der Tuberkuloseerreger, dem jährlich zwei Millionen Menschen zum Opfer fallen und der jedes Jahr viele Millionen Neuinfektionen aufweist. Die Zellwand von *M. tuberculosis* besteht aus verschiedenen Fettsäuren und weist einen hohen Anteil von Mykolsäuren auf. Die damit einhergehende hochgradige Impermeabilität der Zellwand trägt entscheidend zur Virulenz von *M. tuberculosis* bei, indem sie sowohl Schutz vor der Immunreaktion des Wirtsorganismus vermittelt, als auch die Aufnahme von Antibiotika erschwert. Die Synthese mykobakterieller Lipide ist daher ein attraktives Zielobjekt für die Entwicklung neuer Antibiotika. Die erste Reaktion dieser Synthese ist die Carboxylierung von Acetyl-CoA zu Malonyl-CoA und wird durch die multi-funktionelle multimere Acetyl-CoA Carboxylase (ACC) katalysiert. Einige ACC Komplexe, vor allem die der Aktinobakterien, carboxylieren unterschiedliche Substrate und werden daher Acyl-CoA Carboxylasen (YCC) genannt. Sie bestehen aus einer Biotin Caboxylase und einer Biotin-carboxyl Transportprotein Untereinheit (alpha Untereinheit), sowie aus einer Carboxyltransferase Untereinheit (beta Untereinheit). Interessanterweise verfügen Mycobakterien über Gene, die drei alpha Untereinheiten (AccA1 – AccA3), sechs beta Untereinheiten (AccD1 - AccD6) sowie eine epsilon Untereinheit (AccE5) kodieren, während die meisten Lebensformen häufig nicht mehr als zwei YCCs Gene aufweisen. Trotz ihrer wichtigen Rolle im mykobakteriellen Lipidmetabolismus sind die Grundlagen der Funktionsweise dieser Proteinkomplexe weitgehend unbekannt.

Diese Dissertation handelt über die strukturelle und funktionelle Beschreibung der essentiellen Komponenten (AccA3, AccD4 - AccD6, und AccE5) der *M. tuberculosis* YCCs, die für die Synthese der charakteristischen Mykolsäuren verantwortlich sind. Röntgenkristallografie sowie andere komplementäre biophysische und biochemische Methoden wurden genutzt, um die Interaktion zwischen diesen YCC Komponenten zu untersuchen und deren Substratspezifität zu erklären.

Mehrere Strategien zur Herstellung von YCC Komplexen wurden verfolgt um aktive Enzyme zu produzieren. Der Komplex aus Propionyl-CoA Carboxylase (AccA3-AccD5) sowie der Komplex aus Acyl-CoA Carboxylase (AccA3-AccD4), welche

lange Acylketten carboxyliert, wurden koexpremiert und waren biochemisch aktiv. Beide Proteinkomplexe waren jedoch als Komplex nicht stabil und somit nicht für strukturelle Untersuchungen geeignet. Mögliche Gründe hierfür werden in dieser Arbeit ausgiebig beschrieben und analysiert. Das Epsilonprotein AccE5 konnte mit einer aufwendigen Prozedur hergestellt werden. Dabei erwies es sich als ein intrinsisch ungeordnetes Protein. Die biophysikalische Analyse von AccD4 und AccD6 deutet auf eine ungewöhnliche Struktur dieser beiden Untereinheiten hin, die sich von derjenigen der beta Untereinheiten der meisten anderen Aktinobakterien deutlich unterscheidet. Im Gegensatz zu diesen aktinobakteriellen homologen beta Untereinheiten bilden AccD4 und AccD6 keine homohexamere, sondern eine niederhomooligomere Strukturen. Die Kristallisation und Lösung der Röntgenstruktur von AccD6 zeigte ebenfalls, dass diese Untereinheiten Homodimere bilden. Darüber hinaus erlaubt diese Kristallstruktur eine detaillierte Beschreibung der Aminosäuren, die für Substratbindung und Katalyse notwendig sind.

Diese Dissertation liefert einen wichtigen Beitrag zum molekularen und funktionellen Verständnis der *M. tuberculosis* Mykolsäuresynthesemaschinerie. Da YCC beta Untereinheiten potentielle Zielproteine für antimykobakterielle Wirkstoffe sind, liefert diese Arbeit Grundlagen für die zukünftige Erforschung und Entwicklung dieser Wirkstoffe. Darüber hinaus gewährt diese Arbeit auch interessante Einblicke in die Evolution dieser Proteinfamilie.

Publications

This dissertation work resulted in the following manuscripts; reprints of the published manuscripts are attached at the end of the Appendix section.

1) Noens EE, Williams C, **Anandhakrishnan M**, Poulsen C, Ehebauer MT & Wilmanns M (2011) Improved mycobacterial protein production using a *Mycobacterium smegmatis groEL1ΔC* expression strain. *BMC Biotechnology* 11: 27.
2) Holton SJ, **Anandhakrishnan M**, Geerlof A & Wilmanns M (2013) Structural characterization of a D-isomer specific 2-hydroxyacid dehydrogenase from *Lactobacillus delbrueckii* ssp. *bulgaricus*. *Journal of Structural Biology* 181(2): 179-84.
3) **Anandhakrishnan M**, Ehebauer MT & Wilmanns M (2013) Structural analysis of the carboxyltransferase subunit of *Mycobacterium tuberculosis* acetyl-CoA carboxylase: Insights into the substrate specificity code (In preparation).
4) Ehebauer MT, Noens EE, Zimmermann M, Marrakchi H, Laneelle M-A, **Anandhakrishnan M**, Chesnel D, King-Scott S, Holton SJ, Daffé M, Sauer U & Wilmanns M (2013) Catabolism of branched-chain amino acids in mycobacteria: Identification of novel carboxylase activities in the acyl-CoA carboxylase family (In preparation).

Abbreviations

aa	Amino acid(s)
ACC	Acetyl-coenzyme A carboxylase
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
AG	Arabinogalactan
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BC	Biotin carboxylase
BCCP	Biotin carboxyl carrier protein
BCG	Bacillus Calmette–Guérin
β-ME	Beta-mercaptoethanol
BSA	Bovine serum albumin
Cam	Chloramphenicol
CD	Circular dichroism
CHES	N-cyclohexyl-2-aminoethanesulfonic acid
CoA	Coenzyme A
CT	Carboxyltransferase
ddH_2O	Doubly distilled water
DESY	Deutsches Elektronensynchrotron
DMSO	Dimethyl sulphoxide
dNTPs	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ESI-TOF	Electrospray ionization – time of flight
ESRF	European Synchrotron Radiation Facility
ETH	Ethambutol
FAS	Fatty acid synthase
FID	Free interface diffusion
GFP	Green fluorescent protein

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	Hexahistidine
HPSF	High purity salt free
Hyg	Hygromycin
IL	Interleukin
IMAC	Immobilized metal affinity chromatography
INH	Isoniazid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
LB	Luria broth
LIC	Ligase-independent cloning
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MCS	Multiple cloning site
MDR-TB	Multidrug-resistant tuberculosis
MHC	Major histocompatibility complex
MR	Molecular replacement
MW	Molecular weight
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide (reduced)
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCC	Propionyl-coenzyme A carboxylase
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PG	Peptidoglycan
P_i	Inorganic phosphate
PKS	Polyketide synthase
PZA	Pyrazinamide
RIF	Rifampicin

RMSD	Root mean square deviation
rpm	Rotations per minute
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SLS	Static light scattering
SOB	Super optimal broth
SOC	Super optimal catabolite repression medium
Spec	Spectinomycin
SUMO	Small ubiquitin-like modifier
TB[§]	Transformation buffer
TB[#]	Terrific broth
TB	Tuberculosis
TBE	Tris-borate ethylenediamine-tetraacetic acid
TCEP	Tris-(2-carboxyethyl)-phosphine
TEMED	Tetramethylethylenediamine
TEV	*Tobacco etch virus*
TG	Thioglycerol
TLS	Translation, libration, screw
(v/v)	Volume/volume
WHO	World Health Organisation
(w/v)	Weight/volume
XDR-TB	Extremely drug-resistant tuberculosis
YCC	Acyl-coenzyme A carboxylase

1 Introduction

1.1 *Mycobacterium tuberculosis*

Mycobacterium tuberculosis, an aerobic actinobacterium, is the causative pathogen of tuberculosis (TB), one of the main fatal infections in human beings. With an annual global incidence rate of over nine million (Global Tuberculosis Report 2012, World Health Organisation), the rapid emergence of multidrug-resistant (MDR) strains and the ominous crosstalk with human immunodeficiency virus (HIV), the disease is far from being completely eradicated. Diagnostic and treatment delay, inefficient follow-up methods of tuberculosis management and the decline of socioeconomic standards in developing countries have been aggravating the problem by large. The disease has been declared as a global public health emergency as it is believed that one-third of the world's population is infected and one-tenth of these have a lifetime risk of developing disease symptoms (Global Tuberculosis Report 2012, WHO). Epidemiological data underline the urgent requirement of new drugs and a novel vaccine to control this deadly disease.

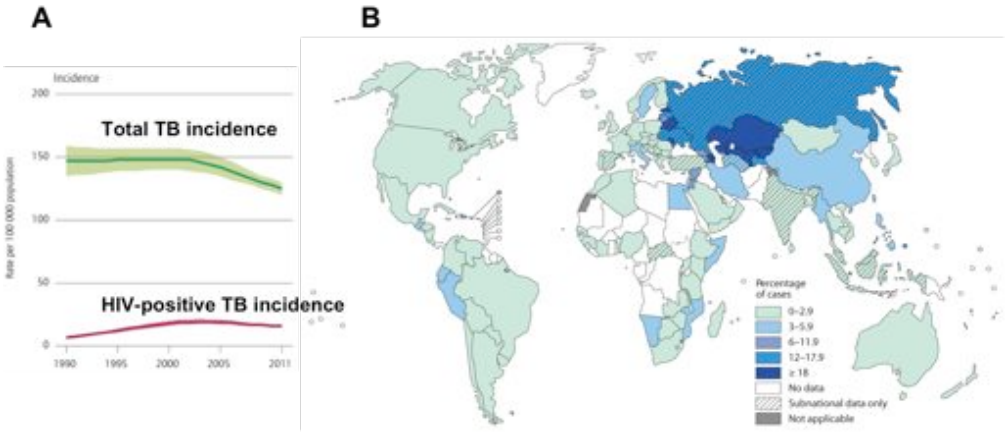

Figure 1-1. TB: two sides of the coin. (A) Measures to control drug-susceptible strains have reduced the total global incidence of TB, although the rate of co-infection with HIV is alarming. **(B)** World map depicting the percentage of new MDR-TB cases in 2011. The failure of health-care systems in some developing countries has fostered the global prevalence of drug-resistant strains. (Adapted, with permission, from the Global Tuberculosis Report 2012, WHO).

1.1.1 Transmission and pathogenesis

The events associated with *M. tuberculosis* infection have been elucidated through studies based on animal models and human TB (Flynn & Chan, 2005, Russell, 2007). All *M. tuberculosis* infections occur by airborne transmission of droplet nuclei containing one or a few viable bacilli produced by a sputum-positive infected individual. Upon inhalation, the droplet nuclei are deposited in the air sacs of the lungs, where they are engulfed by alveolar macrophages, key cells of the innate immune system that act as the primary barrier against infection. Phagocytosis of the bacteria by macrophages is initiated by bacterial contact with macrophage mannose and/or complement receptors (Schlesinger, 1993). Surfactant protein A, a glycoprotein found on alveolar surfaces enhances the binding and uptake of *M. tuberculosis* by upregulating mannose receptor activity (Gaynor*, et al.*, 1995). Cholesterol in cell plasma membranes is thought to be important for the uptake process, as removal of this steroid from human neutrophils hinders the phagocytosis of *M. kansasii* (Peyron*, et al.*, 2000) and similar depletion experiments prevented the entry of *M. bovis* BCG into mouse macrophages (Gatfield & Pieters, 2000). The human toll-like receptor 2 (TLR2) also plays a key role in macrophage activation and hence in *M. tuberculosis* uptake (Noss*, et al.*, 2001).

Upon entry into a host macrophage, *M. tuberculosis* initially resides in an endocytic vacuole called the phagosome. In the event of a normal phagosome-lysosome fusion, these bacteria encounter an adverse environment created by acidic pH, toxic peptides, reactive oxygen intermediates (ROIs) and lysosomal enzymes. Reactive nitrogen intermediates (RNIs) have been known to act strongly against virulent mycobacteria in mouse macrophages (Chan*, et al.*, 1992, Chan*, et al.*, 1995) and alveolar macrophages of a majority of TB-infected patients exhibit inducible nitric oxide synthase (iNOS) activity (Nicholson*, et al.*, 1996). Interestingly, the resistance to RNIs among various strains of *M. tuberculosis* correlates with virulence (Chan*, et al.*, 1992, O'Brien*, et al.*, 1994, Chan*, et al.*, 1995). Given that most macrophage-killing of bacteria occurs in the phagolysosome (Fenton & Vermeulen, 1996), intracellular pathogens have evolved several ways to escape this unfavourable vacuolar microenvironment. It has been demonstrated that *M. tuberculosis* has the ability to prevent phagosome-lysosome fusion (Armstrong & Hart, 1975, Frehel*, et al.*, 1986).

There has also been evidence that virulent mycobacteria can prevent acidification of the phagolysosome, presumably by modulating the activity of a membrane proton pump (Crowle*, et al.*, 1991).

Following phagocytosis, the alveolar macrophages invade the surrounding epithelial layer, giving rise to a localized inflammatory response that results in the chemokine-mediated recruitment of monocytes and neutrophils from adjacent blood vessels (van Crevel*, et al.*, 2002). None of these cell types kill bacteria very efficiently (Fenton & Vermeulen, 1996), but their recruitment provides fresh host cells for the expanding bacterial population. The amorphous bundle of macrophages, monocytes and neutrophils constitutes the granuloma, the hallmark pathologic characteristic of TB. Granuloma formation is an effective means of containing the spread of bacteria. The macrophages in the granulomas undergo differentiation to form multinucleate giant cells, epithelioid cells and foam cells filled with lipid droplets. The development of an acquired immune response triggers the granuloma to adopt a more organized structure: layers of fibrous extracellular matrix material embedded with lymphocytes, surrounding a macrophage-rich centre (Dannenberg, 1994).

M. tuberculosis bacilli are postulated to be unable to multiply within this caseous tissue due to its acidic pH, the low availability of oxygen and the presence of toxic fatty acids; however, some organisms may remain dormant (but alive) for decades within the caseous granuloma (Via*, et al.*, 2008). This enclosed, balanced state is referred to as latent TB and can persist throughout a person's life in an asymptomatic and non-transmissible form. The strength of the host cellular immune response determines if an infection is arrested at this stage or progresses further. In cases of weak or no immune response, disease progression is characterised by the loss of vascularisation, increased necrosis and the accumulation of caseous tissue in the centre of the granuloma (Russell*, et al.*, 2009). Eventually, rupture of the granuloma occurs, releasing thousands of infectious cells into the airways, which can spread as aerosol through a productive cough (Kaplan*, et al.*, 2003).

1.1.2 The fight against TB

Tuberculosis is a disease of antiquity and contrary to the popular belief that it emerged in the agricultural era due to transmission by cows, it appears that the disease has evolved in parallel with human life for thousands of centuries (Smith, *et al.*, 2009). Though the history of anti-tuberculosis chemotherapy dates back to as recent as the 1940s, the competitive interaction between the human immune system and the disease has apparently been a much longer battle.

1.1.2.1 Host defence mechanisms

Understanding host defence mechanisms are of paramount importance in the design and development of strategies to treat and combat disease. The phenomenon of host defence involves successful pathogen recognition, innate and acquired immune mechanisms.

M. tuberculosis activates two germ-line encoded pattern recognition receptors: the Toll-like receptors (TLR) and the nucleotide oligomerization domain (NOD)-like receptors. The TLR1-TLR2 heterodimer and NOD2 specifically recognise a triacylated lipoprotein derivative and a peptidoglycan (muramyl dipeptide) of *M. tuberculosis* respectively (Thoma-Uszynski, *et al.*, 2001, Girardin, *et al.*, 2003), resulting in an NF-κB-mediated inflammatory response and antimicrobial activity (Liu, *et al.*, 2006, Delbridge & O'Riordan, 2007).

Autophagy has been found to be an important innate immune defence strategy (Gutierrez, *et al.*, 2004) and can be induced in macrophages *via* several methods like the activation of the cytokine interferon-gamma (IFN-γ) and TLRs (Delgado, *et al.*, 2008); ubiquitin-derived peptides have direct antimicrobial activity against *M. tuberculosis* and are transported to phagosomes harboring mycobacteria in an autophagy-dependent manner (Alonso, *et al.*, 2007). Serum vitamin D levels also have an influence on TB disease progression and susceptibility (Rook, *et al.*, 1986, Crowle, *et al.*, 1987); under conditions where vitamin D prohormone (25D) is present at sufficient levels, TLR2/1 activation of monocytes leads to a 25-hydroxyvitamin D3-1a-hydroxylase (CYP27b1) and vitamin D receptor (VDR) dependent expression

of the antimicrobial peptide, cathelicidin, and direct microbicidal activity against intracellular *M. tuberculosis* (Liu*, et al.*, 2007, Martineau*, et al.*, 2007).

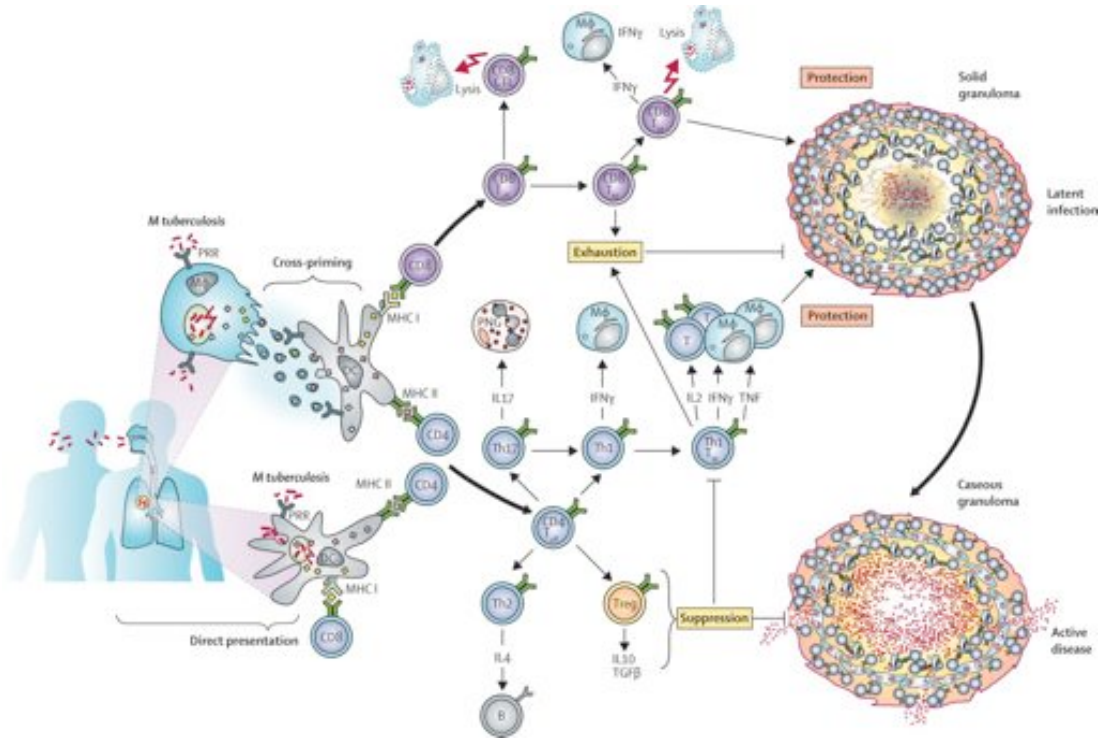

Figure 1-2. The immune response in tuberculosis. Mycobacterial peptides are presented, either directly or by cross-priming, to $CD8^+$ and $CD4^+$ T cells *via* MHC I and MHC II molecules respectively. The stimulated T cells produce a variety of cytokines for activation of T cells, macrophages and polymorphonuclear granulocytes. Some T cells have direct cytolytic function against bacilli whereas some other T cells (T_{reg}) counter-regulate protection mediated by other T cells. Memory T cells that mount a stronger and faster immune response succeed the effector T cells. During active containment in solid granuloma, the bacilli are immune to attack and live in a dormant stage. The suppression of T_h1 cells by cytokines from T_{reg} cells allows resuscitation of bacilli, which leads to granuloma caseation and active disease. (Figure reprinted from (Kaufmann*, et al.*, 2010), with kind permission from Elsevier).

Acquired immunity against *M. tuberculosis* infection is mainly mediated by T lymphocytes and is associated with the emergence of delayed-type hypersensitivity reactions (Dannenberg, 1994). These responses typically occur 4 to 6 weeks post infection. The activation of T lymphocytes by specific mycobacterial antigens leads to a series of immunological events that ultimately limit the replication of *M. tuberculosis*: T cells of the helper/inducer phenotype ($CD4^+$) upregulate populations of antigen-specific effector T cells ($CD4^+$) and cytotoxic T cells ($CD8^+$). The effector T cells secrete various cytokines which activate macrophages and present them with

enhanced mycobactericidal and mycobacteriostatic abilities. The $CD4^+$ cells have been dichotomized into functional subsets based on the respective cytokine profiles (Mosmann*, et al.*, 1986); the T_h1 helper cells produce IFN-γ and interleukin-2 (IL-2), promoting the inflammatory response and cell-mediated immunity; the T_h2 helper cells secrete the cytokines IL-4, IL-5 and IL-10 and drive the immune response towards antibody production (Mosmann*, et al.*, 1986, Mosmann & Coffman, 1989). The cytotoxic T cells, on the other hand, attack and lyse infected macrophages displaying mycobacterial antigens, thereby releasing the engulfed bacilli from their protective niche and exposing them to the killing action of activated macrophages (Kaufmann & Young, 1992).

1.1.2.2 Anti-tuberculosis chemotherapy

The human immune system and *M. tuberculosis* have co-evolved. The capability of the organism to evade the host immune processes is one of the important reasons for the inability of the host to achieve sterile immunity against *M. tuberculosis*, requiring the use of antibiotics to fight the disease.

1.1.2.2.1 First-line drugs

TB can be treated using the first-line drugs mentioned in Table 1-1.

Table 1-1. First-line agents against tuberculosis

Antibiotic	Target of inhibition	Reference
Isoniazid (INH)	Mycolic acid synthesis (InhA, KasA, DfrA)	(Zhang*, et al.*, 1992)
Rifampicin (RIF)	Transcription (RNA polymerase β-subunit)	(Russell & Chopra, 1996)
Ethambutol (ETH)	Arabinogalactan synthesis	(Telenti*, et al.*, 1997)
Pyrazinamide (PZA)	Unknown (FAS-I?)	(Scorpio & Zhang, 1996)

Source: Report No. TDR/PRD/TB/03.1W (WHO, Geneva, 2003)

The first-line therapy is often ineffective for several reasons: lengthy treatment period (that results in lack of compliance and hence failure of treatment), toxic side effects, and the emergence of multidrug-resistant (MDR) and extremely drug-resistant (XDR) TB. MDR-TB is defined as tuberculosis that is resistant to at least INH and RIF (Dalton*, et al.*, 2012) and XDR-TB is defined as TB that has developed resistance to at least INH and RIF as well as to any member of the quinolone family and at least one of the second-line anti-TB injectable drugs: amikacin, kanamycin or capreomycin (WHO press release, 2006). Streptomycin that exhibits anti-tuberculosis activity by means of inhibiting protein synthesis (Blanchard, 1996) is no longer considered a first-line agent because of high rates of resistance.

1.1.2.2.2 Second-line drugs

Anti-tuberculosis agents that are less effective than first-line drugs and/or have toxic side effects and/or are unavailable in many developing countries are classified as second-line drugs. These drugs are administered under special circumstances like resistance to first-line therapy and diagnosis of XDR- and MDR-TB.

Table 1-2. Second-line agents against tuberculosis

Antibiotic class (example)	Target of inhibition	Reference
Aminoglycosides (kanamycin, amikacin)	Protein synthesis	(Doluisio & Swintosky, 1965, Adamis*, et al.*, 2004)
Fluoroquinolones (ciprofloxacin, levofloxacin, moxifloxacin)	DNA gyrase	(Wright*, et al.*, 2000)
Polypeptides (capreomycin, viomycin, enviomycin)	Protein synthesis	(Blumberg*, et al.*, 2003)
Thioamides (ethionamide, prothionamide)	Mycolic acid synthesis	(McIlleron*, et al.*, 2006)
Cycloserine (closerin)	Peptidoglycan synthesis	(Wishart*, et al.*, 2006)
Para-aminosalicylic acid	Folate metabolism, possibly dihydropteroate synthase	(Lehmann, 1946)

Table 1-3. New TB drugs in clinical trials

Compound (other names)	**Chemical class**	**Key features**	**Reference**
Moxifloxacin, gatifloxacin	Fluoroquinolones	Safe (used in other infections); shorter course of therapy	(Cynamon, *et al.*, 2007)
Rifapentine	Rifampicin analog	Longer serum half-life than rifampicin	(Burman, *et al.*, 2001)
Rifalazil (KRM-1648)	Benzoxazinorifamycin	Can be co-administered with drugs that are sensitive to oxidative metabolism	(Fujii, *et al.*, 1995)
SQ109	Ethambutol analog	Superior to ethambutol in *in vitro* and *in vivo* activity	(Jia, *et al.*, 2005)
Bedaquiline (Sirturo, TMC207)	Diarylquinoline	Impressive in vitro and in vivo efficacy against drug-sensitive and drug-resistant strains; targets ATP synthase; the only FDA-approved TB drug in 4 decades	(Andries, *et al.*, 2005)
Sudoterb	Pyrrole	No cross-resistance with existing therapies	(Casenghi, *et al.*, 2007)
Metronidazole	Nitroimidazole	Kills only dormant *M. tuberculosis*	(Wayne & Sramek, 1994)
PA-824	Nitroimidazopyran	No cross-resistance with existing drugs; effective against non-growing cells under hypoxia	(Stover, *et al.*, 2000)
OPC-67683	Nitroimidazole	Inhibits mycolic acid biosynthesis; superior to INH, ETH, RIF and PA-824	(Matsumoto, *et al.*, 2006)

1.1.2.2.3 Promising anti-tuberculosis agents

It is a matter of urgency to develop novel anti-tuberculosis agents mainly because XDR-TB is resistant to first- and second-line drugs. Hundreds of research studies based on genetic approaches to target identification, high-throughput chemical screening, structural biology and virtual screening have identified a few more potential next-generation antituberculars that are currently under clinical testing. Key strong features of some of these compounds have been listed in Table 1-3.

1.1.2.2.4 Emerging drug targets

The availability of the complete genome sequence of *M. tuberculosis* (Cole*, et al.*, 1998) triggered and strengthened the search for new drug targets. Amino acid biosynthesis, especially the shikimate pathway involved in the synthesis of aromatic amino acids in lower life forms, but absent in mammals, is an important target for anti-TB drugs (Ducati*, et al.*, 2007). The PhoPR two-component signal transduction system is a key drug target as it has a major role in the secretion of the immunodominant protein antigens ESAT-6 and CFP-10 and is hence vital for mycobacterial virulence (Ryndak*, et al.*, 2008). The components of siderophore biosynthesis are indispensable for the survival of *M. tuberculosis* in iron-deficit conditions of the host and are therefore valuable drug targets (Ferreras*, et al.*, 2005, Monfeli & Beeson, 2007). Inhibiting the stringent response enzyme can shorten the duration of therapy, as the enzyme is required for the long-time survival of *M. tuberculosis* within the host (Primm*, et al.*, 2000). Another interesting group of targets is constituted by enzymes like isocitrate lyase involved in lipid metabolism as the repertoire of lipophilic molecules is unusually larger in *M. tuberculosis* than in other forms of life (McKinney*, et al.*, 2000, Khasnobis*, et al.*, 2002). A subset of this group comprises of the enzymes responsible for the biosynthesis of the cell wall of *M. tuberculosis,* a unique biological entity whose integrity is of paramount importance for the survival of the bacteria in constrained conditions like within human macrophages. The mode of action of the first-line drugs INH and ETH is indeed the inhibition of cell wall synthesis (Brennan & Crick, 2007).

1.1.3 The cell wall of *Mycobacterium tuberculosis*

Mycobacteria are classified as Gram-positive bacteria, but possess features of both Gram-positive and Gram-negative organisms because of the unique cell wall architecture. Pioneering studies on the chemical nature of the cell wall of *M. tuberculosis* revealed the presence of trehalose-containing compounds, phthiocerols, mycolic acids and other unusual lipids in the mycobacterial cell wall (Mudd*, et al.*, 1941, Goren, 1972, Minnikin*, et al.*, 1982).

1.1.3.1 Biochemical nature of the cell wall

The cell wall of mycobacteria is a unique multilayered structure composed of peptidoglycan (PG), arabinogalactan (AG), mycolic acids, glycolipids and a polysaccharide capsule placed outside the plasma membrane as represented in Figure 1-3. This impermeable, hydrophobic structure confers natural protection against adverse conditions like unfavourable pH, reactive oxygen species, the action of the immune system and many antibiotics, and thus plays an important role in the survival of the pathogen (Briken*, et al.*, 2004, Hunter*, et al.*, 2006). Up to 60% of the dry weight of the cell wall is composed of lipids; moreover, decoding the genome sequence of *M. tuberculosis* revealed that a large proportion of the genes are devoted to the synthesis, processing or breakdown of these lipids (Cole*, et al.*, 1998). Precisely, 233 out of total 3,845 ORFs have been annotated as encoding enzymes for fatty acid metabolism. (Camus*, et al.*, 2002, Ehebauer & Wilmanns, 2011).

The base of the mycobacterial cell wall consists of a plasma membrane (typical of all prokaryotes) that can be resolved into a thick outer layer and a thin inner layer using electron microscopy (Paul & Beveridge, 1992, Zuber*, et al.*, 2008). The thickness of the outer layer is associated with the presence of carbohydrates and phospholipids, including phosphatidylinositol mannosides (PIMs) that are anchor molecules for polysaccharides like lipoarabinomannan and lipomannan in the cell wall (Sibley*, et al.*, 1988, Lemassu & Daffe, 1994, Guerardel*, et al.*, 2002, Pitarque*, et al.*, 2008). The plasma membrane hosts a diversity of other compounds like menaquinones, carotenoids and various glycosylphosphopolyprenols (Brennan & Nikaido, 1995).

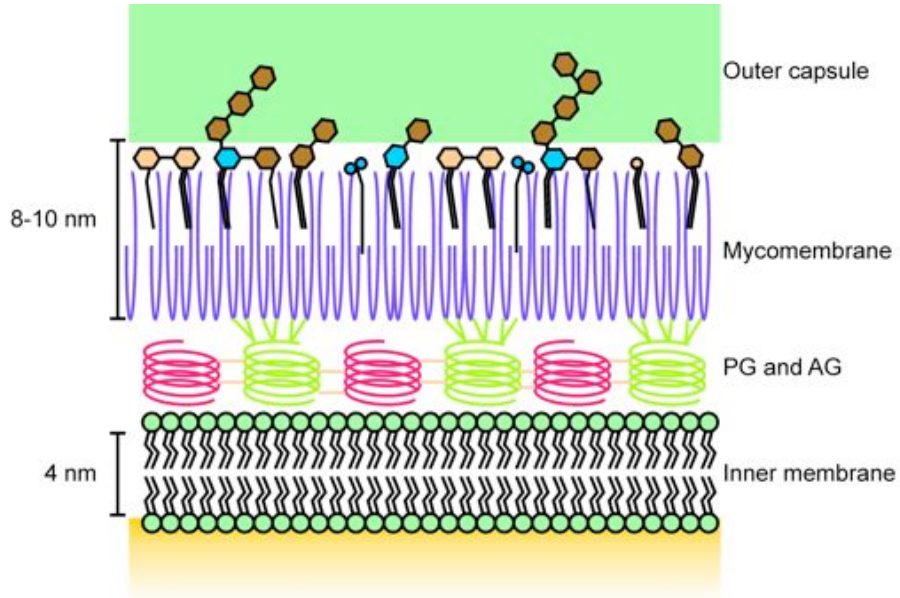

Figure 1-3. Schematic of the mycobacterial cell envelope. The mycobacterial cell wall is a complex biological entity composed of, in addition to the typical prokaryotic inner membrane, an outer capsule and a waxy mycomembrane that is essential for mycobacterial survival.

The peptidoglycan layer is composed of peptide side chains with L-alanyl-D-isoglutaminyl-meso-diaminopimelyl-o-alanine, in which the diaminopimelic acids are amidated (Lederer*, et al.*, 1975). This hydrophobic polymer not only provides the structural framework of the cell wall, but also serves as a permeability barrier against a number of antibiotics. The PG layer is covalently linked to the AG, which is in turn esterified with the mycolic acids that are specialized biomolecules that make up the mycomembrane (Misaki*, et al.*, 1974, McNeil*, et al.*, 1991, Brennan & Nikaido, 1995). Mycolic acids, the hallmark of mycobacteria, are very long-chain α-alkyl, β-hydroxy fatty acids synthesized by the sequential and concerted action of three enzyme systems: fatty acid synthases FAS-I and FAS-II and a polyketide synthase (PKS13) (Glickman*, et al.*, 2000, Vilcheze*, et al.*, 2000, Bhatt*, et al.*, 2005, Bhatt*, et al.*, 2007, Brown*, et al.*, 2007). Mycolic acids are present as free mycolates or as esters of trehalose sugars called trehalose monomycolates (TMM) and trehalose dimycolates (TDM) (Noll & Bloch, 1955, Noll*, et al.*, 1956, Takayama*, et al.*, 2005). These fatty acids not only play a crucial structural role in cell envelope architecture but also are essential for mycobacterial growth (Barry*, et al.*, 1998, Asselineau*, et al.*, 2002).

A number of other surface-exposed lipids like sulfolipids, polyacyltrehaloses, phthiocerol dimycocerosates (PDIM), mannosyl-b-1-phosphomycoketides and diacyl trehaloses intercalate into the cell wall. In addition, the cell wall harbours a number of proteins, including the antigen 85 complex and porins, which constitute the complex cell-wall assembly (Brennan & Nikaido, 1995, Asselineau & Laneelle, 1998, Draper, 1998, Gokhale*, et al.*, 2007).

1.1.3.2 Fatty acid biosynthesis

The complexity of lipid metabolism in mycobacteria is evident from the fact that these organisms harbor both type I and type II FAS systems whilst other actinomycetes (e.g., corynebacteria) possess only a FAS-I evolved for *de novo* fatty acid biosynthesis and most bacteria (including *Streptomyces*) have only FAS-II. The FAS-I system of mycobacteria is dedicated to *de novo* fatty acid biosynthesis whereas the FAS-II is responsible for the elongation of existing long-chain fatty acids to produce the meromycolate part of mycolic acids (Brennan, 2003).

1.1.3.2.1 Fatty acid synthase-I

In mycobacteria, the multifunctional FAS-I polypeptide catalyzes the *de novo* synthesis of long-chain acyl-CoAs ($C_{16:0}$ and $C_{18:0}$) using acetyl-CoA and malonyl-CoA as the starter and extender units respectively (Smith*, et al.*, 2003). These acyl-CoA molecules can either be used directly for the synthesis of membrane phospholipids or further elongated by FAS-II to produce $C_{24:0}$ as in the case of *M. smegmatis* (Peterson & Bloch, 1977) or $C_{26:0}$ as in the cases of *M. bovis* and *M. tuberculosis* (Kikuchi*, et al.*, 1992). In *M. tuberculosis*, the $C_{26:0}$ fatty acid likely gets carboxylated by a dedicated acyl-CoA carboxylase (described in section 1.2.2.5.3) to α-carboxy $C_{26:0}$ fatty acid that could be utilized by PKS13 in the biosynthesis of mycolic acids (Portevin*, et al.*, 2005).

The connector between the FAS-I and FAS-II systems is the condensing enzyme β-ketoacyl-acyl carrier protein (ACP) synthase III known as mtFabH. This enzyme is responsible for the condensation of malonyl-ACP with the acyl-CoA product of FAS-

I to produce 3-ketoacyl-ACP that is further reduced to an acyl-ACP and directed into the FAS-II cycle (Musayev, *et al.*, 2005, Bhatt, *et al.*, 2007).

1.1.3.2.2 Fatty acid synthase-II

Fatty acid elongation by FAS-II requires the substrate malonyl-ACP, the product of a transacylation reaction (between malonyl-CoA and holo-ACP) catalyzed by mtFabD (Kremer, *et al.*, 2001). ACP, as the name suggests, is responsible for shuttling acyl intermediates between enzymes. Acyl-CoA primers are condensed with malonyl-ACP by the action of mtFabH. The resulting acyl-ACP precursor undergoes a cycle of keto-reduction (Marrakchi, *et al.*, 2002), dehydration and enoyl-reduction (Quemard, *et al.*, 1995) catalysed by the enzymes MabA, HadABC and InhA respectively [Figure 1-4]. The acyl-ACP, that is now longer by two carbon units, undergoes a number of iterative cycles of reduction, the only exception being that mtFabH is replaced by the β-ketoacyl-ACP synthases KasA or KasB (Schaeffer, *et al.*, 2001, Kremer, *et al.*, 2002, Slayden & Barry, 2002).

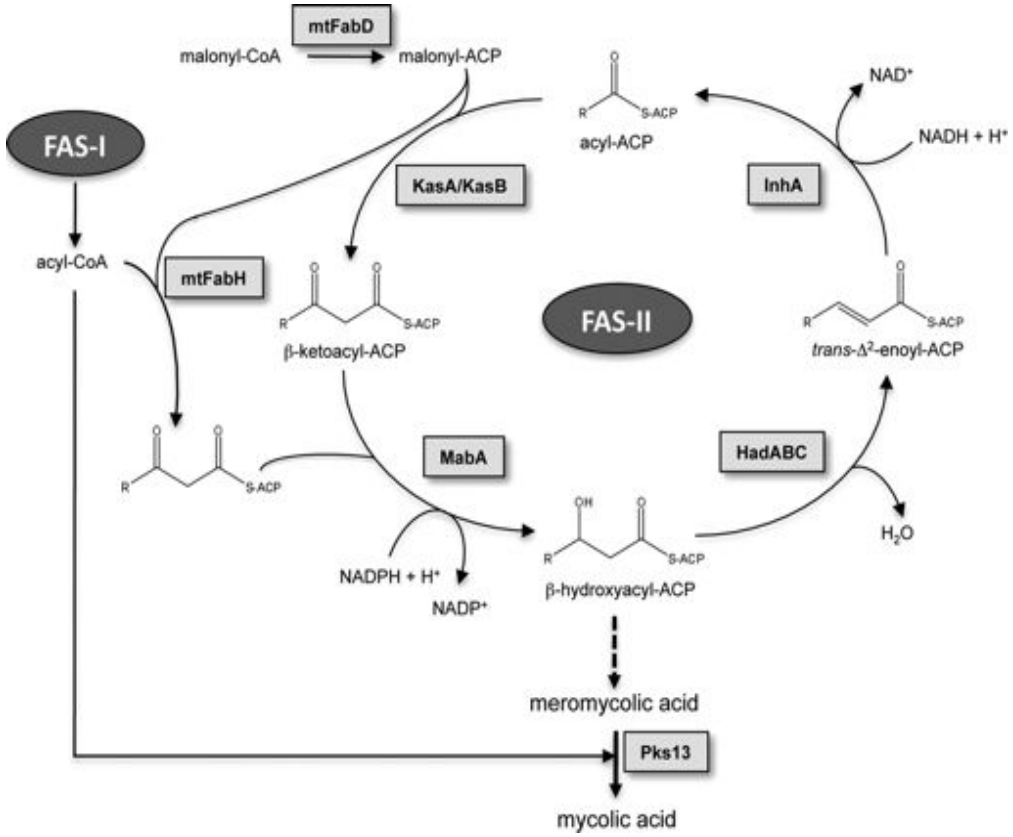

Figure 1-4. Schematic overview of fatty acid biosynthesis in *M. tuberculosis*. Fatty acid biosynthesis in *M. tuberculosis* is carried out by the concerted action of FAS-I and FAS-

II. FAS-I catalyzes *de novo* synthesis of fatty acyl-CoA and FAS-II iteratively produces meromycolates; these moieties are later condensed by PKS13 to form mycolic acids. (Reprinted from (Veyron-Churlet*, et al.*, 2010), with kind permission from the American Society for Biochemistry and Molecular Biology).

The first committed step in long-chain fatty acid biosynthesis is the production of malonyl-CoA, carried out by the enzyme acetyl-CoA carboxylase (ACC) (Wakil*, et al.*, 1983). This enzyme and its close relatives, collectively known as acyl-CoA carboxylases, form the subject matter of this study and will be introduced in more detail in the following section.

1.2 Acyl-CoA carboxylase

1.2.1 Biotin-dependent carboxylases

Acyl-CoA carboxylases (YCC) belong to the large family of biotin-dependent carboxylases: multi-functional enzymes that were discovered more than 50 years ago and are found in all forms of life. Members of this family have diverse functions, including but not limited to metabolism of amino acids, carbohydrates and fatty acids (Wakil*, et al.*, 1983, Cronan & Waldrop, 2002, Tong, 2005, Jitrapakdee*, et al.*, 2008), biosyntheses of mycolic acids, methyl-branched fatty acids (Gago*, et al.*, 2011), and polyketides (Zhang*, et al.*, 2010), autotrophic carbon fixation (Berg*, et al.*, 2010, Pratscher*, et al.*, 2011), urea utilization (Navarathna*, et al.*, 2010, Strope*, et al.*, 2011), methanol assimilation (Smejkalova*, et al.*, 2010), and acetyl-CoA assimilation (Erb*, et al.*, 2007, Alber, 2011, Khomyakova*, et al.*, 2011). These enzymes hence play crucial catabolic and anabolic roles.

1.2.1.1 Functional components

Biotin-dependent carboxylases possess two enzymatic activities and catalyse their reactions in two steps executed by three structural components: biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP) and carboxyltransferase (CT) (Knowles, 1989, Attwood & Wallace, 2002). In many organisms like mycobacteria, BC and BCCP together form the α-subunit while CT, the core catalytic domain, constitutes the β-subunit.

In the first step, BC carries out the ATP-dependent transfer of the carboxyl moiety from bicarbonate to the N1' atom of the biotin cofactor *via* a carboxyphosphate intermediate. *In vivo*, the biotin remains covalently linked (through an amide bond to a conserved Lys side chain) to the BCCP, which facilitates its translocation between active sites. In the second step, the β-subunit catalyses the transfer of the carboxyl group from carboxybiotin to the substrate, which is usually a small compound (such as urea and pyruvate) or a CoA ester of an organic acid. If the substrate is a derivative of a saturated organic acid (such as acetyl-CoA and propionyl-CoA), the α-carbon gets carboxylated. If, on the other hand, the substrate is a derivative of an α-β unsaturated organic acid (such as 3-methylcrotonyl-CoA and geranyl-CoA), the γ-carbon serves as the site of carboxylation.

$$HCO_3^- + \text{Mg-ATP} + \text{enzyme-biotin} \leftrightarrow \text{Enzyme-biotin-}CO_2^- + \text{Mg-ADP} + P_i$$

$$\text{Enzyme-biotin-}CO_2^- + \text{acyl-CoA (substrate)} \leftrightarrow \text{Enzyme-biotin} + \text{carboxyacyl-CoA}$$

To carry out the two steps outlined above, the cofactor biotin has to visit the active sites of both BC and CT. Earlier, it was believed that this requirement could be met by a 'swinging-arm' model, which supports the translocation of biotin by about 30 Å, in turn made possible by the flexible biotin-BCCP connection that comprises of eight methylene groups and ten rotatable single bonds (Perham, 2000). This swinging arm, in its fully extended form, can reach a length of about 16 Å (the distance between the N1' atom of biotin and the Cα-atom of the biotinylated Lysine residue). However, the swinging-arm model has recently been replaced by a 'swinging-domain' model as structures of a few biotin-dependent carboxylases (St Maurice*, et al.*, 2007, Xiang & Tong, 2008, Huang*, et al.*, 2010, Huang*, et al.*, 2011, Fan*, et al.*, 2012) have shown that the distance between the active sites of BC and CT ranges from 55 to 85 Å, thereby rendering the swinging-arm model insufficient for biotin to reach both sites. The swinging-domain model explains the translocation of the entire BCCP domain during catalysis.

1.2.1.2 Classification of biotin-dependent carboxylases

Biotin-dependent carboxylases can be grouped on the basis of component architecture and the chemical nature of the substrate acted upon. The BC, BCCP and CT

components can be present as individual subunits, as in the case of bacteria, or can be fused into a large, multi-domain enzyme as in eukaryotes. Intermediates between these two extreme scenarios also exist in nature.

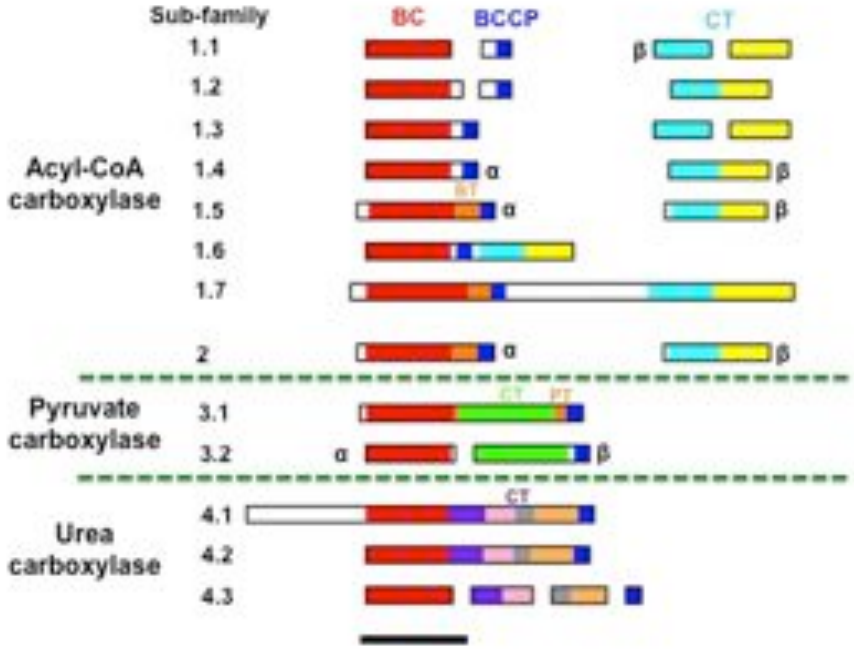

Figure 1-5. Classification of biotin-dependent carboxylases. Biotin-dependent carboxylases can be classified into 4 groups based on substrate nature and 13 sub-families based on domain architecture. Homologous domains are shown with the same colour. The proteins are drawn to size; the black bar at the bottom corresponds to 500 residues. (Adapted from (Tong, 2013), with kind permission from Springer Science and Business Media).

Most of the characterised biotin-dependent carboxylases fall under the subfamily of acyl-CoA carboxylases (YCCs) that utilize CoA-esters of small organic acids as substrates. Some members of this subfamily have distinct substrate preferences while other members can act on a wider group of substrates. There are a few additional members whose substrate preferences are yet to be studied, but have been identified and classified based on sequence information alone (Tong, 2005). The primary sequences of the YCC β-subunits share high levels of amino acid sequence conservation, presumably because all these enzymes bind esters of CoA. Despite high sequence homology, recent data (Huang, *et al.*, 2011) have shown that differences exist in the organization of their components and have led to the understanding that YCCs are of two distinct lineages: one (subfamilies 1.1 through 1.7) that includes acetyl-CoA carboxylase (ACC), propionyl-CoA carboxylase (PCC), etc. and the other (subfamily 2) that consists of methylcrotonyl-CoA carboxylase (MCC) and geranyl-

CoA carboxylase (GCC) [Figure 1-5]. Besides YCCs, two other major subfamilies of biotin-dependent carboxylases are constituted by pyruvate carboxylases (PC) and urea carboxylases (UC), designated as subfamilies 3.x and 4.x respectively.

1.2.2 Actinobacterial acyl-CoA carboxylases

Actinobacteria, a dominant phylum of the bacterial domain, constitute a heterogeneous group of GC-rich Gram positive bacteria that include economically important genera like *Corynebacterium, Mycobacterium, Nocardia, Rhodococcus* and *Streptomyces*. Acetyl-CoA carboxylase complexes purified from actinomycetes were found to have activity on a wider variety of substrates; these enzymes have hence been generally referred to as YCCs (Erfle, 1973, Henrikson & Allen, 1979, Hunaiti & Kolattukudy, 1982). A unique feature of some actinobacterial YCC complexes is the presence of, in addition to the α- and β-subunits, a recently discovered ε-subunit (Diacovich*, et al.*, 2002, Gago*, et al.*, 2006).

1.2.2.1 Physiological roles of actinobacterial YCCs

The availability of huge volumes of sequence data from genome sequencing projects has facilitated the physiological, biochemical, biophysical and structural characterisation of YCCs from several actinobacteria (Diacovich*, et al.*, 2002, Gande*, et al.*, 2004, Gago*, et al.*, 2006, Oh*, et al.*, 2006, Daniel*, et al.*, 2007, Gande*, et al.*, 2007). YCC complexes from actinomycetes not only generate malonyl-CoA for fatty acid biosynthesis, but also catalyze the carboxylation of a diverse range of substrates, such as propionyl-, butyryl- or long-chain acyl-CoAs, thereby producing precursors for the biosynthesis of various complex lipids and polyketides. There seems to exist a direct correlation between the number of putative YCC genes in an actinobacterium and the diversity of its lipids (Gago*, et al.*, 2011). The genome of *Corynebacterium glutamicum* carries genes encoding an α-subunit (*accBC*), four β-subunits (*accD1 – D4*) and an ε-subunit (*accE*), whereas the *Streptomyces coelicolor* genome harbors four genes coding for putative α-subunits (*accA1, accA2, pccA, SCO4381*), four genes coding for putative β-subunits (*accB, pccB, SCO2776, SCO4380*) and two genes encoding ε-subunits (*accE, pccE*) (Bramwell*, et al.*, 1996, Rodriguez & Gramajo, 1999, Rodriguez*, et al.*, 2001, Diacovich*, et al.*, 2004). While most other life forms

possess one or two YCC-related genes, the *M. tuberculosis* genome (Cole*, et al.*, 1998) encodes an unusually high number of putative YCC genes: three α-subunits (*accA1, accA2, accA3*), six β-subunits (*accD1 – D6)*, and one ε-subunit (*accE5*) (Gago*, et al.*, 2006, Oh*, et al.*, 2006). The three α-subunit genes are conserved across mycobacteria, except *M. leprae* wherein only the ortholog of *accA3* is found. This observation, together with transposon mutagenesis experiments in *M. tuberculosis* (Sassetti*, et al.*, 2003, Griffin*, et al.*, 2011), has shown that *accA3* encodes the α-subunit of the essential YCC complexes in *M. tuberculosis*. The different β-subunits presumably act on different fatty acyl-CoA extender units as substrates and hence contribute to the rich diversity of mycobacterial lipids. Existing literature on the biochemical and structural properties of a few YCCs of *S. coelicolor* and *M. tuberculosis* have been reviewed in more detail under sections 1.2.2.4 and 1.2.2.5 respectively.

1.2.2.2 Phylogeny of the β-subunits

Phylogenomic analyses aimed at gaining insights into the early evolution of biotin-dependent carboxylases have suggested that the last common bacterial ancestor had three biotin-dependent carboxylases, one of which probably belonged to the large family of CoA-bearing-substrate-carboxylases, defined based on protein domain composition and phylogenetic analysis (Lombard & Moreira, 2011). These studies have further indicated that the cenancestor (the common ancestor of all organisms) had two different carboxylases capable of the specific carboxylation of pyruvate and the non-specific carboxylation of several CoA-derivatives respectively. These enzymes may have had crucial roles in diverse metabolic pathways like the ancestral versions of anaplerosis, fatty acid biosynthesis, gluconeogenesis, etc. (Lombard & Moreira, 2011). Phylogenetic analysis based only on sequences of the putative β-subunits from several actinomycetes has identified six clearly distinguishable groups (Gago*, et al.*, 2011).

Group I includes the mycobacterial β-subunits AccD1 and AccD2 that constitute YCC complexes which, based on genomic organization, have been implicated in fatty acid degradation (Cole*, et al.*, 1998). The *accA1* and *accD1* genes are organized as an operon in a locus next to *fadD35* and *fadE10* while the *accA2* and *accD2* genes are

found clustered together with *fadE12*, *fadE13* and *echA7*. Therefore, it was assumed that these complexes are involved in fatty acid β-oxidation; however recent lipidome analyses of the corresponding *M. smegmatis* knockouts have negated this assumption by demonstrating no involvement whatsoever in lipid metabolism (Ehebauer, *et al*., unpublished). Furthermore, data from enzymatic assays and metabolomic analyses of the knockouts show that at least one of these complexes possesses methylcrotonyl-CoA carboxylase activity and hence might be a key player in leucine catabolism (Ehebauer, *et al*., unpublished). Group I also includes two *S. coelicolor* β-subunits, one of which is part of a partially characterised PCC complex (Bramwell*, et al.*, 1996). There are no CT orthologs from corynebacteria in group I.

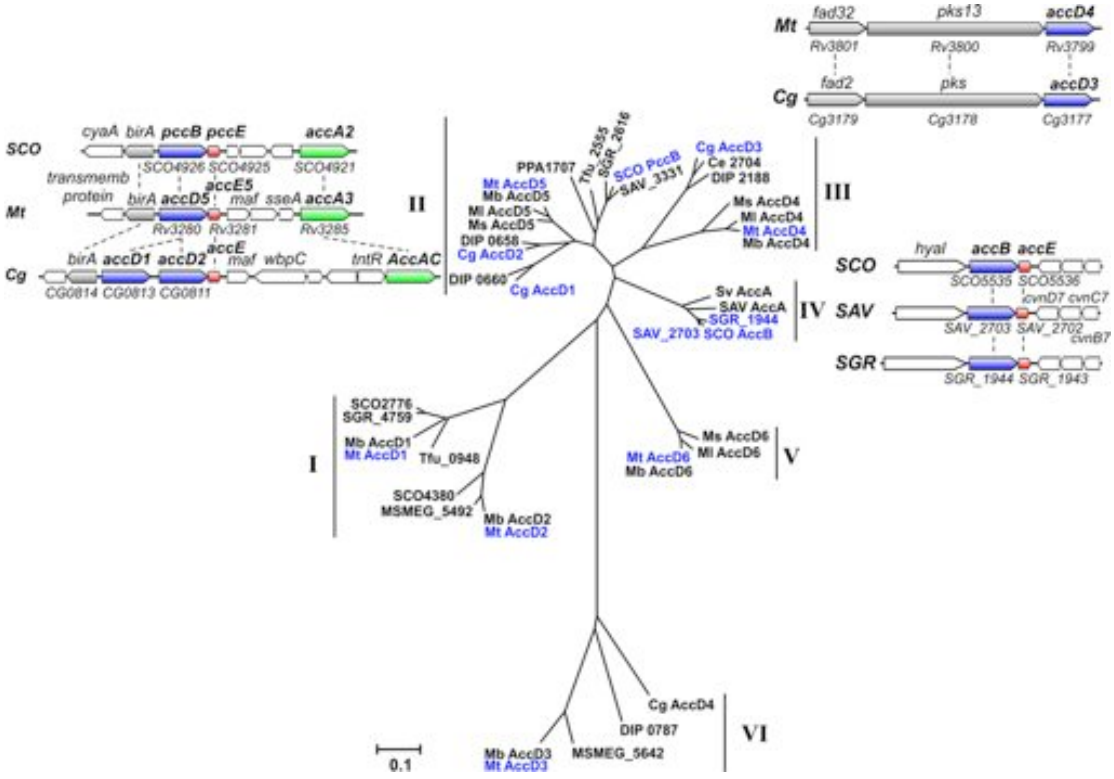

Figure 1-6. Genetic organization and phylogenetic analysis of actinomycete YCC β-subunits. The unrooted phylogenetic tree shows evolutionary relationships between YCC CT subunits of selected actinomycetes. The tree is drawn to scale; the lengths of the branches are indicative of the inferred evolutionary distances. The bar at the bottom indicates the relative number of substitutions per site. The insets depict the synteny and genetic organization of the β-subunits. Cg, *Corynebacterium glutamicum*; Cg, *C. efficiens*; DIP, *C. diptheriae*; Mt, *M. tuberculosis*; Mb, *M. bovis*; Ml, *M. leprae*; Ms, *M. smegmatis*; SCO, *Streptomyces coelicolor*; SAV, *S. avermitilis*; SGR, *S. griseus*; Sv, *S. venezuelae*; PPA, *Propionibacterium acnes*; Tfu, *Thermobifida fusca*. (Reprinted from (Gago*, et al.*, 2011), with kind permission from John Wiley and Sons).

Group II encompasses representatives from all organisms analysed, implying that members of this group might carry out fundamental functions in actinobacteria. Members of this group are the mycobacterial AccD5 (whose encoding gene *accD5* is arranged in a locus with *accA3, accE5* and *birA*, the gene for biotin ligase) [Figure 1-6], the *S. coelicolor* PccB (whose encoding gene *pccB* shares a genetic locus with *accA2, pccE* and *birA*) and the *C. glutamicum* AccD2 (whose encoding gene *accD2* is clustered with *accBC, accE*, *birA* and *accD1* that codes for a β-subunit highly homologous to AccD2, suggesting the occurrence of a gene duplication event. The nomenclature of *C. glutamicum* YCCs is different from that of mycobacterial YCCs; *C. glutamicum* AccD1 and AccD2 are actually orthologous to mycobacterial AccD5 and relatively more distantly related to mycobacterial AccD1 and AccD2.

Group III contains enzymes that are involved in mycolic acid production, suggesting that the corresponding β-subunits have highly specialized functions (Gande*, et al.*, 2004, Portevin*, et al.*, 2005, Gande*, et al.*, 2007). The mycobacterial member of this group is AccD4, the β-subunit of a long-chain YCC that likely provides α-C_{26} carboxylic acid required in the final condensation step of mycolic acid biosynthesis (Trivedi*, et al.*, 2004). The representatives of *C. glutamicum* in this group are AccD3 and AccD2 (orthologous to the mycobacterial AccD4 and AccD5), components of the complex AccBC-AccD2-AccD3-AccE that provides extender units for the biosynthesis of corynomycolic acid. (Gande*, et al.*, 2004, Gande*, et al.*, 2007).

Group IV includes AccB, the β-subunit of an essential YCC (AccA1/AccA2-AccB-AccE) that carboxylates acetyl-CoA in *S. coelicolor* (Rodriguez*, et al.*, 2001, Diacovich*, et al.*, 2002). Although this complex is promiscuous in terms of substrate specificity and has a similar specificity constant for acetyl-, propionyl- and butyryl-CoA *in vitro*, its main physiological role seems to be the generation of malonyl-CoA for the biosynthesis of fatty acids and the polyketide actinorhodin (Rodriguez & Gramajo, 1999, Rodriguez*, et al.*, 2001, Diacovich*, et al.*, 2002). Members of group IV are absent in mycobacteria and corynebacteria.

Group V contains representatives only from mycobacteria. The mycobacterial β-subunit AccD6 serves as the CT component of the essential AccA3-AccD6 complex, whose biological function is to carboxylate acetyl-CoA to malonyl-CoA, the key

extender unit in iterative fatty acid biosynthesis (Daniel, *et al.*, 2007, Kurth, *et al.*, 2009). This complex does not require an ε-subunit for its catalytic activity.

Group VI is only distantly related to groups I-V and includes 'outlier' members like the less-characterised mycobacterial AccD3 and the *C. glutamicum* AccD4 (Gande, *et al.*, 2004). These members seem unlikely to be directly involved in lipid metabolism; their physiological functions remain unclear.

1.2.2.3 Quaternary structure of acyl-CoA carboxylases

All YCC complexes reported so far possess an α-subunit : β-subunit stoichiometry of 1:1. The quaternary structures of YCC complexes of *M. tuberculosis* (Gago, *et al.*, 2006), *M. smegmatis* (Haase, *et al.*, 1982), *S. coelicolor* (Diacovich, *et al.*, 2002) and *Myxococcus xanthus* (Kimura, *et al.*, 1998) have all been deciphered to be $\alpha_6\beta_6$ whereas *Sacharopolyspora erythraea* YCC has a quaternary structure of $\alpha_4\beta_4$ (Hunaiti & Kolattukudy, 1982). The presence of an additional ε-subunit enhances the activity of some of these complexes (Rodriguez, *et al.*, 2001, Diacovich, *et al.*, 2002, Gago, *et al.*, 2006, Oh, *et al.*, 2006); however the exact stoichiometry and mode of interaction of this subunit with the other two subunits is unknown. The β-subunit mostly forms a stable hexamer (Diacovich, *et al.*, 2004, Holton, *et al.*, 2006, Lin, *et al.*, 2006) while the α-subunit exists both as a hexamer and a trimer (Gago, *et al.*, 2006).

1.2.2.4 Acyl-CoA carboxylases of *Streptomyces coelicolor*

The structural basis of actinobacterial YCC β-subunit function and the role of the ε-subunit in YCC activity were first illustrated in *S. coelicolor* (Diacovich, *et al.*, 2002). Genetic and biochemical analyses of actinomycete YCCs were also initiated with work on enzymes from this organism. The genetic organization and the kinetic properties of an ACC and a PCC from *S. coelicolor* have been studied in detail (Rodriguez & Gramajo, 1999, Rodriguez, *et al.*, 2001, Diacovich, *et al.*, 2002). The two complexes have specific β- and ε-subunits (AccB-AccE and PccB-PccE) while sharing the same α-subunit, AccA2. The complexes are clearly different in their substrate specificities and both require the corresponding ε-subunit for maximal

enzyme activity (Diacovich*, et al.*, 2002). The crystal structure of hexameric PccB (Diacovich*, et al.*, 2004) revealed similarity in structural features to related enzymes like the *Propionibacterium shermanii* 12S transcarboxylase (Hall*, et al.*, 2003) and the β-subunit of yeast ACC (Zhang*, et al.*, 2003). Each monomer of PccB has two domains (N- and C-domains) with the conventional crotonase fold (Holden*, et al.*, 2001). The structural analysis, besides highlighting the existence of strong dimeric didomain interactions and the presence of 32-fold symmetry with a threefold axis perpendicular to a twofold axis, pinpointed for the first time a hydrophobic biotin-binding pocket (Diacovich*, et al.*, 2004).

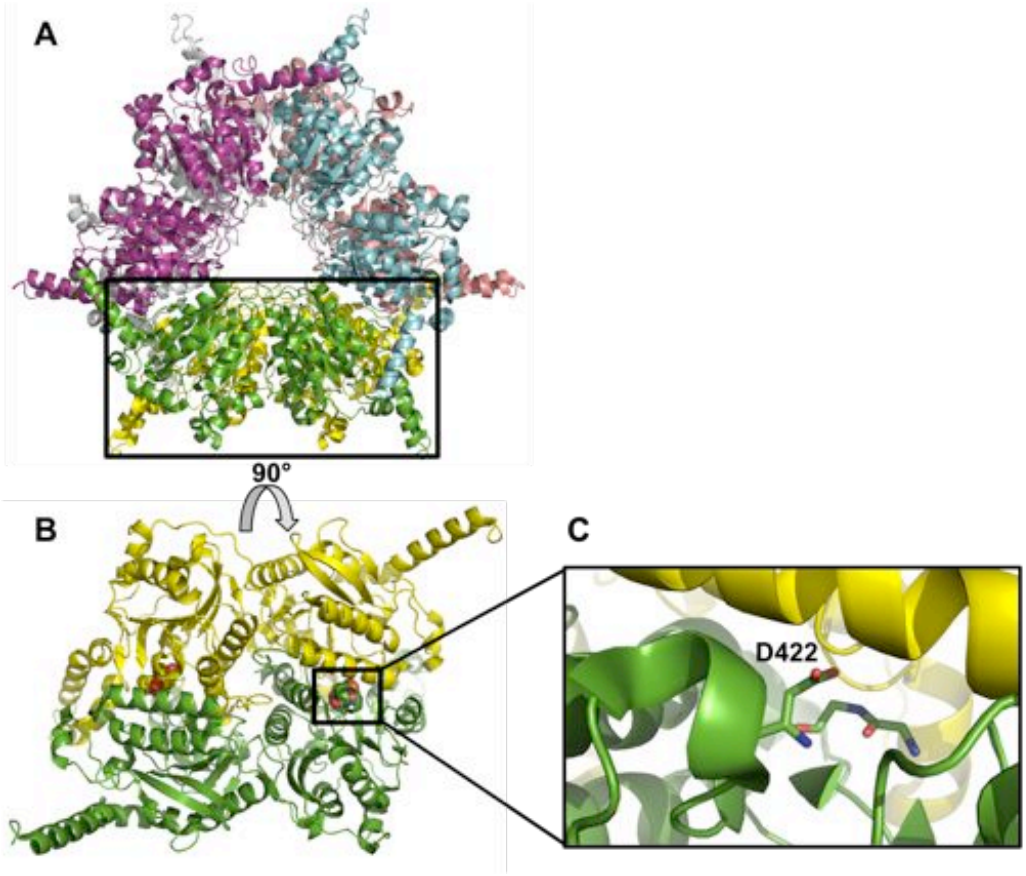

Figure 1-7. Crystal structure of *S. coelicolor* PccB. (A) The overall structure of PccB hexamer, with each monomer in a different colour. **(B)** The dimeric didomain interaction that is crucial for PccB function; the catalytic site (key residues shown as spheres) lies at the interface of two monomers (in green and yellow, respectively). **(C)** Closer view of the substrate-binding pocket with the key substrate specificity determinant residue (D422) and the two conserved glycines (that provide a catalytically important oxyanion hole) shown in 'stick' mode with carbon, oxygen and nitrogen atoms coloured green, red and blue respectively. (PDB ID: 1XNY)

The biotin-propionyl-CoA cocrystal structure of PccB showed that biotin and propionyl-CoA bind perpendicular to each other at the active site, where two key oxyanion holes were identified. Like in previously characterised homologs, the acyl-CoA binds predominantly to the N-domain of the first monomer, while biotin binds predominantly to the C-domain of its dimeric partner. The β-subunit substrate specificity code, i.e. how the β-subunits potentially differentiate one substrate from another, has been explained with the crystal structures of active site mutants of *S. coelicolor* PccB (Diacovich*, et al.*, 2004, Arabolaza*, et al.*, 2010). These mutational analyses have shed light on the key molecular determinant of substrate specificity and have further demonstrated that a single mutation in the substrate-binding pocket is sufficient to interchange the substrate specificities of *S. coelicolor* ACC and PCC (Arabolaza*, et al.*, 2010).

1.2.2.5 Acyl-CoA carboxylases of *Mycobacterium tuberculosis*

The *M. tuberculosis* genome codes for an unusually high number of YCC subunits; some of their properties have been tabulated below (Cole*, et al.*, 1998, Sassetti*, et al.*, 2003, Portevin*, et al.*, 2005, Gago*, et al.*, 2006).

Table 1-4. The YCC subunits of *M. tuberculosis*

Subunit	Protein	Gene	Active expression	Essential?	Substrate
α-subunit	AccA1	*Rv2501c*	Late log phase	No	Biotin
	AccA2	*Rv0973c*	Late log phase	Yes	Biotin
	AccA3	*Rv3285*	Log phase	Yes	Biotin
β-subunit	AccD1	*Rv2502c*	Late log phase	No	Methylcrotonyl-CoA
	AccD2	*Rv0974c*	Late log phase	No	Unknown
	AccD3	*Rv0904c*	Unknown	Unknown	Unknown
	AccD4	*Rv3799c*	Log phase	Yes	C_{26}-CoA
	AccD5	*Rv3280*	Log phase	Yes	C_3-CoA
	AccD6	*Rv2247*	Log phase	Yes	C_2-CoA
ε-subunit	AccE5	*Rv3281*	Log phase	Yes	---

The YCC 'interactome' of *M. tuberculosis* seems to contain fewer YCC complexes than the theoretically possible number of combinations (Ehebauer, *et al.*, unpublished). In other words, each characterised β-subunit interacts with only one of the three α-subunits. Based on genetic organization and data from pull-down experiments (Ehebauer, *et al*., unpublished) in model organisms of *M. tuberculosis*, it is now known that *M. tuberculosis* YCC complexes are likely five in number: AccA1-AccD1, AccA2-AccD2, AccA3-AccD4, AccA3-AccD5-AccE5 and AccA3-AccD6. YCCs of *M. tuberculosis* are of very high biological and pharmaceutical interest mainly because 50% of *M. tuberculosis* YCC subunits (namely AccA3, AccD4, AccD5, AccD6 and AccE5) are validated drug targets (Mdluli & Spigelman, 2006); this study focuses on the characterisation of these subunits.

1.2.2.5.1 *M. tuberculosis* propionyl-CoA carboxylase

The *M. tuberculosis* PCC is an essential complex and is constituted by the subunits AccA3, AccD5 and AccE5 (Gago*, et al.*, 2006). The function of this complex was predicted based on homology to PccB from *S. coelicolor* and was confirmed by kinetic experiments (Gago*, et al.*, 2006). The most likely physiological role of this enzyme is to generate methylmalonyl-CoA for the biosynthesis of branched-chain fatty acids because, *in vitro*, it displays five-fold higher preference to propionyl-CoA than to acetyl-CoA (Gago*, et al.*, 2006). The crystal structure of AccD5 has shed light on the active site architecture of this enzyme (Holton*, et al.*, 2006, Lin*, et al.*, 2006). As in the case of the *S. coelicolor* counterpart, the *M. tuberculosis* PCC exhibits maximal catalytic activity only in the presence of the ε-subunit AccE5 (Gago*, et al.*, 2006). How the three subunits of *M. tuberculosis* PCC interact with each other and how the ε-subunit modulates the activity of the complex remain to be discovered, as there has been no clear molecular evidence for the functional role of the ε-subunit.

1.2.2.5.2 *M. tuberculosis* acetyl-CoA carboxylase

The ACC of *M. tuberculosis* is composed of the subunits AccA3 and AccD6. The physiological role of this complex, the generation of malonyl-CoA for fatty acid and mycolic acid biosynthesis, was inferred based on the location of *accD6* in the *fasII* operon (*fabD-acpM-kasA-kasB-accD6*) and confirmed by experiments on a

conditional mutant of *accD6* in *M. smegmatis* (Kurth, *et al.*, 2009). The ACC complex of *M. tuberculosis* reconstituted *in vitro* has been shown to act on both acetyl-CoA and propionyl-CoA, with a threefold higher efficiency in carboxylation of the former (Daniel, *et al.*, 2007). Unlike in the case of PCC, the ε-subunit AccE5 displayed inhibitory effect on the carboxylase activity of the AccA3-AccD6 complex (Daniel, *et al.*, 2007). The essential nature of AccD6 (Sassetti, *et al.*, 2003, Kurth, *et al.*, 2009) makes it an attractive target for the design of novel anti-TB therapeutics. Based on *in vitro* inhibitory effects of the hit compounds determined *via in silico* inhibitor screening carried out against the AccD5 active site (Lin, *et al.*, 2006), a ligand NCI-172033 capable of potentially inhibiting AccD6 has been identified (Kurth, *et al.*, 2009). However, the stoichiometry of the *M. tuberculosis* AccA3-AccD6 carboxylase and the structural basis of its function remain unclear.

1.2.2.5.3 *M. tuberculosis* long-chain acyl-CoA carboxylase

The relatively poorly characterised long-chain acyl-CoA carboxylase of *M. tuberculosis* is composed of the subunits AccA3 and AccD4. AccD4 is encoded by *accD4*, a genomic neighbour of the genes *pks13 (Rv3800c)* and *fadD32 (Rv3801c)*, whose products are, respectively, a polyketide synthase and an acyl-AMP ligase involved in mycolic acid biosynthesis (Trivedi, *et al.*, 2004). The existence of AccD4 orthologs only in the mycolic acid producing bacteria (Gago, *et al.*, 2011) and the presence of *accD4* in the vicinity of the mycolic acid synthetic machinery suggested that AccD4 is the CT subunit of a long-chain YCC that is responsible for the generation of long-chain acyl carboxylic acid involved in the Claisen condensation step of mycolic acid biosynthesis. Based on the interaction of AccD4 with AccA3 and AccD5 identified by coimmunoprecipitation studies in cell-free extracts of *M. smegmatis*, it was proposed that these three proteins together form an YCC complex (Portevin, *et al.*, 2005). This hypothesis, however, has not been substantiated by experimental data. Detailed biochemical and structural characterisation of AccD4 is requisite to fully comprehend its role in the production of mycolic acids.

1.3 Aims of this study

Given the indispensability of fatty- and complex mycolic-acids for the growth and survival of *M. tuberculosis*, its lipid metabolism is being viewed as a major repertoire of potential drug targets. Acyl-CoA carboxylases, being responsible for the catalysis of the first rate-limiting step of mycolic-acid biosynthesis, are of special therapeutic interest because of two main reasons: (1) these enzymes are encoded by genes known to be essential for mycobacterial survival and (2) these enzymes bear limited homology (in terms of sequence, domain architecture and presumably three-dimensional structure) to their counterparts in human beings.

This study was aimed at the biophysical and X-ray crystallography-based structural characterisation of lipid-metabolizing acyl-CoA carboxylase complexes (AccA3-AccD5-AccE5 / AccA3-AccD4 / AccA3-AccD6) of *M. tuberculosis*. It was expected that such structural and complementing biophysical analyses would help address questions concerning the stoichiometries and modes of interaction between the subunit components, besides providing detailed insights into the molecular basis of carboxylase function.

Another objective of the work was to investigate, by X-ray crystallography, the molecular determinants of β-subunit substrate specificity. The β-subunits AccD5 and AccD6 share moderate levels of sequence identity and are known to possess varying strengths of affinity to substrates that are close chemical relatives; AccD4 is the only YCC β-subunit that has been implicated to act on long-chain acyl-CoA substrates. A part of this study was an attempt to elucidate the molecular reasons for this difference in substrate preferences.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals used in this study were of analytical grade and were purchased from Carl Roth GmbH, Sigma-Aldrich Corporation, Serva Electrophoresis GmbH and Merck Group unless otherwise stated. Microbial growth media components were bought from Melford. Crystallisation screens were purchased from Hampton Research and Jena Bioscience.

2.1.2 Fine biochemicals

Molecular biology grade biochemicals like restriction enzymes, DNA polymerases, PCR materials, DNA ligases, DNA molecular weight markers, etc. for molecular cloning were purchased from New England Biolabs (NEB) and Thermo Fisher Scientific Inc. Oligonucleotides were bought from Eurofins MWG Operon in lyophilized form with HPSF-purified quality, diluted in double distilled water to a final concentration of 10 pmol/µl and frozen until required. Expression vector backbones were provided by the EMBL-Hamburg vector collection [Appendix I]. Protein molecular weight markers were purchased from Carl Roth, NEB and Thermo Fisher Scientific; gel filtration standards were procured from Bio-Rad Laboratories, Inc.; DNaseI and protease inhibitor cocktail were obtained from Serva.

2.1.3 Genomic DNA and bacterial strains

Mycobacterium tuberculosis strain H37Rv and *Mycobacterium smegmatis* strain mc^2155 genomic DNA were kind gifts from the Kaufmann group at the Max Planck Institute for Infection Biology, Berlin. Bacterial strains for cloning and expression were used from the EMBL Hamburg bacterial strain collection [Appendix I]. Owing to its fast-growing nature, relative ease of handling and well-established genetics, the non-pathogenic mycobacterium *M. smegmatis* has often been used as a near-native expression host for the purification of *M. tuberculosis* proteins (Goldstone*, et al.*,

2008). However, the major disadvantage of using this expression system to produce polyhistidine-tagged recombinant proteins is the undesired co-purification of the Hsp60 chaperone GroEL1, which contains an intrinsic histidine-rich C-terminus. The strain of *M. smegmatis* used in this study harbors a mutant version of *groEL1* created in-house (Noens*, et al.*, 2011) by removing the coding sequence for the histidine-rich C-terminus. This novel *M. smegmatis* expression strain *groEL1ΔC* solves the problem of GroEL1 contamination and hence increases the efficiency and speed of protein purification.

2.1.4 Laboratory consumables

Table 2-1. List of laboratory consumables used in the study

Consumable	**Supplier(s)**
Assay plates (96-well)	Greiner Bio-One
Centrifugal ultrafiltration devices	Sartorius, Corning
Chromatography columns (polyprep)	Bio-Rad
Crystallisation plates (24-well)	Nextal, Qiagen, Crystalgen
Dialysis cassettes	Thermo Fisher Scientific
Dialysis membranes from regenerated cellulose	Roth, Spectrum Laboratories
Disposable plastic cuvettes	Brand
Electroporation cuvettes	Biozym Scientific
Filtration membranes from cellulose acetate	GVS Filter Technology
Injection needles	B. Braun
Inoculation loops and needles	Greiner Bio-One
PCR tubes	Brand
Pipette tips	Sarstedt
Plastic Petri dishes	Sarstedt
Plastic syringes	Henke-Sass Wolf, B. Braun
Plastic tubes	Sarstedt, Greiner Bio-One
Serological pipettes	Greiner Bio-One
Siliconized glass cover slides	Hampton Research
Syringe filters	Sartorius, Millipore
Vacuum-driven filtration systems	Millipore, Nalgene

2.1.5 Molecular cloning kits

The DNA NucleoSpin Plasmid kit, purchased from Macherey-Nagel, was used for plasmid purification. For gel extraction and purification of PCR products, Wizard SV Gel and PCR Clean-Up system, purchased from Promega Corporation was used.

2.1.6 Chromatography materials

Ni-NTA agarose resin for protein purification was purchased from Qiagen. Superose™ 6 10/300 GL, HiLoad™ 16/60 Superdex™ 75 prep grade, and HiLoad™ 16/60 Superdex™ 200 prep grade prepacked Tricorn columns and chromatography media were purchased from GE Healthcare Life Sciences.

2.2 Methods

This section of the thesis provides a general description of the experimental procedures used in the study. Protein-specific modifications to the protocols, if any, are mentioned in Chapter 3.

2.2.1 Preparation of competent cells

2.2.1.1 Preparation of chemically competent *E. coli* cells

Competence, the ability of bacteria to take up exogenous genetic material from the environment, can be artificially induced by causing passive and transient permeability of the cell through exposure to suitable chemicals (most commonly, calcium chloride) or to an electric field. In this study, high-efficiency chemically competent *E. coli* for cloning and expression purposes was prepared by the Inoue method (Inoue*, et al.*, 1990). Briefly, an overnight pre-culture (of DH5α-T1R cloning strain or of expression strains) grown at 37 °C was used to inoculate a larger volume of SOB medium [Appendix II] (containing suitable antibiotic, if required) and incubation was carried out overnight at 21 °C. When an OD_{600} of 0.4 – 0.5 was reached, cells were harvested in pre-chilled tubes by centrifugation (2500 rpm, 20 min, 4 °C) and gently resuspended in 80 ml ice-cold Transformation Buffer ($TB^{§}$) [Appendix II]. After a brief incubation on ice, cells were collected by centrifugation (2500 rpm, 20 min, 4

°C), gently resuspended in 20 ml ice-cold TB[§] and treated with 1.5 ml DMSO. Aliquots of 100 µl were dispensed into pre-chilled microfuge tubes, flash-frozen with liquid nitrogen and stored at -80 °C until further use. All the above steps were carried out under sterile conditions. Cells were tested for background and DNA contamination and the transformation efficiency was evaluated using a plasmid DNA sample of known concentration.

2.2.1.2 Preparation of electrocompetent *M. smegmatis* cells

Mycobacterium smegmatis has been classified as a risk group 2 (S2) organism. Therefore, in accordance with established laboratory safety guidelines, all open work with *M. smegmatis* cells were carefully performed in a laminar flow hood in a biological safety level 2 (S2) laboratory. For preparation of electrocompetent cells, a single colony of *M. smegmatis groEL1ΔC* cells (Noens*, et al.*, 2011) was used to inoculate a small volume (5 ml) of 7H9 complete medium [Appendix II] and incubated with shaking at 37 °C for 2 to 3 days. This pre-culture was used to inoculate a larger volume (50 ml) of 7H9 complete medium and incubated overnight at 37 °C. After cooling the culture on ice for a couple of hours, cells were collected by centrifugation (4000 rpm, 10 min, 4 °C). Cells were resuspended in 50 ml ice-cold 10% (v/v) glycerol and recovered by centrifugation (4000 rpm, 10 min, 4 °C). The above step was repeated twice. The resulting pellet was resuspended in 2 ml ice-cold 10% (v/v) glycerol, aliquoted into 100 µl fractions, flash-frozen in liquid nitrogen and stored at -80 °C until future use. The entire preparation was carried out under sterile conditions.

2.2.2 Molecular cloning

2.2.2.1 Ligase-dependent cloning

Conventional ligase-dependent cloning experiments involve the insertion of a DNA fragment into a plasmid vector containing a multiple cloning site, a drug resistance gene marker, an inducible promoter to drive protein expression, and if necessary, a suitable tag for affinity purification of the expressed protein(s).

2.2.2.1.1 Preparation of insert

DNA fragments to be inserted into the vectors of choice were prepared by PCR. In this study, genomic DNA of *M. tuberculosis* strain H37Rv or a different plasmid harboring the gene(s) of interest was used as DNA template for PCR. Each PCR mix contained 0.1 – 0.2 µg of template DNA, 0.5 µM of each primer (forward and reverse) designed to amplify region of interest and to engineer appropriate restriction sites that enable generation of suitable overhangs [Appendix I], 0.2 mM of each of the four dNTPs, 1 – 3% DMSO, 1X final concentration of suitable DNA polymerase buffer (+$MgCl_2$) and prescribed concentration of the DNA polymerase (*Pfu* DNA polymerase or Phusion® high-fidelity DNA polymerase) in a total reaction volume of 50 µl made up using nuclease-free H_2O. The enzyme was added immediately prior to the reaction. Appropriate positive and negative control reactions were included and PCR was carried out in a Mastercycler® Personal or a Mastercycler® Gradient thermal cycler (Eppendorf) by the following general protocol with minor modifications, if necessary:

Initial denaturation	97 °C, 2 - 3 min	
Denaturation	97 °C, 30 s	x 30 cycles
Primer annealing	55 °C / 60 °C, 30 s	
Primer extension	72 °C, 2 min per 1 kb for *Pfu*; 20 s per 1 kb for Phusion	
Final extension	72 °C, 10 min	
Storage	4 °C	

The resulting PCR products were mixed with DNA sample buffer [Appendix II] and visualized by agarose gel electrophoresis [Appendix II], a technique that separates polynucleotides based on size, performed at constant voltage (100 V) using a Consort EV243 electrophoresis system (Wolflabs). The size of the polynucleotide of interest was verified by comparison with a DNA ladder loaded on the same gel; the PCR products were purified, quantified by measuring absorbance at 260 nm using an ND-1000 NanoDrop spectrophotometer (PEQLAB) and frozen until further use.

2.2.2.1.2 Restriction digestion and ligation

Vector DNA and purified PCR products were digested with restriction endonucleases, using enzyme concentration, incubation temperature and duration as prescribed by the

manufacturer. The digested products were electrophoresed and the linearised materials were excised and purified. Vector and insert were mixed in a stoichiometric ratio of approximately 1:3 and ligated using DNA ligase, according to the manufacturer's instructions.

2.2.2.1.3 Transformation

2.2.2.1.3.1 Heat-shock transformation of chemically competent *E. coli*

For each transformation, a frozen aliquot (100 µl) of competent *E. coli* cells was thawed on ice. One or a few µl of DNA (plasmid DNA/ligation mix) was added to the cells, mixed gently and incubated on ice for 20 – 30 min. The cells were administered a brief heat shock (42 °C, 90 sec) and immediately placed on ice for a few min. This was followed by addition of approx. 500 µl of SOC medium [Appendix II] and incubation, with shaking, at 37 °C for 45 – 60 min. The transformation mix was spun down (3000 rpm, 2 min); the cell pellet was resuspended in SOC medium (approx. 150 µl), plated out on an LB-agar plate supplemented with appropriate antibiotic and incubated overnight at 37 °C. For co-expression experiments with two different plasmids, 1 µl of each plasmid encoding a component of the complex (AccA3-AccD5 / AccA3-AccD4 / AccA3-AccD6) was added to an aliquot of competent *E. coli* cells and the transformation protocol was performed as described above.

2.2.2.1.3.2 Transformation of electrocompetent *M. smegmatis*

For each transformation, a frozen aliquot (100 µl) of electrocompetent *M. smegmatis groEL1ΔC* cells was thawed on ice. The expression construct solution (1 µl) was pipetted into a pre-chilled tube and added with 80 µl of competent cells and 80 µl of 10% (v/v) ice-cold glycerol. The suspension was vortexed, centrifuged briefly and transferred into an electroporation cuvette. Electroporation was performed thrice at 2.5 kV using the Electroporator 1000™ (Stratagene) without intermediate cooling. After the electroporation, the cuvette was immediately cooled on ice. Ice-cold 7H9 complete medium [Appendix II] (1 ml) was added; the suspension was transferred to a 2-ml microfuge tube and incubated, with shaking, at 37 °C for 2 to 4 hours. The

suspension (approx. 100 µl) was plated out on 7H10-agar plates [Appendix II] containing 50 µg/ml hygromycin. The plate was incubated at 37 °C for 3 to 4 days.

2.2.2.1.4 Confirmation of construct integrity

Positive transformants from the transformed LB agar plates were identified by colony PCR using reaction mixtures as described in section 2.2.2.1.1 and single colonies as DNA templates. Bacteria from each colony pool were reserved to inoculate small volumes of LB containing suitable antibiotic. Plasmid DNA was isolated from cultures corresponding to the positive transformants and incorporation of insert was verified by restriction analysis of the plasmids. Sequencing of plasmid DNA samples was outsourced to Eurofins MWG Operon.

2.2.2.2 Ligase-independent cloning

LIC is a simple two-step method to produce expression constructs that makes use of 3' to 5' activity of T4 DNA polymerase to create specific 10-15 base single overhangs in the expression vector (Chen*, et al.*, 2000). LIC vectors used in this study contain the gene coding for eGFP flanked by two *Bsa*I sites. In step 1 of the method, the vector is linearised by restriction using *BsaI*, thereby removing the eGFP gene. In step 2, the linearised vector is treated with T4 DNA polymerase in the presence of dTTP. As a consequence of 3' to 5' activity of the polymerase, bases are removed from both 3'-ends until the first T base is reached. This protocol results in two specific overhangs (of 10 and 12 bases respectively) in the LIC vector, which facilitate specific, ligase-independent annealing.

2.2.2.2.1 Preparation of LIC vector and insert

The LIC vector was linearised by *BsaI* digestion performed as per the instructions of the manufacturer. The linearised vector was gel-purified, treated with T4 DNA polymerase in the presence of 2.5 mM dTTP and incubated at room temperature for 30 min. The polymerase was inactivated by heating at 75 °C for 20 min. The insert was obtained by PCR amplification (as explained in section 2.2.2.1.1) using suitable LIC primers [Appendix I] designed as follows:

Forward primer (5' to 3'): CAGGGCGCC**ATG**-gene of interest
Reverse primer (5' to 3'): GACCCGACGCGG**TTA**-gene of interest (rev. comp.)

The forward primer contained the complementary overhang (underlined), the start codon (in bold) and a reasonably long overlap with the gene of interest. The reverse primer contained the complementary overhang (underlined), one or more stop codons (in bold) if no C-terminal tag was used and a suitably long overlap with the reverse complement of the gene of interest. The PCR product was purified and treated (as above) with T4 DNA polymerase in the presence of 2.5 mM deoxyadenosine triphosphate.

Table 2-2. List of constructs for protein expression in *E. coli* / *M. smegmati*s

Protein (Gene, UniProt ID)	**Construct ID**	**Residue range**	**Vector backbone**
AccA3 (Rv 3285, P96890)	M02[#]	Full-length	pETM-11
	M03	Full-length	pMyNT
	M17	Full-length	pCDF-13
	M18	Full-length	pCDF-11
	M24	1 – 454 aa	pETM-11/LIC
	M25	1 – 471 aa	pETM-11/LIC
	M26	1 – 480 aa	pETM-11/LIC
AccD5 (Rv3280, P96885)	M08	Full-length	pETM-11/LIC
	M15	Full-length	pCDF-13
	M16	Full-length	pCDF-11
	M32	Full-length	pnEK
AccD5-AccE5 (Rv3280-81, P96885-P96886)	M09	Full-length	pMyNT
AccE5 (Rv3281, P96886)	M10	Full-length	pETZ2-1a/LIC
	M11	Full-length	pETSUMO/LIC
	M12	1 – 68 aa, 90 – 178 aa	pETSUMO/LIC
	M19	Full-length	pCDF-13
	M20	Full-length	pCDF-11
	M13[§]	Full-length	pKW08 (N-ter GFP)
	M14[§]	Full-length	pKW08 (C-ter GFP)
AccD4 (Rv3799c, O53578)	M06[#]	Full-length	pETM-11
	M21	Full-length	pCDF-13
	M22	Full-length	pCDF-11
	M33	Full-length	pnEK
AccD6 (Rv2247, P63407)	M27[*]	Full-length	pET-28a
	M34	Full-length	pnEK

[#]Prepared by Simon Holton. [*]Provided by collaborators at Lionex GmbH, Braunschweig. [§]Provided by collaborators at Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw.

2.2.2.2.2 Annealing of LIC vector and insert

The treated vector and insert DNA were mixed in a molar stoichiometry of 1:3, added with 1 µl of 25 mM EDTA and incubated at room temperature for 5 min. The annealing mixture was used to transform *E. coli* DH5α competent cells and construct integrity was verified as explained in sections 2.2.2.1.3 and 2.2.2.1.4 respectively.

2.2.3 Heterologous overexpression of proteins

2.2.3.1 Large-scale expression of proteins in *E. coli*

Optimal conditions (inducer concentration, post-induction incubation temperature and duration) for soluble overexpression of proteins in *E. coli* were identified by small-scale expression trials and later used for large-scale expression. An overnight pre-culture (grown from a single colony of cells/glycerol stock of cells bearing the expression construct) was used to inoculate large volumes (typically 500 or 750 ml in baffled 2-litre Erlenmeyer flasks) of Terrific Broth (TB$^{\#}$) [Appendix II] supplemented with appropriate antibiotic(s) to select for transformed cells of the relevant expression strain. The cultures were incubated at 37 °C, with shaking at 200 rpm, until an OD_{600} of 0.8 to 1.0 was reached. The cultures were cooled down and protein expression was induced by the addition of IPTG to a suitable final concentration (typically 0.1, 0.25 or 0.5 mM). The incubation was continued at 200 rpm and a lower temperature (typically between 18 °C and 30 °C) for another 12 to 16 hours. The cells were harvested by centrifugation at 5500 rpm for 30 min using an Avanti centrifuge J20-XP (Beckman Coulter) and the cell pellets were stored at -20 °C until further use.

2.2.3.2 Large-scale expression of proteins in *M. smegmatis*

A saturated pre-culture (grown, for a couple of days, from a single colony of *M. smegmatis groEL1ΔC* cells bearing the expression construct) was used to inoculate large volumes (typically 500 ml in baffled 2-litre Erlenmeyer flasks) of 7H9 expression medium [Appendix II] supplemented with 50 µg/ml hygromycin. The cultures were incubated at 37 °C, with shaking at 220 rpm, and cell growth was followed spectrophotometrically. The generation time for *M. smegmatis* is about 3 hours. When an OD_{600} of approx. 3.0 was reached, protein expression was induced by

the addition of 4.55 ml of 220X acetamide solution [Appendix II] per litre of culture. The incubation was continued (at 37 °C, 220 rpm) for another 16 hours. The cells were harvested by centrifugation for 60 min at 6000 rpm using a JLA-81000 rotor in an Avanti centrifuge J20-XP (Beckman Coulter) and the pellets were stored at -20 °C until use.

2.2.4 Purification of proteins

Proteins were purified in two steps:
(1) Immobilized metal-affinity chromatography (IMAC) is an effective, widely used method of protein purification (Porath*, et al.*, 1975). It is based on the property of His and Cys side-chains of proteins to interact strongly with divalent transition metal cations. Recombinant proteins expressed in tandem with a stretch of His residues (a 'polyhistidine tag') can be separated from the majority of non-tagged proteins in the supernatant of the cell lysate by adsorbance on a solid matrix embedded with divalent metal cations (e.g., Ni^{2+} or Co^{2+} immobilized with nitrolotriacetic acid [NTA] on agarose beads). The bound protein can be eluted competitively by application of high concentrations of imidazole, by protonation of the polyhistidine tag by pH-change, or by detachment of the bound cations from the solid phase using metal chelators like EDTA.

(2) Size-exclusion chromatography (SEC) is a method of separation of molecules based on differences in their hydrodynamic volume (their "size") (Porath & Flodin, 1959). The material from which the protein of interest is to be purified is allowed to traverse a bed of porous beads (of particular dimensions) made of material such as cross-linked agarose or dextran, usually held in a vertical column. Separation of molecules is achieved owing to the fact that molecules of different sizes migrate differently through the column: large molecules enter the bead pores with low probability and as a result, elute out quickly; whereas smaller molecules can enter the bead pores with a higher probability and need to pass a larger volume before eluting from the column.

2.2.4.1 IMAC purification of polyhistidine-tagged proteins

Cell pellets were freeze-thawed, resuspended in lysis buffer [Appendix III] (approx. 5 ml per gram of cell pellet) and treated with DNaseI and protease inhibitor cocktail prepared according to the manufacturer's instructions. The resuspended cells, being cooled on ice, were lysed by ultra-sonication. *E. coli* cells were sonicated by three 3-min pulse cycles (0.5 sec on + 0.5 sec off), with 2-min pauses between cycles, in a Bandelin Sonopuls sonicator equipped with a VS 70T probe operating at 45% power. When *M. smegmatis* was used for protein production, the cell pellets were lysed by four 3-min pulse cycles (0.5 sec on + 0.5 sec off), with 2-min pauses between cycles, in a Bandelin Sonopuls sonicator equipped with a MS 73 probe operating at a maximum of 40% power. To avoid producing aerosols, lysis of *M. smegmatis* cell pellets was performed in closed vessels (50-ml tubes with a hole in the lid that just allows the sonicator probe). Cell debris was removed by centrifugation (19000 rpm, 4 °C, 60 min) using an SS-34 rotor in an RC-26 Plus centrifuge (Sorvall). The soluble cellular fraction was passed through a 0.45-µm syringe filter and loaded onto Ni-NTA resin pre-equilibrated with binding buffer [Appendix III] in a polyprep chromatography tube. The resin was allowed to rock gently on a platform shaker for about an hour, to allow efficient binding of the polyhistidine-tagged recombinant protein to the Ni-NTA resin. The resin was then washed with several bed volumes of binding buffer and wash buffer(s) [Appendix III] to get rid of non-specifically bound contaminant proteins. His-tagged protein was eluted using elution buffer [Appendix III] containing high concentrations of imidazole. To facilitate conformational stability of proteins, purification protocols were carried out at 4ºC (cold room/cold cabinet) and proteins were handled carefully on ice.

2.2.4.2 Size-exclusion chromatography

Final purification, buffer exchange and assessment of the oligomeric status of affinity-purified proteins were done by size-exclusion chromatography. Proteins purified by affinity chromatography were concentrated and loaded on to a Superose™ 6 10/300 GL or a HiLoad™ 16/60 Superdex™ 200 prep grade column, pre-equilibrated with SEC buffer [Appendix III], using an ÄKTA purification system (GE Healthcare Life Sciences). Protein loading and elution were performed at constant flow rates; eluates were collected as equal-volume fractions in 18-mm collection

tubes or in 96-well deep well blocks. The oligomeric states of the proteins of interest were inferred by comparison with the elution volumes of gel filtration calibration standards run under similar conditions. Whenever required, affinity tags on the purified proteins were removed by treatment with TEV protease, an efficient enzyme that has a recognition site between the target protein and the affinity tag in the constructs used in this study. Typically, protein: TEV protease amount ratio of 50:1 was used for the cleavage process (a few hours at room temperature or overnight at 4 ºC).

2.2.4.3 Evaluation of protein biochemical purity

Electrophoresis is the most widely used technique for the separation and analysis of proteins based on their propensity to migrate in an electric field. The commonly used method of electrophoresis employs the anionic detergent SDS, which binds uniformly to most proteins, about one SDS molecule per two amino acid residues. SDS imparts a large net negative charge to the proteins, thereby masking the intrinsic charge of the proteins and rendering a similar charge-to-mass ratio to each protein. Additionally, binding of SDS and heating of samples prior to electrophoresis help in the denaturation of secondary and non–disulfide–linked tertiary structures of proteins and hence, most proteins assume a similar shape. Therefore, during SDS-PAGE proteins are separated almost exclusively on the basis of molecular weight, with heavier proteins migrating slower than lighter ones. The purity of samples is evaluated visually and the molecular weights of proteins are estimated by comparison to molecular weight standards electrophoresed on the same gel.

In this study, protein samples were tested for purity by SDS-PAGE using the Mini-Protean III Tetracel vertical protein electrophoresis system (Bio-Rad). Depending on the molecular weight of the protein(s) of interest, appropriate concentrations (typically 7.5%, 12% or 15%) of acrylamide were used in the preparation of the polyacrylamide gels [Appendix II]. Samples were mixed with SDS-PAGE sample buffer [Appendix II], heated at 75 ºC for 10 min and electrophoresed at constant voltage (180V – 200V). Gels were thoroughly washed with ddH_2O to remove SDS, stained using the Coomassie blue safe stain [Appendix II] and destained using ddH_2O.

2.2.4.4 Protein quantification and concentration

Estimation of protein concentration was performed by measuring light absorption at 280 nm using an ND-1000 NanoDrop spectrophotometer (PEQLAB), the underlying working principle of which is the Beer-Lambert law (Beer, 1852). The Beer-Lambert law explains the relationship of protein molar concentration (c, mol.l^{-1}) with the path length of light (d, cm), the absorption (A) and the molar extinction coefficient (ε, L.mol^{-1}.cm^{-1}) of the protein. The extinction coefficient (ε) of the protein is the sum total of the extinction coefficients of the absorbing species (Tyr, Trp, disulphides) multiplied by their respective number per molecule of the protein.

$$c = A / (\varepsilon . d)$$

$$\text{where, } \varepsilon = (n_{Tyr} . 1280) + (n_{Trp} . 5690) + (n_{disulphide} . 120)$$

In this study, proteins were concentrated as required, by centrifugation (at 4000 rpm) using centrifugal ultrafiltration tubes of suitable volume (0.5 ml, 6 ml or 20 ml) and molecular weight cut-off (10 kDa or 30 kDa) in a refrigerated 5415R or 5810R bench-top centrifuge (Eppendorf).

2.2.4.5 Verification of protein identity

Mass spectrometry is a sensitive analytical technique that helps determine the masses of molecules by measuring the mass-to-charge (m/z) ratios of the corresponding ions (Fenn*, et al.*, 1989). In a typical mass spectrometry experiment, the sample is ionized by the effect of an impact of an electron beam or a vaporized ionic solution. These ionized particles are sorted by mass in a mass analyzer chamber that supports an isotropic electromagnetic field. The sorted ions are monitored by a detector that allows the calculation of the intensities of each 'mass-by-charge' species present in the sample. A useful application of mass spectrometry in protein science is peptide finger printing wherein the protein sample, whose identity has to be established/verified, is subjected to specific proteolytic cleavage and the resulting peptides are analysed by electrospray ionization time-of-flight or matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) (Pappin*, et al.*, 1993). The experimental peptide peak list is then computationally compared with a theoretically

generated list of peptides corresponding to protein sequences from a standard database (e.g., UniProt); with the help of specialized software, the results of the comparison are analysed statistically to arrive at candidate proteins.

Analytical characterisation of molecular weights of purified protein samples in this study was done using a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems) at the in-house Sample Preparation and Characterisation (SPC) facility. For a MALDI-TOF experiment, protein sample at a concentration of around 1 mg/ml in a low-salt buffer (<50 mM salt) was mixed with matrix solution (50% acetonitrile, 0.1% trifluoroacetic acid, sinapinic acid to saturation) at a ratio of 1:2, 2:1 or 1:1 and spotted onto sample plates. The measurement was done after the samples dried out. Confirmation of protein identity by peptide mass fingerprinting was done by the Proteomics Core Facility at EMBL, Heidelberg.

2.2.5 Reconstitution of protein complexes

Attempts to reconstitute YCC complexes for crystallography and kinetic studies, were made using the following methods:

2.2.5.1 *In vitro* reconstitution by purification and co-elution

One of the presumably simplest and most common ways to assemble a protein complex *in vitro* is to purify the constituent proteins individually and to incubate them together in appropriate stoichiometric ratios. It has been known that many YCCs exist as dodecameric complexes possessing 1:1 molecular stoichiometry with regard to the numbers of α- and β-subunit protomers (Haase*, et al.*, 1982, Kimura*, et al.*, 1998, Diacovich*, et al.*, 2002, Gago*, et al.*, 2006). Individual components of *M. tuberculosis* YCCs (AccA3 and AccD5 / AccA3 and AccD4) were purified to homogeneity, mixed together at equimolar concentrations and incubated overnight at 4ºC to facilitate the formation of protein complexes. Formation of a high molecular weight YCC complex was verified by subjecting each protein mixture to analytical chromatography on Superose™ 6 10/300 GL.

2.2.5.2 *In vivo* reconstitution by co-expression

Heterologous expression of components of a complex in the same host cell is another useful strategy to obtain the holocomplex in a form suitable for structural and/or functional characterisation. *E. coli* is the expression system of choice for co-expression trials as it has advantages like the possibility of obtaining large quantities of protein, the absence of certain post-translational modifications that might hinder crystallisation and the availability of well-established high-throughput platforms that enable parallel testing of multiple constructs, strains and culture conditions (Berrow, *et al.*, 2006, Tolia & Joshua-Tor, 2006, Perrakis & Romier, 2008, Bieniossek, *et al.*, 2009). Co-expression can also provide means of facilitating proper folding of the interacting protein partners, thereby overcoming folding and solubility problems, if any, associated with individually expressed proteins (Li, *et al.*, 1997, Romier, *et al.*, 2006). Co-expression can be achieved using either single or multiple constructs; when more than one expression construct is co-transformed into a single host cell, the vectors generally harbor different antibiotic resistance selection markers (Perrakis & Romier, 2008, Zeng, *et al.*, 2010) and compatible origins of replication (Johnston, *et al.*, 2000, Perrakis & Romier, 2008). When a single plasmid is used for co-expression, it can be either monocistronic (a separate promoter for each gene that is transcribed in a distinct mRNA) or polycistronic (a common promoter for multiple genes that are transcribed in the same mRNA). In the present study, a series of constructs for single protein expression of binary complexes were created using pET- and pCDF-based vectors [Figure 2-1], which harbor different resistance markers (kanamycin and spectinomycin) and compatible origins of replication (Busso, *et al.*, 2011).

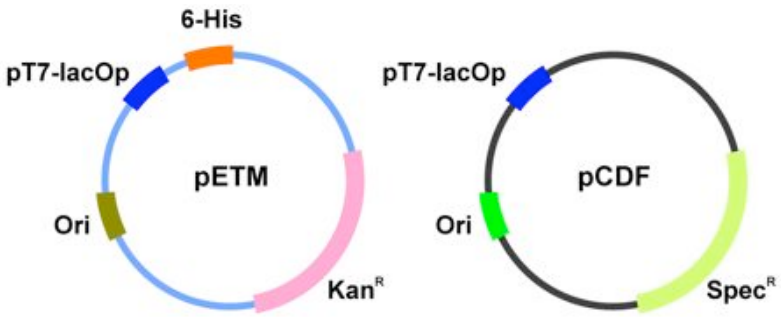

Figure 2-1. Representation of vector backbones used in co-expression constructs. Ori, origin of replication; pT7-lacOp, T7 promoter-*lac* operator system; 6-His, region coding for the hexahistidine affinity tag; Kan^R, kanamycin resistance gene; Spec^R, spectinomycin resistance gene.

The constructs were used to co-transform expression strains of *E. coli* as described in section 2.2.2.1.3.1. Following primary growth, overexpression of the two proteins was induced by the addition of IPTG; as one of the interacting partners was expressed with a polyhistidine tag, the potential binary complex assembled *in vivo* was purified by IMAC followed by SEC, as explained earlier.

2.2.6 Kinetic characterisation

Complementary to structural studies on enzymes, it is usually of scientific interest to examine if the *in vitro* catalytic activity of the material under study is intact. Some of the enzyme complexes reconstituted in this study were tested for their enzymatic activity by standard assays.

2.2.6.1 Michaelis-Menten kinetics

It is often assumed that biochemical reactions involving a single substrate follow, irrespective of conformity to the underlying assumptions, Michaelis-Menten kinetics, probably the simplest and most popular model of enzyme kinetics. The Michaelis-Menten equation of enzyme kinetics is an algebraic expression of the relationship between substrate concentration and the velocity of the enzymatic reaction; it is based on the hypothesis that the rate-limiting step in enzymatic reactions is the breakdown of the enzyme-substrate complex to product and free enzyme and on the Briggs-Haldane steady state approximation (Michaelis*, et al.*, 2011). The Michaelis-Menten equation, the kinetic rate equation for a one-substrate enzymatic reaction is as follows:

$$V_0 = (V_{max}\,[S]) / (K_m + [S])$$

where, V_0 and V_{max} are the initial and the maximum velocity of the enzymatic reaction respectively, [S] is the substrate concentration and K_m is the Michaelis constant, the substrate concentration at which V_0 equals half of V_{max}.

2.2.6.2 Assay for carboxylase activity

Carboxylase activity of YCC enzyme complexes (AccA3-AccD5 / AccA3-AccD4) was estimated by a coupled enzyme assay that follows the rate of ATP hydrolysis spectrophotometrically (Diacovich*, et al.*, 2002).

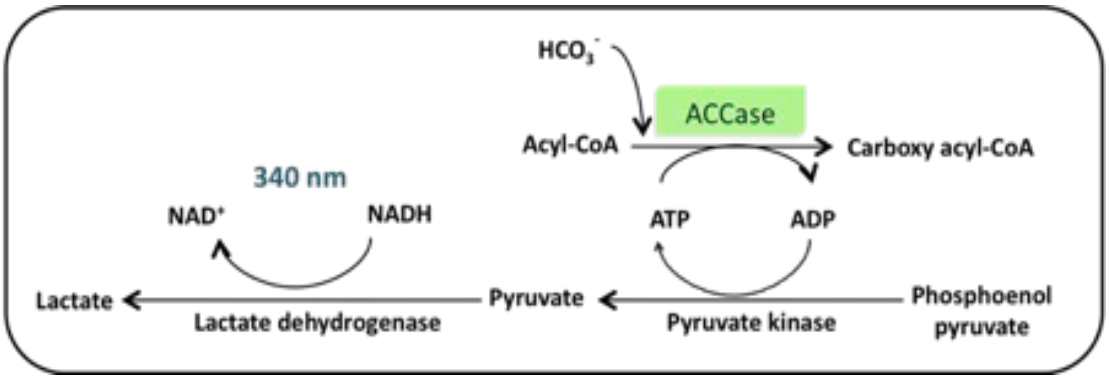

Figure 2-2. Schematic of the coupled enzyme reaction used to measure carboxylase activity. YCC (also denoted as ACCase) activity can be assayed by spectrophotometrically monitoring the oxidation of NADH, coupled to the consumption of ATP.

In the presence of bicarbonate and ATP, the YCC enzyme carboxylated the substrate acyl-CoA to produce carboxy acyl-CoA. The production of ADP during the reaction was coupled to pyruvate kinase and lactate dehydrogenase, and the oxidation of NADH was probed at 340 nm. The optimal enzyme concentration to be used in the assay for the establishment of kinetic parameters was determined by preliminary trial experiments. The assay mixture contained 7 units of pyruvate kinase, 10 units of lactate dehydrogenase, 50 mM $NaHCO_3$, 3 mM ATP, 0.5 mM phosphoenol pyruvate, 0.2 mM NADH, 0.3 mg/ml BSA, 100 mM K_2HPO_4 pH 7.6 and 5 mM $MgCl_2$ and varying concentrations of substrate. Reactions were initiated by the addition of YCC enzyme to the assay mixture and were maintained at 30 °C. Data were acquired using a monochromator-based multimode microplate reader (Infinite® M1000 PRO, Tecan). The kinetic parameters K_m and V_{max} were determined by fitting the mean velocities *versus* the substrate concentration to the Michaelis-Menten equation of enzyme kinetics using nonlinear regression analysis, executed by the program Prism 5 (GraphPad Software™).

2.2.7 Biophysical characterisation

2.2.7.1 Static light scattering

The phenomenon of static light scattering (SLS) is very useful in the accurate experimental determination of the absolute molecular mass of a protein sample in solution through exposure of the sample to low intensity laser light (690 nm) (Debye,

1947). During an SLS experiment, the intensity of the light scattered by the sample is recorded as a function of angle. The molar mass, root mean square radius, and second virial coefficient (A_2) may be calculated from the measured scattering intensities. SLS, in principle, eliminates ambiguities in molecular weight determination using SEC elution volumes arising from unusual shapes of analyte molecules. In this study, SLS measurements were made in the chromatography mode using a miniDAWN Tristar SLS detector (Wyatt Technology) designed to collect scattering data from three different angles, placed in the flow path of the Äkta Purifier and pre-equilibriated with the corresponding SEC buffer. Data analysis was performed using the software ASTRA version 5.3.4.11 as per manufacturer instructions.

2.2.7.2 Thermofluor

Thermofluor is a fluorescence-based thermal stability assay that monitors the effects of buffer conditions on temperature-dependent protein unfolding (Pantoliano*, et al.*, 2001). It presents a valuable method to study protein behaviour (solubility, stability and conformational homogeneity) as a function of the chemical environment (pH, salt nature and concentration, presence of additives, etc.) and hence to identify condition(s) amenable to crystallisation (Ericsson*, et al.*, 2006). Protein was added to 96 different conditions of a sample stabilization screen (prepared by the in-house SPC facility) in a PCR plate; each well contained the following components: 16 µl ddH_2O + salt (of varying concentration), 5 µl buffer (of varying chemical nature and concentration), 2 µl diluted SYPRO® orange dye (Invitrogen), 2 µl protein sample (20 – 100 µM). Thermal denaturation analysis was performed in an ICycler thermal cycler with a MyIQ optical module (Bio-Rad) using a gradient of 1 °C/min from 5 °C to 95 °C. Data was analysed using Prism 5 (GraphPad Software™).

2.2.7.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a spectroscopic technique wherein the CD (i.e., the difference in the absorption of left-handed and right-handed circularly polarized light) of chiral molecules measured over a range of wavelengths is used to draw inferences on the physiochemical properties of these molecules. CD

measurements made in the far-UV range provide information on the secondary structure or conformation of proteins and are therefore useful in structural, kinetic and thermodynamic studies (Greenfield, 2007). For the assessment of secondary structure content by this method, proteins were dialyzed overnight against 20 mM sodium phosphate buffer (pH 7.5 or 8.0) and then diluted in the same buffer to final concentrations of 0.1 – 0.5 mg/ml. CD spectra between 180 and 260 nm wavelength were recorded in 0.5 nm steps on a Chirascan CD spectrometer (Applied Photophysics) at room temperature in a quartz cuvette of 0.1 cm path length; raw data were converted to molar ellipticity (deg cm^2 $dmol^{-1}$).

2.2.8 Protein crystallisation

Obtaining crystals of diffraction-quality is often a major bottleneck for macromolecular X-ray crystallography, because crystallisation is a complex process that involves multiple parameters. Protein crystals can be grown from highly concentrated solutions of pure proteins by various methods aimed at forcing the protein to a certain level of supersaturation by the controlled addition of precipitants or/and changes in ionic strength, protein concentration, etc.

2.2.8.1 Principles

2.2.8.1.1 The protein crystallisation phase diagram

The process of crystallisation proceeds in two distinct phases that require different conditions: nucleation and growth (Ataka & Asai, 1990, Chayen, 2005). Upon the formation of a critical nucleus, spontaneous growth of the crystal follows, provided the existence of an amenable environment. As crystallisation involves phase transition, it can be illustrated by a phase diagram that shows which state of the protein (liquid, crystalline or amorphous precipitate) is stable under a given set of crystallisation parameters. A typical phase diagram is produced experimentally by testing two parameters at a time, thus representing a two-dimensional slab of the multiparametric space of the phenomenon of crystallisation.

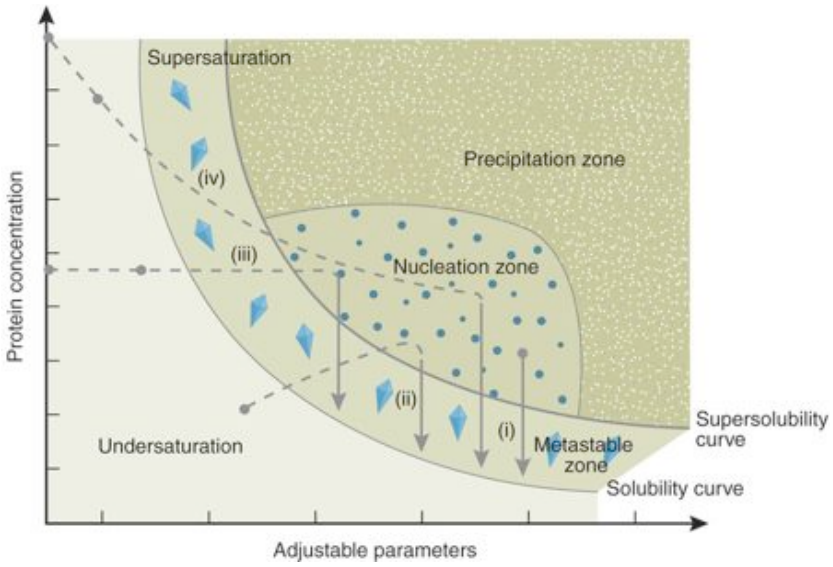

Figure 2-3. The protein crystallisation phase diagram. The solubility curve demarcates conditions that favour spontaneous nucleation (or phase separation or precipitation or crystal growth) from those under which the protein remains completely soluble. Each of the major crystallisation methods [(i) microbatch (ii) vapour diffusion (iii) dialysis (iv) free interface diffusion (FID)] adopts a different route to reach the nucleation zone (dashed lines) and the metastable zone (solid arrows). The starting conditions are represented by the filled circles. (Reprinted by permission from Macmillan Publishers Ltd: (Chayen & Saridakis, 2008)).

A phase diagram consists of four areas: (1) the undersaturation zone, where the protein is completely soluble and does not crystallize; (2) the nucleation zone, an area of moderate supersaturation where spontaneous formation of crystal nuclei takes place; (3) the metastable zone, an area of low supersaturation best suited for well-ordered crystal growth and stability, but unsuitable for further nucleation; (4) the precipitation zone, an area of high supersaturation where the protein precipitates, mostly irreversibly (Ataka & Asai, 1990, Chayen*, et al.*, 1996). Phase diagrams clearly represent the physical chemistry operating behind crystallisation and are hence useful in the design and fine-tuning of crystallisation experiments.

2.2.8.1.2 Protein crystallisation methods

There are several procedures and devices to cause supersaturation in a protein solution, generally by the gradual increase in the concentration of a precipitant such as polyethylene glycol (PEG) or salt. In addition to the most common methods (microbatch, vapour diffusion, dialysis and free interface diffusion), other methods available for attaining a solubility minimum are bulk crystallisation, evaporation,

liquid bridge, sequential extraction, pH-induced crystallisation, temperature-induced crystallisation and effector addition (McPherson, 1999).

(i) In microbatch experiments, crystallisation trials are maintained under low-density paraffin oil, which protects the drops from evaporation, contamination and mechanical shock (Chayen*, et al.*, 1990). As samples are set at their final concentration right at the start of experiment, unlike in diffusion-based methods, the conditions are constant within a normal duration of microbatch crystallisation experiment. This method is unsuitable only for crystallisation trials involving small volatile organic molecules that dissolve into oil.

(ii) In the most widely used vapour diffusion method, the protein solution is maintained as either a sitting or hanging drop that equilibrates against a reservoir of mother liquor containing crystallisation agents at known initial concentrations. As equilibration between the drop and the reservoir proceeds, conditions keep varying and the crystallisation parameter space is sampled. Besides transportation problems, this method has a few disadvantages like changes in pH due to volatile ions, and indeterminable and uncontrolled changes in drop volume and composition during the course of the experiment.

(iii) The dialysis method of crystallisation involves the separation of the protein from the crystallisation agent by a semi permeable membrane that selectively allows the precipitant molecules to gradually mix with the protein molecules.

(iv) In the FID technique, the protein solution and the precipitant are placed side-by-side; owing to slow diffusion of one solution into the other, a concentration gradient that varies with time is created. The system self-selects the optimal supersaturation levels for crystal nucleation and growth. This technique is generally useful as a fine-tuning strategy rather than an initial screening method (McPherson, 1999).

Each of the above-mentioned methods of crystallisation adopts a different mechanistic route to attain the desired level of supersaturation; these routes have been indicated by arrows in Figure 2-3. In this study, the vapour diffusion method of crystallisation was exclusively used.

2.2.8.2 Screening and optimization

The appropriate concentration of proteins, an important variable for crystallisation experiments, was identified using a pre-crystallisation test (PCT, Hampton Research) performed in hanging-drop format (1 µl protein + 1 µl PCT solution). Initial screening for crystallisation conditions by the sitting-drop vapour diffusion method was performed at the high-throughput crystallisation facility at the EMBL Hamburg (Mueller-Dieckmann, 2006). Commercially available crystallisation screens (such as Classic I and II from Qiagen and JBScreen Classic from Jena Bioscience) were used in these experiments. Drops of volume 400 – 600 nl were set, maintained at constant temperature (20 °C) and monitored periodically. Conditions that yielded crystalline hits were optimized manually by the hanging-drop vapour diffusion method. Optimisation experiments involved the variation of parameters like composition of the mother liquor (concentration of precipitant, pH, additives), crystallisation temperature (20 °C/ 4 °C) and drop volume. Crystallisation protocols of individual proteins are described in detail in the results and discussion chapter.

2.2.8.3 Reductive methylation

Reductive methylation of free amino groups, especially of residues found on the surface, has been shown to promote crystallisation of proteins presumably by means of immobilization of flexible side chains (Walter*, et al.*, 2006, Kim*, et al.*, 2008). It has also been demonstrated that methylation of proteins neither affects their biochemical function nor causes significant conformational changes (Rypniewski*, et al.*, 1993, Kobayashi*, et al.*, 1999). In an attempt to improve its crystallisability, the protein AccD4 was methylated by an adaptation of a standard reductive methylation protocol. Briefly, 1 mg of pure protein was dialyzed overnight against a buffer free of amino groups or alcohols [Appendix III]. The protein was then added with 20 µl of 1M dimethylamine-borane complex (ABC) and 40 µl of 1M formaldehyde and the reaction mixture was incubated with gentle shaking at 4 °C for 2 h. The above step was repeated and the incubation was continued for two hours at 4 °C. The mixture was left overnight at 4 °C after the addition of a final aliquot of 10 µl ABC. The methylated protein was concentrated and analyzed by SEC using a Tris-based gel filtration buffer [Appendix III] that helps quench the methylation reaction while

washing out the excessive reagents. The methylated protein was used in crystallisation screens after its homogeneity was verified by MALDI-TOF.

2.2.9 X-ray crystallography

X-ray crystallography, the most powerful means for macromolecular three-dimensional structure determination, was used for the structural characterisation of proteins in this study. The contribution of X-ray crystallography to the field of structural biology has been enormous because, as a technique, it is not restricted by molecular size limitations (unlike NMR); it enables structural elucidation at a very high resolution (unlike cryo-electron microscopy) with only minor destructive effects on the integrity of the crystalline sample (unlike electron crystallography). As a detailed illustration of the theory behind the tool is beyond the scope of the thesis, a limited general explanation (Wlodawer*, et al.*, 2008) of diffraction concepts and of the phasing method (molecular replacement) used in the study will be provided in the following sections.

2.2.9.1 Theory

2.2.9.1.1 Diffraction by crystals

Scattering of X-rays by electrons in the molecules of the investigated sample forms the basis of structural crystallography. The symmetric, periodic arrangement of molecules in unit cells throughout the entire volume of a crystal allows the crystal to act as a three-dimensional diffraction grating that results in enhanced scattering of X-rays in some directions and annihilated scattering in others. The fate of the scattering is determined by the geometry and content of the unit cell of the crystal and by the wavelength of the incident radiation. The geometrical conditions that give rise to phase agreement in the scattered radiation and hence to constructive interference can be described by Bragg's law (Bragg, 1913):

$$\mathbf{2d_{\mathit{hkl}}\ \sin\theta = n\lambda}$$

where, d_{hkl} is the interplanar spacing in a lattice, θ is the glancing angle of the electromagnetic wave, λ is the wavelength of the electromagnetic wave and n is a

whole number. The implication from Bragg's law, that every order of diffraction can be considered as reflection in a set of lattice planes (denoted by the reciprocal indices *hkl*), is a central theme of X-ray crystallography.

The characteristics of a wave (i.e., its amplitude and phase) diffracted from a set of crystal lattice planes (whose Miller indices are *h*, *k*, *l*) can be described by a mathematical function known as the structure factor $F_{(hkl)}$:

$$\mathbf{F_{(hkl)} = |F_{(hkl)}| \exp (i\alpha_{hkl}) = \Sigma_j f_j \exp [2\pi i (hx_j + ky_j + lz_j)]}$$

where, i is the imaginary number $\sqrt{-1}$, α_{hkl} is the phase of the diffracted wave, f_j is the scattering factor of the j^{th} atom whose positional coordinates are x_j, y_j, z_j, and the summation is over all atoms of the unit cell. The electron density ρ, the probability of existence of an electron at a specific location in three-dimensional space, can be obtained by the inverse Fourier transform of the above structure factor equation.

$$\mathbf{\rho (xyz) = V^{-1} \sum_{hkl} |F_{(hkl)}| \exp (-2\pi i (hx + ky + lz) + i \alpha_{hkl})}$$

where, *V* is the volume of the unit cell. A diffraction experiment facilitates the measurement of only the volume of the unit cell and the reflection intensities *I(hkl)*. These measured reflection intensities can be used in the calculation of the structure factor amplitudes $|F_{(hkl)}|$, while also taking into account the absorption correction factor and the radiation source-dependent combined geometry and polarization factor. However, the phase information of the reflections is missed out during the experiment, thereby resulting in a complication known as the crystallographic 'phase problem' in the calculation of electron density.

The phase problem can be solved by mathematical strategies (molecular replacement, difference Fourier synthesis, or direct methods) or by various experimental approaches (isomorphous replacement, anomalous dispersion, or their combinations); several combinatorial techniques (that utilize a mathematical strategy in addition to an experimental approach) have also been developed. Among these, molecular replacement was exclusively used as the phase determination method in this study.

2.2.9.1.2 Phase determination by molecular replacement

Molecular replacement (MR) is conceptually the simplest method of phase determination. If the target protein shares considerably high sequence homology (at least 25% sequence identity) with a protein (probe) of known three-dimensional structure, starting phases can normally be derived by computationally placing the probe molecule in the correct orientation and position in the unit cell of the target protein and calculating its theoretical diffraction pattern using the electron density equation. The probe molecule is then oriented and moved until the experimental patterns match the theoretical ones. Positioning of the molecule in three-dimensional space requires a six-dimensional search with three variables each for rotational and translational searches. As six-dimensional searches (three rotation angles and three translations) are computationally expensive, the process is normally broken up into two segments: a rotation search followed by a translation search.

The most common strategy of searching uses Patterson functions (Patterson, 1935). Patterson peaks, the interatomic distance vectors calculated from the target (P') and probe (P) molecules are superimposed in order to identify overlaps. This operation can be described by the following equation:

$$\mathbf{P' = P[R] + T}$$

where, [R] is a matrix describing the rotational parameters that bring the probe molecule into maximum coincidence with the target, and T is the translation vector that maximizes the overlap.

2.2.9.1.2.1 Patterson rotation function

Three rotational parameters are used to score the agreement between two Patterson functions as they are rotated with respect to each other. Only the Patterson function relatively close to the origin is considered, in order to enrich for intramolecular vectors and to reduce the contribution from the intermolecular counterparts. During the rotational search, the model has to be placed in a 'box' of appropriate dimensions so that all intermolecular vectors are outside the search volume; the box size must be at least the largest molecular radius plus the search sphere radius. The match can be

evaluated as a correlation coefficient or product function (RF) in Patterson space, or as the equivalent to the Patterson product function in reciprocal space.

$$\mathrm{RF}(\mathbf{R}) = \int_{r_{\min}}^{r_{\max}} P_{\mathrm{observed}}(\mathbf{u}) P_{\mathrm{model}}(\mathbf{R}, \mathbf{u})\, \mathrm{d}\mathbf{u}$$

RF is the product of the observed crystal Patterson $P_{observed}(u)$ and the rotated model Patterson $P_{model}(R, u)$ integrated over all points (u) in Patterson space within a sphere of radius r_{max} centered on the origin and omitting the origin peak out to a radius r_{min}. When a crystal has rotational symmetry operators (i.e., belongs to a space group other than the triclinic P1 that lacks symmetry), the search function respects the fact that the corresponding Patterson also has 'cross-vectors' between atoms belonging to different molecules related by symmetry. The variables of the rotation matrix are three angles designated by either of two alternative geometrical systems: Eulerian angles and Polar angles.

2.2.9.1.2.2 Patterson translation function

Once the optimal orientation of the model is deduced from the rotation search, the model structure factors can be calculated for every possible shift vector 't'. The translational search function then probes for correlation between the experimentally measured intensities and the Patterson cross-vectors of the symmetry-related molecules of the probe as it is moved within the cell. When the probe is optimally placed, the translation function has peaks at values corresponding to the translation vectors between the symmetry related molecules. The Patterson translation function for a translation T is defined as the product of the observed and model Pattersons, integrated over the whole cell,

$$T2(\mathbf{t}) = \int_V \left[P_{\mathrm{observed}}(\mathbf{u}) - \sum_{j=1}^{N_{\mathrm{sym}}} P_{jj}(\mathbf{u}) \right] \left[P_{\mathrm{model}}(\mathbf{u}, \mathbf{t}) - \sum_{j=1}^{N_{\mathrm{sym}}} P_{jj}(\mathbf{u}) \right] \mathrm{d}\mathbf{u}.$$

where, $P_{observed}(u)$ is the crystal Patterson at point u, $P_{model}(u, t)$ is the model Patterson shifted by the search vector 't' and the P_{jj} terms are the calculated self-vectors.

2.2.9.2 Data collection

X-ray diffraction data corresponding to crystals of the potential AccA3-AccD5 complex and of AccD6 were collected at synchrotron radiation microfocus beamlines at the ESRF (Grenoble, France) and PETRA III, EMBL-DESY (Hamburg, Germany) respectively [Table 2-3].

Table 2-3. Beamlines used for diffraction experiments

Synchrotron	Beamline	Energy	Wavelength	Detector
ESRF	ID23-2	14.2 keV	0.8726 Å	MarMOSAIC 225
PETRA III	P14	10 keV	1.2395 Å	PILATUS 6M

For data collection, the crystal was carefully harvested from the crystallisation drop, treated with a suitable cryoprotectant (mother liquor mixed with 10-15% (v/v) glycerol), mounted on a nylon crystal loop (Hampton Research) and flash-cooled either in liquid nitrogen (for transportation to the ESRF) or directly in the cryo-stream at the beamline (diffraction experiment at PETRA III, EMBL-DESY). The main purpose of flash-cooling (Haas & Rossmann, 1970) is to overcome the problem of radiation damage to the sample by avoiding ice formation that is bound to happen in slower cooling methods. The mounted crystal was exposed to X-rays after appropriate alignment using the option of 'three-click centering'. All diffraction experiments were performed at a temperature of 100 K using the single axis rotation method.

2.2.9.3 Structure solution and refinement

On the basis of diffraction spots from a few initial images, the crystal was indexed using the program *iMOSFLM* (Battye*, et al.*, 2011) and an appropriate data collection strategy was identified using *BEST* (Bourenkov & Popov, 2010). The programs *XDS* (Kabsch, 2010) / *iMOSFLM* (Battye*, et al.*, 2011) and *POINTLESS* (Collaborative Computational Project, 1994) were used in the integration of the diffraction data and in the determination of the Laue and space group of the crystal respectively. The integrated data were scaled and merged using *SCALA* (Evans, 2006). Using the scaled and merged intensities (in MTZ format) and the molecular weight of a monomer calculated from the protein sequence/number of residues, the Matthews coefficient

was calculated in order to estimate the number of molecules in the asymmetric unit (Collaborative Computational Project, 1994).

Initial phases for the structures were determined by MR using *Phaser* (McCoy*, et al.*, 2007) that performs maximum likelihood molecular replacement. The initial models were improved by alternative cycles of manual model building with *COOT* (Emsley & Cowtan, 2004) and maximum-likelihood refinement using *REFMAC5* (Murshudov*, et al.*, 1997). During restrained refinement, non-crystallographic symmetry (NCS) restraints were applied and in the later stages of refinement, TLS parameters were also employed. The progress of refinement was monitored using R_{free} (estimated using 5% of reflections set aside) and validation tools in *COOT* (Emsley & Cowtan, 2004, Emsley*, et al.*, 2010). The stereochemical qualities of the structural models were verified using the program *PROCHECK* (Laskowski*, et al.*, 1993) integrated in the *CCP4* suite. Structural analysis and graphical visualization were performed using *PyMOL* (The PyMOL Molecular Graphics System, Version 1.5.0.3 Schrödinger, LLC).

2.2.10 Small angle X-ray scattering

Small angle X-ray scattering is a structural biology method, which, unlike X-ray crystallography, is not restricted to elucidating low energy conformations of macromolecules within lattice planes of crystals. Although the resolution range of SAXS data is low (about 50 Å to 10 Å), it provides complementary information about overall molecular shape, conformation, folding, unfolding and aggregation. A key merit of SAXS is that it neither requires highly ordered crystalline sample nor is associated with size limitations like those inherent in electron microscopy and NMR studies.

In this study, synchrotron radiation small angle X-ray scattering data were collected at the X33 beamline of EMBL/DESY (Roessle*, et al.*, 2007) using a PILATUS 1M pixel detector at a sample-detector distance of 2.7 m and a wavelength of $\lambda = 1.5$ Å; the range of momentum transfer $0.01 < s < 0.6$ Å^{-1} was covered ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). For each sample, several solute concentrations in the range 16.1 to 0.7 mg/ml were measured. To monitor for possible radiation damage, eight

successive 15-second exposures of protein solutions were compared and no significant changes were observed. The data were normalized to the intensity of the transmitted beam and radially averaged; the scattering of the buffer was subtracted and the difference curves were scaled for protein concentration. The low angle data measured at lower protein concentrations were extrapolated to infinite dilution and merged with the higher concentration data to yield the final composite scattering curves. The data processing steps were performed using the program package *PRIMUS* (Konarev*, et al.*, 2003).

The forward scattering I(0) and the radius of gyration R_g were evaluated using the Guinier approximation (Guinier, 1939) assuming that at very small angles ($s < 1.3/R_g$) the intensity is represented as $I(s) = I(0) \exp(-(sR_g)^2/3)$. These parameters were also computed from the entire scattering patterns using the indirect transform package *GNOM* (Svergun, 1992), providing also the pair distribution function of the particle p(r) and the maximum size D_{max}. The molecular mass (MM) of the solute was evaluated by comparison of the forward scattering with that from reference solution of bovine serum albumin (MM = 66 kDa). The excluded volume of the hydrated particle was computed from the small angle portion of the data ($s < 0.25$ Å^{-1}) using the equation (Porod, 1982):

$$V = 2\pi^2 I(0) / \int_0^\infty s^2 I_{\exp}(s) ds$$

Prior to this analysis, an appropriate constant was subtracted from each data point to force the s^{-4} decay of the intensity at higher angles following the Porod's law (Porod, 1982) for homogeneous particles. For globular proteins, Porod (i.e. hydrated) volumes in nm^3 are about 1.7 times the MMs in kDa (Petoukhov & Svergun, 2013).

3 Results and Discussion

3.1 Characterisation of components of the propionyl-CoA carboxylase (PCC) of *Mycobacterium tuberculosis*

There have been extensive efforts towards understanding the mechanisms and regulation of lipid biosynthesis in *M. tuberculosis*; however, the enzymes that catalyze the formation of the building units of these complex biomolecules have not been completely characterised. One of the most important building/extender units is methylmalonyl-CoA, which is employed by key biosynthetic enzymes such as mycoseric acid synthases (Mathur & Kolattukudy, 1992) and multifunctional polyketide synthases (Sirakova*, et al.*, 2001, Gande*, et al.*, 2004) for the iterative extension of complex acyl chains. The common products of these enzymes are surface-exposed lipids like mycosides and phthiocerols that are unique to pathogenic species of mycobacteria and have been shown to play important roles in cell wall structure and permeability (Camacho*, et al.*, 2001), virulence and host-pathogen interactions (Cox*, et al.*, 1999, Sirakova*, et al.*, 2003). Methylmalonyl-CoA is generally produced in the cell either by the methylmalonyl-CoA mutase-driven isomerization of succinyl-CoA (Hunaiti & Kolattukudy, 1984) or by the propionyl-CoA carboxylase-driven carboxylation of propionyl-CoA. The *M. tuberculosis* genome harbors two genes coding for the putative methylmalonyl-CoA mutase subunits (Cole*, et al.*, 1998) that are known to be non-essential for mycobacterial survival (Sassetti*, et al.*, 2003) although the physiological relevance is not precisely known. On the other hand, at least one component of the relatively better-studied propionyl-CoA carboxylase (that catalyses the carboxylation of propionyl-CoA to methylmalonyl-CoA) is known to be essential for mycobacterial survival (Sassetti*, et al.*, 2003). Therefore, PCC is responsible for probably the only available route used by *M. tuberculosis* for the generation of the building blocks of certain essential lipids. Surprisingly, a comprehensive understanding of the structural and functional properties of this key enzyme is lacking to date.

Sub-chapter 3.1 describes the characterisation of AccA3, AccD5 and AccE5 – the component subunits of *M. tuberculosis* propionyl-CoA carboxylase. Various strategies tested for the *in vitro* assembly of the AccA3-AccD5 complex, kinetic characterisation, crystallisation experiments, and efforts towards the biophysical characterisation of AccE5 are elucidated.

3.1.1 Results

3.1.1.1 AccA3 and AccD5: formation of a binary complex

Full-length versions of *M. tuberculosis* AccA3 and AccD5 proteins could be individually overexpressed in moderate amounts (AccA3: 1 – 2 mg per litre of culture; AccD5: 2 – 5 mg per litre of culture) in *M. smegmatis groEL1ΔC* by acetamide-based induction using constructs M03 and M09 respectively [Table 2-2]. However, no overexpression of AccE5 could be observed with construct M09, as verified by SDS-PAGE and MALDI-TOF. Profiles from SEC experiments indicated the existence of AccD5 as a stable hexamer in solution. The α-subunit AccA3, on the other hand, eluted as a mixture of oligomers, with trimers possibly being the predominantly present species; significant amounts of AccA3 protein were also lost as very high-molecular weight aggregates in the void volume during SEC trials with Superose™ 6 10/300 GL.

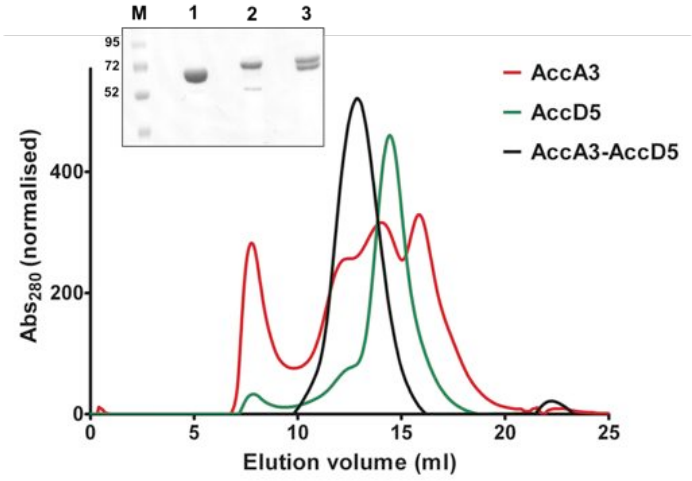

Figure 3-1. Purification of the AccA3-AccD5 complex. SEC elution profiles of AccA3, AccD5 and the AccA3-AccD5 complex. Molecular weight standards used and respective elution volumes (ml) in brackets: 670 kDa (12.1), 158 kDa (15.5). **Inset:** Peak analysis of (1) AccD5 hexamer, (2) AccA3 trimer and (3) AccA3-AccD5 complex by SDS-PAGE (12%); lane M: protein marker with standard molecular weights (in kDa).

The polyhistidine affinity tag at the N-terminal of AccD5 was cleaved by TEV protease treatment. Untagged AccD5 could be pulled down by polyhistidine-tagged AccA3 immobilised on Ni-NTA resin, indicating a direct physical interaction between the two proteins. Hexameric AccD5 and trimeric AccA3, upon mixing in an equimolar ratio, form a high-molecular weight binary complex, as observed by SEC [Figure 3-1]. Comparison with molecular weight standards showed that this complex is a dodecamer (six monomers each of AccA3 and AccD5) with a total molecular weight of approximately 700 kDa.

3.1.1.2 AccA3-AccD5 is an active carboxylase

The AccA3-AccD5 complex reconstituted *in vitro* was found to be a catalytically active carboxylase based on a spectrophotometric assay that utilizes a coupled enzyme reaction and monitors ATP hydrolysis at 340 nm. Preliminary trials identified an optimal enzyme concentration of 0.06 mg/ml (i.e., 0.5 µM) for the determination of kinetic parameters of the carboxylase reaction of AccA3-AccD5 on the known substrate, propionyl-CoA. The Michaelis-Menten kinetic parameters for the carboxylase reaction were calculated as: V_{max} = 1.33 ± 0.05 mM min^{-1} mg^{-1} and K_M = 0.13 ± 0.019 mM. The high efficiency of propionyl-CoA carboxylation of AccA3-AccD5 was evidenced by the turnover number k_{cat} = 0.045 s^{-1}, calculated based on the assumption that the number of molecular active sites is 6, in accordance with the hexameric arrangement of the core catalytic subunit AccD5.

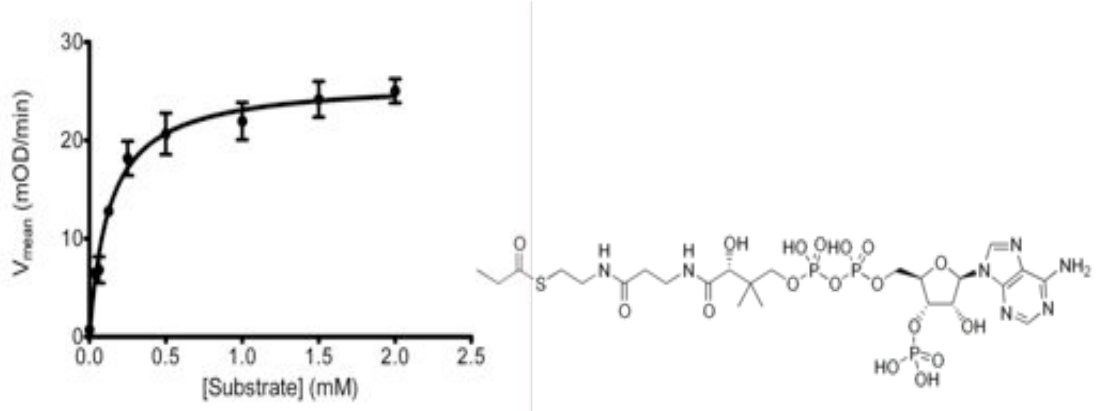

Figure 3-2. Carboxylase activity of AccA3-AccD5. AccA3-AccD5 assembled *in vitro* exhibits efficient carboxylase activity on propionyl-CoA (chemical structure shown on the right). Mean velocity of the reaction has been plotted as a function of the concentration of propionyl-CoA; error bars have been calculated based on data from three independent experiments.

3.1.1.3 Crystallisation

Pre-crystallisation tests identified the optimal concentration of AccA3-AccD5 for crystallisation as 4 mg/ml. In the crystallisation (sitting-drop vapour diffusion) trials that AccA3-AccD5 was subjected to, crystals of different morphologies appeared in several conditions in a duration of two to three weeks.

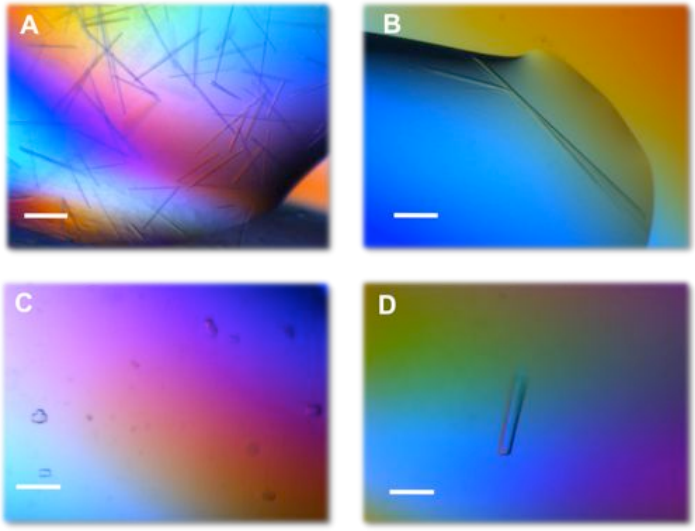

Figure 3-3. Crystallisation of the putative AccA3-AccD5 complex. (A) Short needles that were grown in 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 6.5, 25% (w/v) PEG 3350. **(B)** Long needles that were grown in 0.1 M MES pH 6.5, 30% (w/v) PEG 4000. **(C)** Small cuboidal crystals that appeared with 0.1 M HEPES pH 7.0, 20% (w/v) PEG 8000. **(D)** Rod-shaped crystal that was grown in 0.1 M HEPES-sodium pH 7.5, 1.6 M ammonium sulphate, 2% (w/v) PEG 1000. The white bars indicate a length of 50 µm.

Crystals were briefly soaked in suitable cryoprotectants (mother liquor containing 15% (v/v) glycerol additionally) and flash-cooled in liquid nitrogen until they were tested for X-ray diffraction. Crystals of type B [Figure 3-3] were very fragile, so only short fragments of the long needles were mounted, cryoprotected and flash-cooled.

3.1.1.4 X-ray data collection, structure solution and refinement

Crystals of the putative AccA3-AccD5 complex were tested at the ID23-2 beamline of the ESRF, Grenoble; only the crystal (D) was of good diffraction quality. Diffraction data were collected to a maximum resolution of 2.49 Å at a detector distance of 306 mm; the data were indexed, scaled and merged using *iMOSFLM* (Battye*, et al.*, 2011) and *SCALA* (Evans, 2006) [Table 3-1]. The crystal belonged to the orthorhombic space group $P2_12_12_1$, with a space group confidence of 0.925, Laue

group confidence of 0.987 and systematic absence probability of 0.950, as estimated by the program *POINTLESS* (Collaborative Computational Project, 1994). The structure was solved by molecular replacement with *Phaser* (McCoy*, et al.*, 2007) using the crystallographic structure of a monomer of AccD5 (PDB ID: 2BZR) as the search model. The emergence of unbiased electron density features indicated a correct molecular replacement solution. However, it was observed that the data corresponded to protomers of only AccD5; unfortunately, even after several rounds of maximum-likelihood refinement using *REFMAC5* (Murshudov*, et al.*, 1997), positive electron density features that could accommodate protomers of AccA3 in the crystal structure did not appear.

Table 3-1. Crystallographic statistics of AccD5

Crystallographic parameter	**Value**
Data collection and processing	
Wavelength [Å]	0.8726
Resolution range [Å]	53.53 – 2.49 (2.63 – 2.49)
Space group	$P2_12_12_1$
Unit cell dimensions [Å]	175.8, 218.8, 258.9
Average mosaicity [°]	0.33
Number of measured reflections	2164770 (282860)
Number of unique reflections	329407 (44209)
Multiplicity	6.6 (6.4)
Mean $I/\sigma(I)$	6.7 (1.9)
Completeness [%]	95.3 (88.4)
[1]R_{sym}	0.06 (0.37)
Values in parentheses correspond to the highest resolution shell.	
Refinement	
[2]R_{conv} / [3]R_{free}	0.204 / 0.257
Average B-factor [Å^2]	27.93
Root mean square deviations	
Bond lengths [Å]	0.016
Bond angles [°]	1.719
Ramachandran plot analysis	
Favoured [%]	94.8
Additionally allowed [%]	5.2

[1] $Rsym = \sum_h \sum_j \left| I_{h,j} - \langle I_h \rangle \right| \Big/ \sum_h \sum_j I_{h,j}$ where $I_{h,j}$ is the intensity of the j^{th} observation of unique reflection h.

[2] $Rconv = \sum_h \left\| Fo_h \right| - \left| Fc_h \right\| \Big/ \sum_h \left| Fo_h \right|$ where Fo_h and Fc_h are the observed and calculated structure factor amplitudes for reflection h.

[3] R_{free} is equivalent to R_{conv}, but is calculated using a 5% disjoint set of reflections excluded from the maximum likelihood refinement stages.

3.1.1.5 Crystal structure of AccD5

The crystal structure of AccD5 (in a space group different from $P2_12_12_1$) has previously been characterised (Holton*, et al.*, 2006, Lin*, et al.*, 2006); so only a brief description of the structure will be provided in the following sub-sections. A more detailed comparative analysis of the structural features of AccD5 with those of its structural relatives will be presented in sub-chapter 3.3. It is noteworthy that the structure of AccD5 determined in this study is either of moderately higher resolution or is more complete than each of the AccD5 structural models that have been described earlier (Holton*, et al.*, 2006, Lin*, et al.*, 2006).

3.1.1.5.1 Overall molecular architecture

The asymmetric unit of the crystal consists of 12 monomers of AccD5 related to each other by operations of crystallographic and non-crystallographic symmetry. These monomers are arranged in the form of two homohexamers related by symmetry arising from crystallographic packing. The biological assembly of the protein is a 350-kDa hexamer, similar to the CT domain of *Streptomyces coelicolor* (Diacovich*, et al.*, 2004) and the core catalytic 12S domain of *Propionibacterium shermanii* transcarboxylase (Hall*, et al.*, 2003). The AccD5 homohexamer is a roughly donut-shaped structure, made of two layers of tightly stacked trimers A-B-C and D-E-F [Figure 3-4], giving rise to 32-fold symmetry with a 3-fold axis perpendicular to a 2-fold axis.

The structural differences observed among the six monomers are negligible (a root mean square deviation of approximately 0.35 Å over 520 equivalent Cα atoms in each pair). Each monomer measures a length of approximately 80 Å in the longest dimension and is composed of two subdomains – the N-terminal and C-terminal subdomains – that possess the conventional crotonase fold [Figure 3-4B]. One loop region in each monomer (residues 470 to 489) could not be modeled with confidence because of the lack of corresponding positive electron density, presumably due to a high degree of conformational flexibility.

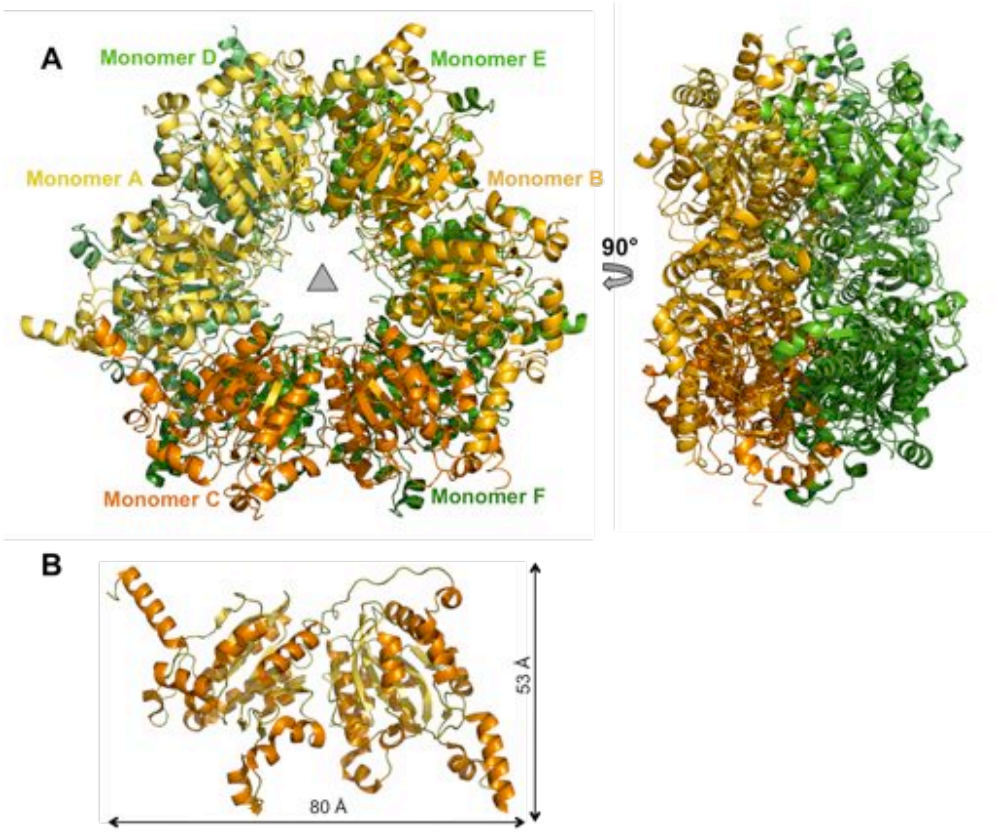

Figure 3-4. The crystal structure of AccD5. (A) The ribbon diagram of the AccD5 hexamer, arranged as two rings of trimers. The two rings are coloured in shades of orange and green respectively. The grey triangle indicates the three-fold axis of symmetry that associates monomers A, B and C as well as D, E and F. The hexamer is approximately 145 Å long in its largest dimension and the central pore is approximately 20 Å in diameter. **(B)** The structure of one monomer of AccD5, with helices, strands and loops coloured orange, yellow and olive respectively. The monomer is composed of N-terminal and C-terminal subdomains that interact strongly with the C-terminal and N-terminal subdomains respectively of the adjacently placed monomer in the other trimeric ring.

Buried surface area refers to the solvent-accessible surface area (in square Å) of monomeric units that becomes solvent-inaccessible upon the corresponding assembly formation and is therefore, an indicator of the strength of the association. Three-dimensional assembly and interface analysis performed using the *PDBePISA* tool (Krissinel & Henrick, 2007) revealed that the fraction of buried surface area resulting from the formation of units A-D, B-E and C-F are significantly higher than that in the cases of A-B, B-C or C-A [Table 3-2].

Table 3-2. Interface analysis of AccD5 protomers

No. of monomers	Composition	Total surface area [Å^2]	Buried surface area [Å^2]	Fraction buried [%]
6	ABCDEF	98050	44730	45.62
3	ABC	63390	8250	13.01
3	DEF	62630	8510	13.59
2	AD	38820	9270	23.88
2	BE	37940	9210	24.28
2	CF	38340	9190	23.97
2	AB	47821	1378	2.88
2	BC	47278	1372	2.90
2	CA	48171	1374	2.85
2	DE	47425	1403	2.96
2	EF	47407	1454	3.07
2	FD	47448	1399	2.95

The intra-dimeric interactions between monomeric members of the dimer pairs A-D, B-C and E-F are likely stronger than the inter-dimeric interactions (AD-BE, BE-CF, CF-AD) and the intra-trimeric associations within A-B-C and D-E-F. Only values corresponding to the A-B-C-D-E-F hexamer have been listed in the table above; nevertheless, nearly identical interface characteristics were observed for the G-H-I-J-K-L hexamer.

3.1.1.5.2 The active site

The active site of AccD5, as in other CT subunits characterised so far, lies at the conserved dimeric interface and is composed of two pairs of oxyanion-stabilizers: Gly 193 - Gly 194 (of monomer 1) and Gly 434' - Ala 435' (of monomer 2) [Figure 3-5A]. By analogy with other CT subunits (Hall*, et al.*, 2003, Diacovich*, et al.*, 2004), the backbone amide groups of the former pair might be responsible for making hydrogen bonds with the carbonyl group of the acyl-CoA substrate; the amide groups of the latter pair might be involved in forming hydrogen bonds with the carbonyl group of the carboxybiotin transported into the β-subunit active site by the BCCP domain of the α-subunit. In this way, residues from both the monomers of the β-subunit contribute to the stable binding of the substrate, the cofactor and probably the reaction intermediates, thereby facilitating the catalytic mechanism. Such an arrangement emphasizes the indispensability of the formation and maintenance of the dimeric interface for successful carboxyl moiety transfer to occur. The conserved

residues Gly 433', Tyr 432', and Ala 431' seem to be suitably positioned to allow a certain degree of flexibility in order to accommodate the substrate and reaction intermediates in the active site of the enzyme. The tentative substrate-binding pocket lies at the dimeric interface, adjacent to the catalytic site described above. Major contribution to the substrate-binding pocket comes from monomer 1 [Figure 3-5B]. Fully/partially conserved residues like Ala 153, Ala 155, Ile 157, Leu 164, Tyr 167, and Phe 120 are strategically arranged to give rise to a predominantly hydrophobic patch that could bind the acyl-CoA substrate molecule.

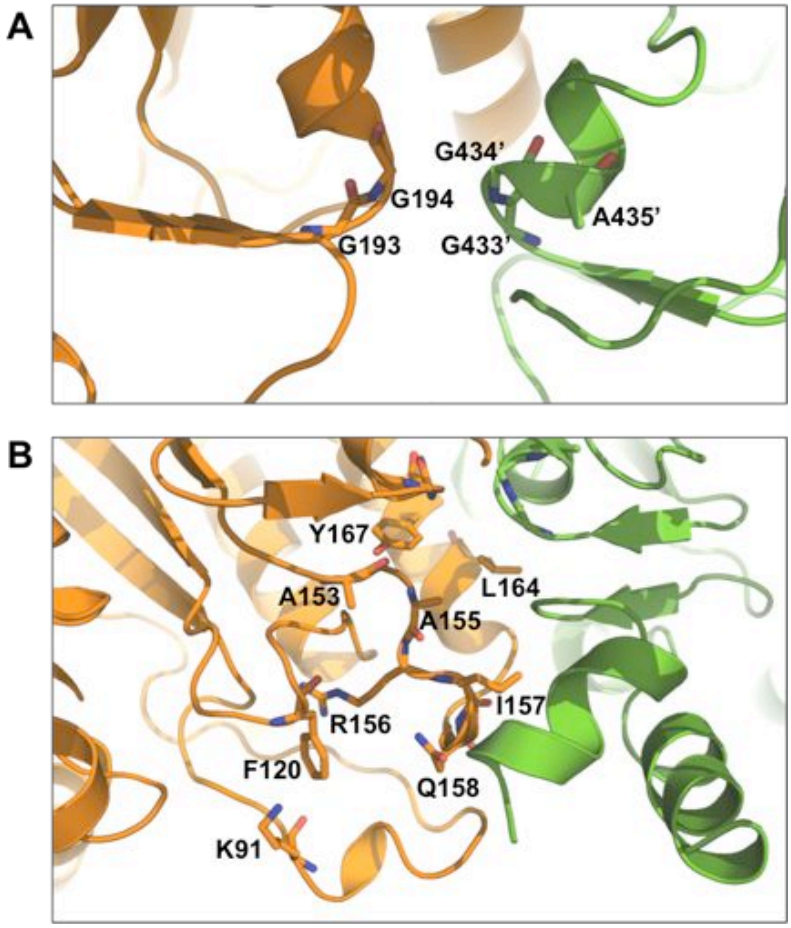

Figure 3-5. The active site of AccD5. **(A)** Secondary structural elements of monomers A and D are coloured orange and green respectively. Conserved residues that are integral to catalytic function are shown in the 'stick' mode; carbon, nitrogen and oxygen atoms are coloured green, blue and red respectively. **(B)** Fully/partially conserved residues that are most likely to have a role in substrate-binding and stabilization are represented as sticks; carbon, nitrogen and oxygen atoms are coloured orange/green, blue and red respectively. The residues described in (A) are seen in the top part of the figure.

The partially conserved polar residues Lys 91, Arg 156, and Gln 158 might utilize their side chains to stabilize different parts of the substrate molecule. Certain residues

from monomer 2 that could potentially be involved in binding of the bisphosphate group of the substrate through main-chain as well as side-chain interactions could not be modeled with confidence, as they are part of a flexible loop (residues 470 – 489). It is highly likely that the conformational mobility of this loop will be relatively restricted in the event of substrate-binding. It is interesting to note that Lys 91, that is distant from the active site in terms of primary sequence and is closer to the N-terminal of monomer 1, might get suitably positioned to act in conjunction with residues of the C-terminal subdomain of monomer 2 to bring about efficient substrate-binding and stabilization. The cofactor-binding pocket is located in close vicinity to the key catalytic residues. The binding of biotin is expected to be carried out predominantly by residues of monomer 2 [Figure 3-6]. The substrate and cofactor should therefore reach the catalytic centre *via* opposite sides of the dimeric interface.

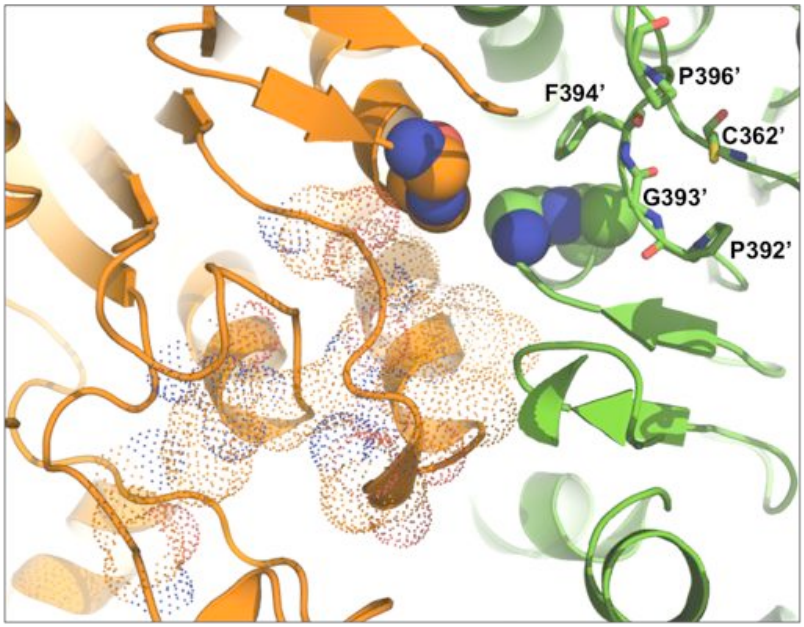

Figure 3-6. The cofactor-binding site of AccD5. The residues responsible for cofactor-binding are predominantly from monomer 2 and are represented here in the 'stick' mode with carbon, nitrogen, oxygen and sulphur atoms coloured green, blue, red, and yellow respectively. The catalytic site residues from both the monomers are shown as spheres with colouring scheme as in Figure 3-5. The tentative substrate-binding residues from monomer 1 are shown as dots for reference.

Similar to the cases of *S. coelicolor* PccB and *P. shermanii* transcarboxylase 12S domain, the conserved residues Gly 393', Phe 394', and Pro 396' give rise to a hydrophobic environment suitable for carboxybiotin-AccD5 interaction. This arrangement favours the precise binding of carboxybiotin in such a fashion that the

ureido moiety of biotin points directly into the oxyanion hole generated by residues Gly 433' and Gly 434'. The crystal structure of AccD5 and comparison with structurally related homologs have helped reiterate that the modes of binding of acyl-CoA substrate and carboxybiotin should be conserved and ubiquitous among biotin-dependent carboxyltransferases, wherein the dimeric, di-domain associations are essential for activity.

3.1.1.6 Electron microscopy of AccA3-AccD5

The AccA3-AccD5 complex was subjected to preliminary investigation by negative staining electron microscopy (in collaboration with the Sachse group, EMBL Heidelberg). At all concentrations of proteins tested, no clear uniformity/pattern was observed in the shapes of the particles [Figure 3-7]. The incubation of AccA3-AccD5 with the substrate analog CoA (at a molar concentration thrice as that of AccA3-AccD5) did not lead to improvement in sample quality.

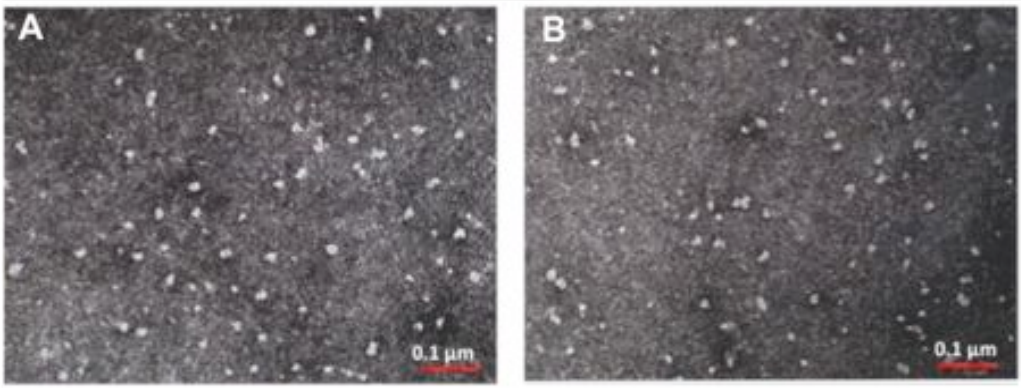

Figure 3-7. Negative staining electron microscopy of AccA3-AccD5. Negative staining electron microscopy images of AccA3-AccD5 in the **(A)** absence of any ligand **(B)** presence of the substrate analog CoA. The red bars correspond to a length of 0.1 µm.

Acquisition of useful electron microscopy data was complicated by the fact that the AccA3-AccD5 complex was conformationally heterogeneous, as observed by visual analysis of the images. Large variations in the dimensions of the particles underlined the possibilities of sample polydispersity and non-native homotypic interactions (facilitated by the inherently flexible BCCP domain) between AccA3 protomers of adjacently located complexes.

3.1.1.7 Overexpression in *E. coli*

It was reasoned that contamination from endogenous host proteins could contribute to the biochemical heterogeneity of the AccA3-AccD5 complex. Peptide mass fingerprinting analysis (Proteomics Core Facility, EMBL Heidelberg) of the samples indeed highlighted the presence of, along with a few host chaperone proteins, the PCC α- and β-subunits (with approximately 55% sequence coverage in each case) of *M. smegmatis* in preparations of *M. tuberculosis* AccA3-AccD5. The amino acid sequences of the major *M. smegmatis* 'contaminant' proteins are provided in the following box; the peptides identified with confidence by mass spectrometry are highlighted in red.

```
>gi|118471103|ref|YP_890604.1| propionyl-CoA carboxylase
subunit beta [Mycobacterium smegmatis str. MC2 155]

MTNKTTAELL AELREKLELA KEPGGEKAVA KREKKGIPSA RARINALLDP
GSFIEIGALA KTPGDPNALY GDGVVTGRGT IDGRPVGVFS HDQTVFQGSV
GEMFGRKVAR LMEWVAMVGC PIIGINDSAG ARIQDAVTSL AWYAELGRRH
EMLRGLVPEI SLIFGKCAGG AVYSPIQTDL LVAVRDQGYM FITGPDVIKD
VTGEDVTFDE LGGADEQAKR GNIHKVVNSE AEAYQYVRDY LSFLPSNHFD
NPPIVNPGME PEITPHDLEL DSIVPDADNM AYDMHEILLR IFDDGDVFEI
AEQRGPAMIT AFARVDGHPV GVIANQPMVL SGAIDNEASD KAASFIRFCD
SYNLPLVFVV DTPGAMPGVA EEKGGIIKRG GRFFNAIVEA DVPKVTVIIR
KAYGGGYAVM GSKQLSADLN FAWPTARIAV IGAEGAAQLL VKRFPDPNAP
EVQKIRDDFI EGYNLNMATP WIAAERGYID AVIQPHETRL LLRKSLRLLR
DKQNGPKVQR KHGLLPL


>gi|118469108|ref|YP_886179.1| acetyl-/propionyl-coenzyme A
carboxylase subunit alpha [Mycobacterium smegmatis str. MC2
155]

MANHASSKIS KVLVANRGEI AVRVIRAAKD AGLASVAVYA EPDADAPHVR
LADEAFALGG QTSAESYLVF EKILDAAEKS GANAIHPGYG FLSENADFAQ
AVIDAGLIWI GPSPQSIRDL GDKVTARHIA ARAKAPLVPG TPDPVKDADE
VVAFAKEHGV PVAIKAAFGG GGRGMKVART LEEIPELFES ATREAIAAFG
RGECFVERYL DKPRHVEAQV IADQHGNVVV AGTRDCSLQR RFQKLVEEAP
APFLTDAQRK EIHESAKRIC KEAGYYGAGT VEYLVGQDGL ISFLEVNTRL
QVEHPVTEET SGIDLVRQQF KIANGEPLDI TEDPTPRGHS FEFRINGEDA
GRGFLPAPGP VTKFVAPTGP GVRMDSGVET GSVIGGQFDS MLAKLIVTGA
TREEALERSR RALAEFTVEG LATVIPFHRA VVSDPAFIGD GEKFDVHTRW
IETEWNNTVE PFTGGDPIEE EDTVPRQTVV VEVGGRRLEV SLPGDLAIGG
GGGAAAPGVV RKKPKPRKRG GGGAKAASGD AVTAPMQGTV VKVAVEEGQE
VSAGDLVVVL EAMKMENPVT AHKDGTITGL AVEAGAAITQ GTVIAEIK
```

Sample biochemical heterogeneity resulting from using a strain of *M. smegmatis* as the expression host could be unfavourable for crystallisation of the AccA3-AccD5 complex. In order to overcome this problem, an expression strain of *E. coli* (BL21 star

(DE3) pRare2) was chosen as a far-native host for the heterologous production of the proteins AccA3 and AccD5 using constructs M02 and M08 respectively [Table 2-2]; high levels of overexpression could be achieved (AccA3: 5 – 10 mg per litre of culture; AccD5: 15 – 20 mg per litre of culture) by induction with IPTG. As observed with the *M. smegmatis* expression system, AccD5 eluted as a single symmetrical peak corresponding to a hexameric quaternary structure [Figure 3-8]. AccA3, on the other hand, eluted as a mixture of oligomers (hexamers possibly being the predominant species), in addition to forming very high-molecular weight aggregates.

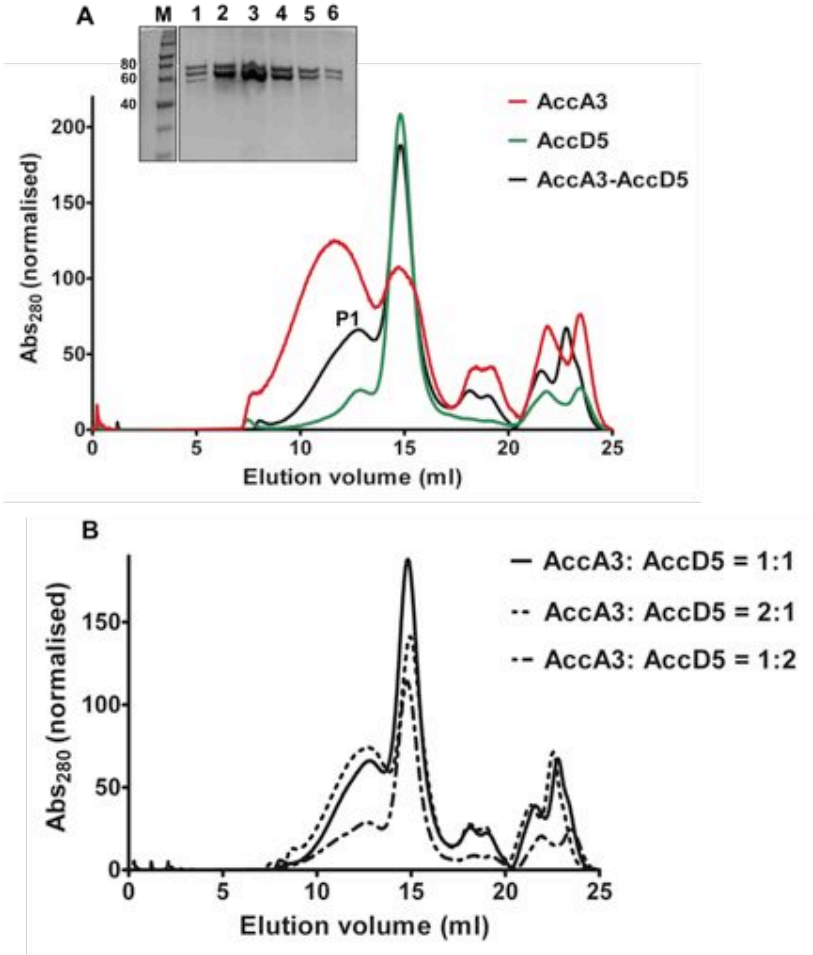

Figure 3-8. Overexpression in *E. coli* and purification of AccA3 and AccD5. (A) SEC elution profiles of AccA3, AccD5 and the unstable AccA3-AccD5 complex on Superose™ 6 10/300 GL. Molecular weight standards and elution volumes (ml) in brackets: 670 kDa (12.1), 158 kDa (15.5). **Inset:** Peak analysis of AccA3-AccD5 complex by SDS-PAGE (12%); lanes 1 – 6: equidistant fractions corresponding to the leading 'shoulder' peak P1 of the AccA3-AccD5 chromatogram; lane M: protein marker with standard molecular weights (in kDa). **(B)** SEC elution profiles of the AccA3-AccD5 complex following mixing of its components in different molar ratios.

Incubation of purified samples of hexameric AccA3 and AccD5 together in equimolar concentrations resulted in the formation of a high-molecular weight complex observed as a broad, asymmetrical peak P1 in the SEC profile [Figure 3-8A]. This complex was unstable and composed of more than one oligomeric state in dynamic equilibrium with each other. Incubation of AccA3 and AccD5 in different stoichiometric ratios also did not favour the formation of a stable complex suitable for crystallisation trials [Figure 3-8B].

3.1.1.8 Co-expression

As purification of individual components and mixing did not result in successful *in vitro* reconstitution, co-expression was attempted as a strategy for the *in vivo* reconstitution of the AccA3-AccD5 complex. For this, a series of monocistronic, IPTG-inducible constructs were made using vector backbones with compatible origins of replication [Table 3-3]; in each pair tested, one of the components was expressed in tandem with a polyhistidine tag that could facilitate isolation of the component together with its interacting partner by imidazole elution.

Table 3-3. Constructs used for co-expression of PCC components

S. No.	**Construct 1** (with N-ter His tag)	**Construct 2** (with no tag)
1	*accD5* in pETM-11	*accA3* in pCDF-13
2	*accA3* in pETM-11	*accD5* in pCDF-13
3	*accD5* in pETM-11	*accE5* in pCDF-13
4	*accA3* in pETM-11	*accE5* in pCDF-13

For each pair of constructs, overexpression trials were performed wherein cells from induced and uninduced cultures were lysed by sonication and the supernatant and pellet fractions of the lysates were sampled. These experiments showed that there were unequal levels of overexpression of the PCC components [Figure 3-9]. In the cases of construct pairs (1) and (2), the polyhistidine-tagged member was preferentially highly expressed, while the non-tagged protein was expressed only in lower levels, as observed by SDS-PAGE; protein identity corresponding to the labeled bands [Figure 3-9A, B] was confirmed by peptide mass fingerprinting

analyses. In the cases of construct pairs (3) and (4), overexpression of the ε-subunit AccE5 could not be achieved under any conditions (of inducer concentration and post-induction incubation temperature) tested; only the proteins AccA3 and AccD5 were overexpressed.

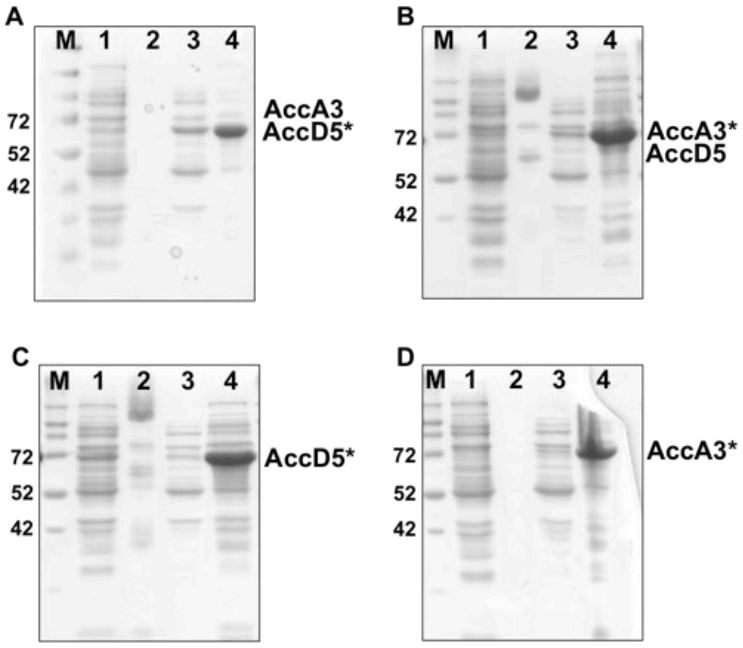

Figure 3-9. Co-expression trials of *M. tuberculosis* PCC components. In gel images **(A)**, **(B)**, **(C)** and **(D)**, lanes (1) and (2): cell pellet and cell supernatant respectively from non-induced cultures; lanes (3) and (4): cell pellet and cell supernatant respectively from induced cultures; lane M: protein marker with standard molecular weights (in kDa). Approximate theoretical molecular weights of non-tagged versions of the proteins are as follows: AccA3: 64 kDa, AccD5: 59 kDa, AccE5: 19 kDa. The polyhistidine-tagged protein of each pair is marked with a (*).

Upon scale-up of the expression and protein purification, a complex of AccA3-AccD5 could not be observed in the SEC profiles [Figure 3-10]. The elution profiles of AccA3-AccD5 co-expressed were similar to those of AccA3 or AccD5 purified individually, depending on which of the proteins was polyhistidine-tagged. In other words, a high-molecular weight complex suitable for crystallisation could not be obtained by the co-expression strategy tested, as comparable levels of overexpression of the components could not be accomplished.

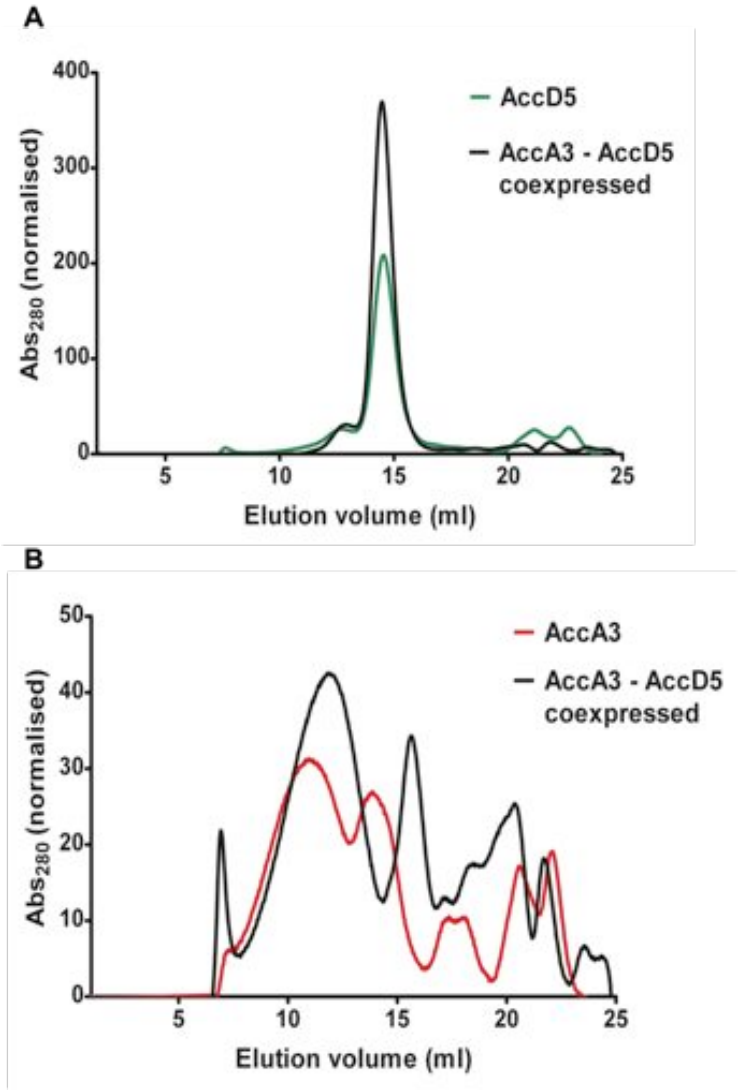

Figure 3-10. SEC profiles of co-expressed *M. tuberculosis* PCC components. (A) The elution profile (on Superose 6) of co-expressed AccA3-AccD5 (polyhistidine tag on AccD5) overlaid with that of AccD5. **(B)** The elution profile (on Superose 6) of co-expressed AccA3-AccD5 (polyhistidine tag on AccA3) superposed with that of AccA3. Elution volumes of molecular weight standards are the same as in Figure 3-8.

3.1.1.9 Attempts to improve AccA3 behaviour

The inability of AccA3 and AccD5 to assemble into a stable complex could have stemmed from the fact that the preparation of AccA3 was a heterogeneous mixture of several oligomeric states that existed in dynamic equilibrium with each other, presumably because of the inherent conformational flexibility of its BCCP domain. Several attempts were made to obtain full-length AccA3 in a stable, monodisperse form. Repeated thermofluor experiments aimed at the identification of optimal buffer conditions for stable AccA3 preparation did not provide interpretable melting curve data. So, a number of AccA3 purification trials (using construct M02) were performed

with varying buffer conditions, in a low-throughput fashion. The effects of different pH conditions (pH 6.5, 7.0, 7.5, 8.0 and 8.5, at a fixed salt concentration of 200 mM) and different concentrations of salt (100 mM, 300 mM) on the oligomerisation behaviour of AccA3 were analysed. However, these experiments did not help arrive at an absolutely optimal buffer condition for AccA3 monodispersity [Figure 3-11]. At pH 6.5 and 8.5, the protein formed dense, cloudy precipitate in the initial stages of the purification. In all other conditions tested, AccA3 existed in multiple oligomeric states in dynamic equilibrium with each other, in addition to forming heavy aggregates that were lost in the void volume.

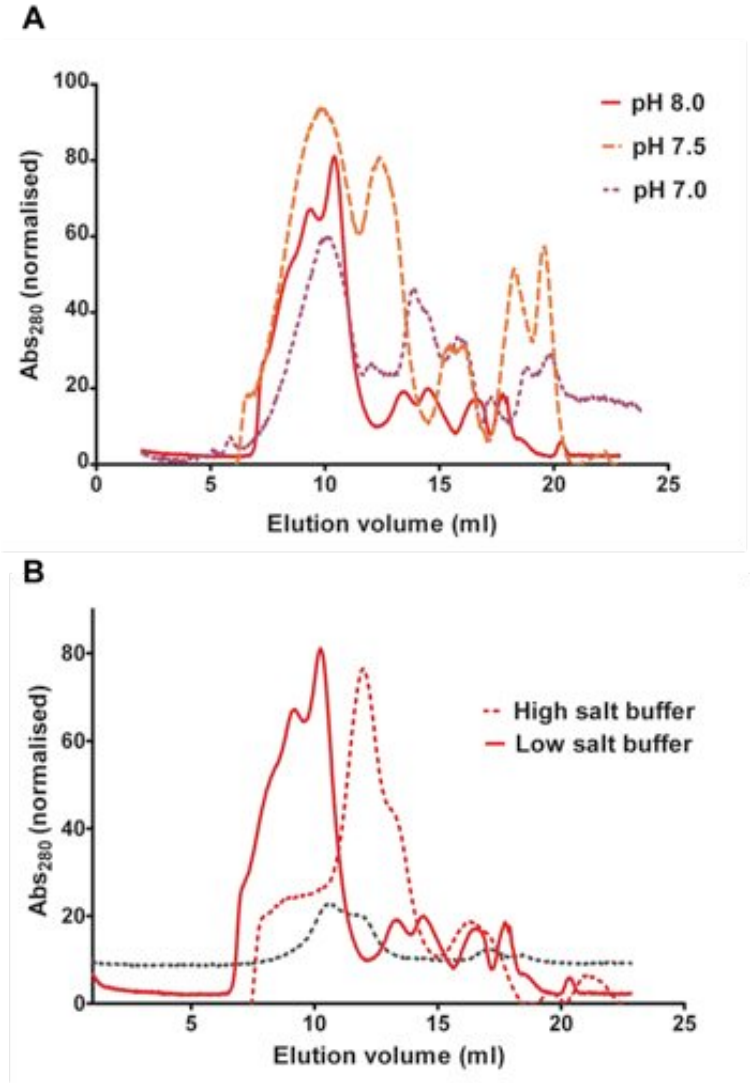

Figure 3-11. Purification of AccA3 under different buffer conditions. (A) The elution profiles of AccA3 purified under different conditions of pH. **(B)** The elution profiles of AccA3 purified under different conditions of ionic strength; high salt: 300 mM KCl, low salt: 100 mM KCl). The black dotted chromatogram corresponds to the dominant peak fraction of the 'high salt buffer' trial reloaded on Superose 6; multiple oligomeric states

existed in dynamic equilibrium with each other. Elution volumes of molecular weight standards are the same as in previous figures.

Truncated versions of AccA3, with the putative BCCP domain deleted, were sub-cloned to obtain AccA3 in a form devoid of the ability to make undesirable homotypic interactions that were presumably unfavourable for complex formation. In order not to disrupt secondary structural elements, three ΔBCCP variants – AccA3 ΔBCCP1, ΔBCCP2 and ΔBCCP3 (constructs M24, M25 and M26 respectively, Table 2-2) of different lengths (1 – 454 aa, 1 – 471 aa and 1 – 480 aa) were made based on multiple sequence alignment analysis and a homology model of AccA3 constructed by an automated process (*SWISS-MODEL*, (Arnold*, et al.*, 2006)) using the crystal structure of the α-subunit of an unnatural chimeric PCC (PDB ID: 3N6R) as template.

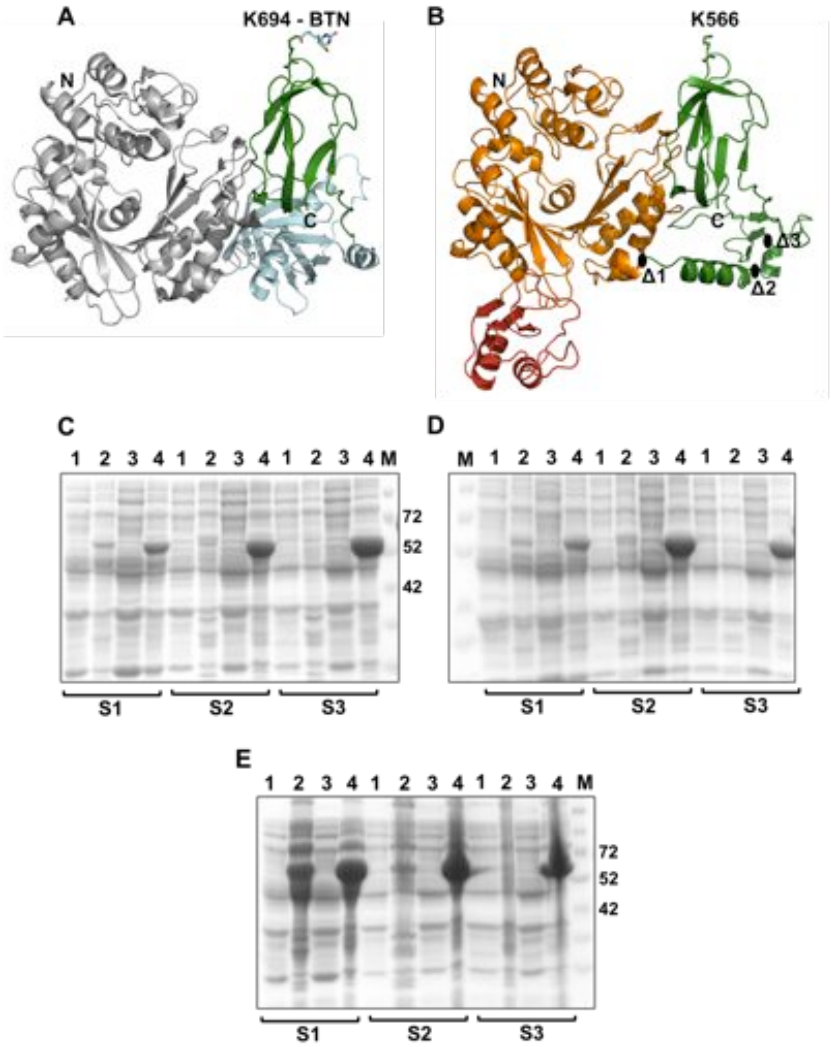

Figure 3-12. Overexpression of AccA3 ΔBCCP. (A) Ribbon representation of the crystal structure of the α-subunit of *Ruegeria pomeroyi* (PDB ID: 3N6R-A) used as a

template for AccA3 homology modeling. The core BC domain, the BCCP domain and the unique BT domain (BC-CT interaction domain, absent in the model of AccA3) are coloured grey, green and cyan respectively. Biotin (BTN) and the key lysine residue to which it is covalently linked are shown in 'stick' representation with carbon, nitrogen and oxygen atoms coloured green/cyan, blue and red respectively. The N- and C-termini of the protein are indicated. **(B)** The homology model of AccA3 with the putative core BC domain and BCCP domain coloured orange and green respectively. The region highlighted in red is a part of the BC domain whose counterpart is unmodeled in the *R. pomeroyi* structure, presumably because of lack of conformational order. The conserved lysine that is expected to act as the site of biotinylation is shown as sticks with colouring scheme as in (A). The C-termini of the constructs AccA3 ΔBCCP1, ΔBCCP2, and ΔBCCP3 are indicated as black ovals Δ1, Δ2, and Δ3 respectively. **(C), (D) and (F)** Overexpression trials with constructs AccA3 ΔBCCP1, ΔBCCP2, and ΔBCCP3 respectively. S1, S2 and S3 refer to expression trials in the expression strains *E. coli* BL21 (DE3) CodonPlus-RP, BL21 star (DE3), and BL21 star (DE3) pRare2 respectively (Appendix I). Lanes 1 and 2: supernatant and cell pellet respectively of non-induced cultures; lanes 3 and 4: supernatant and cell pellet respectively of induced cultures; lanes M: protein marker with standard molecular weights (in kDa).

A number of conditions (concentration of inducer, post-induction incubation temperature and duration, expression strains) were tested for the overexpression of the ΔBCCP constructs. Although high levels of overexpression were achieved with all the constructs, the proteins were present in the insoluble pellet fraction of the cells [Figure 3-12]; optimal conditions for obtaining the proteins in soluble form for further analyses could not be identified.

3.1.1.10 The epsilon subunit AccE5

The ε-subunit AccE5 of *M. tuberculosis* PCC has been proposed to be responsible for the modulation of the catalytic activity of the complex (Gago*, et al.*, 2006, Oh*, et al.*, 2006). However, the mode and stoichiometry of interaction and the functional basis of the role of AccE5 in activity modulation are not known. In this study, attempts were made to express, purify and perform preliminary biophysical characterisation of AccE5.

3.1.1.10.1 Purification of AccE5

Efforts to overexpress AccE5 in *E. coli* with routinely used pET vectors and *M. tuberculosis* H37Rv genomic DNA as template were not successful. After repeated trials, AccE5 could be successfully overexpressed as a fusion protein, in combination with a solubility enhancer tag – the Z-tag or small ubiquitin-like modifier (SUMO-)

tag – in addition to a polyhistidine tag, using constructs M10 and M11 respectively [Table 2-2]. These constructs were prepared using a synthetic *accE5* gene in a commercial vector backbone (GeneArt) as the DNA template. In the synthetic gene, the codon usage had been adapted to the codon bias of *E. coli* genes and further optimization measures (removal of negative cis-acting sites such as splice sites, poly (A) signals, TATA-boxes, etc. that might adversely influence protein expression, adjustment of GC content to prolong mRNA life) to allow high and stable expression rates in *E. coli* had been undertaken.

Protocols for the affinity purification of AccE5-His_6-Z-tag and AccE5-His_6-SUMO-tag fusion proteins by imidazole-based elution were optimized [Figure 3-13]. AccE5 was separated from the respective solubility tag by site-specific cleavage using TEV protease. During the cleavage of the Z-tag, AccE5 formed a dense, insoluble precipitate; however, in the case of the SUMO-tag, the protein remained soluble after the removal of the tag and was used for further experiments. SEC experiments using Superdex 75 prep grade chromatography media indicated that AccE5 exists as a dimer in solution. However, the protein was very unstable and generally underwent degradation/precipitation (at room temperature and at 4 °C) in a few hours after preparation.

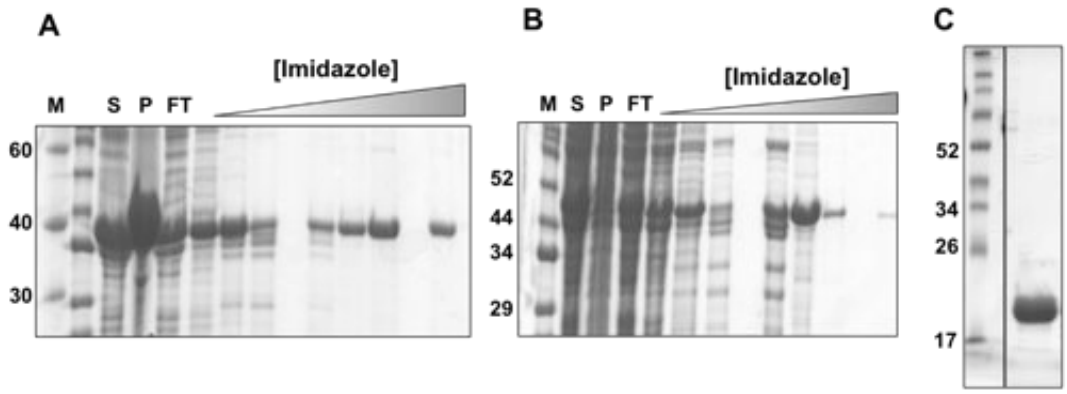

Figure 3-13. Purification of AccE5. SDS-PAGE (15%) analysis of **(A)** fractions from the affinity purification of AccE5-His_6-Z-tag fusion protein. **(B)** fractions from the affinity purification of AccE5-His_6-SUMO-tag fusion protein. **(C)** pure eluate of untagged AccE5 from an SEC experiment. Lanes M: protein marker with standard molecular weights (in kDa); lanes S: supernatant of cell lysate; lanes P: insoluble pellet of cell lysate; lanes FT: flow-through during loading of supernatant of cell lysate onto Ni-NTA resin.

3.1.1.10.2 AccE5 lacks secondary structure

In silico analysis of the primary structure of AccE5 indicated that more than 70% of the protein is prone to be intrinsically disordered [Figure 3-14 inset]. This prediction was indeed verified by circular dichroism (CD) spectroscopy measurements [Figure 3-14] performed on AccE5 at different concentrations and at two different temperatures (4 ºC and 20 ºC). In all cases tested, AccE5 exhibited CD spectra typical of randomly coiled proteins. A concentration-dependent effect on the values of molar ellipticity was observed; varying values of molar ellipticities at different concentrations suggested the possibility of protein aggregation under the experimental conditions. These experiments, nevertheless, indicated that the lack of secondary structure was protein concentration- and temperature-independent.

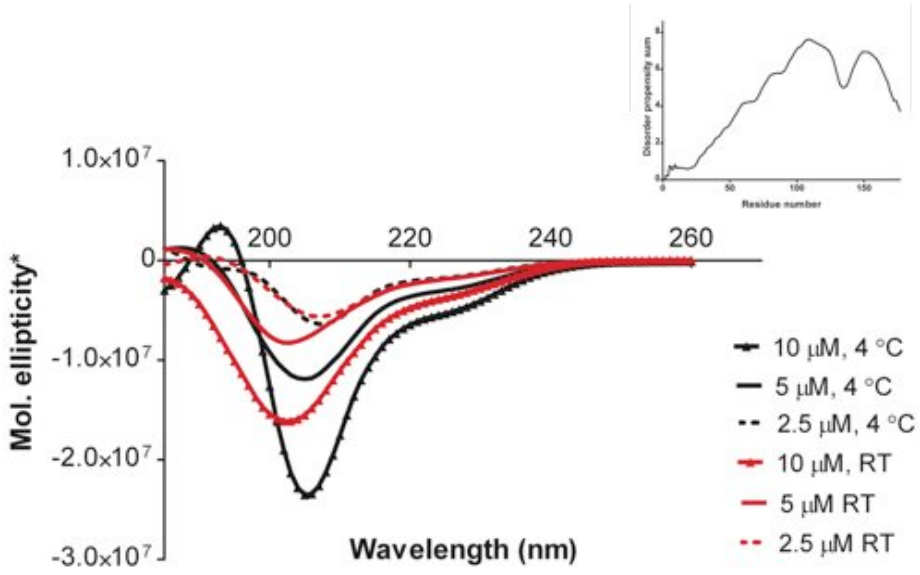

Figure 3-14. AccE5 is an intrinsically disordered protein. CD spectra of AccE5 (measured at 4 ºC and 20 ºC) at three different concentrations: 10 μM, 5 μM, and 2.5 μM. Mol. ellipticity*: Molar ellipticity in deg cm^2 $dmol^{-1}$. **Top right:** Intrinsic disorder propensity of AccE5, as predicted by the *GlobPlot 2* web service (Linding*, et al.*, 2003).

The intrinsically disordered nature of AccE5 explained the difficulties experienced in the overexpression, purification and handling of the protein. As a consequence of these technical constraints, AccE5 could not be subjected to further structural or functional characterisation during the study.

3.1.2 Discussion

In this part of the study, a dodecameric complex of the *M. tuberculosis* essential propionyl-CoA carboxylase components AccA3 and AccD5 was successfully reconstituted by overexpression in a suitably engineered expression strain of the near-native non-pathogenic host *M. smegmatis* (Noens*, et al.*, 2011). Unlike in the case of the *S. coelicolor* acetyl-CoA carboxylase enzyme, where interaction between the α-subunit AccA2 and β-subunit AccB occurs only in the presence of the ε-subunit AccE (Diacovich*, et al.*, 2002), the *M. tuberculosis* PCC components AccA3 and AccD5 are able to interact even in the absence of the ε-subunit AccE5.

Interestingly, the activity of *M. tuberculosis* AccA3-AccD5 (reconstituted as above) was found to be multiple folds higher than that of the *M. tuberculosis* AccA3-AccD5-AccE5 complex (reconstituted after overexpression in and purification from *E. coli*) studied earlier, wherein K_M = 0.24 ± 0.035 mM for propionyl-CoA (Gago*, et al.*, 2006). Based on the data published by Gago *et al.,* the V_{max} and k_{cat} of the *E. coli*-derived complex could be calculated to be as low as 1.38 μM min^{-1} mg^{-1} and 4.7x10^{-5} s^{-1} respectively, whereas the corresponding values for the *M. smegmatis*-derived complex reconstituted in this study are V_{max} = 1.33 ± 0.05 mM min^{-1} mg^{-1} and k_{cat} = 0.045 s^{-1}. In the study by Gago *et al*., the presence of AccE5 at a 5:1 ratio with respect to the concentrations of AccA3 or AccD5 was shown to enhance the catalytic activity at least 20-fold; hence the ε-subunit AccE5 was reported to be essential for maximal carboxylase activity of the enzyme and has been proposed to be responsible for inducing an enzyme conformational change that enhances the substrate's access to the active site or for altering the oligomeric composition of the AccA3-AccD5 complex to make it more functional (Gago*, et al.*, 2006). It is interesting to note that the complex reconstituted in this study could achieve such high level of activity even in the absence of AccE5. The only explanation that could be attributed to this high activity is that the usage of the near-native expression host ensures proper protein folding and processing, thereby leading to optimal catalytic efficiency. These results can be considered to serve as proof of principle that near-native expression hosts are usually most optimal for functional studies on proteins.

Crystallisation experiments using the AccA3-AccD5 complex, however, did not result in structural data of the complex; during the process of crystallisation, the core catalytic β-subunit AccD5, being an apparently more compactly packed component, had crystallized preferentially over molecules of AccA3-AccD5 or AccA3 alone. This result points to the possibility that the interaction between AccA3 and AccD5 is too weak and/or transient to give rise to a symmetrical crystal lattice arrangement. Given the molecular weight and the expected dimensions of the PCC complex and the predicted conformational flexibility of the α-subunit, the inability of AccA3-AccD5 to readily crystallize was not totally unexpected. During this study, C.S. Huang and co-workers solved the first crystal structure of an $\alpha_6\beta_6$ holoenzyme of PCC (Huang*, et al.*, 2010); this structure corresponds to an unnatural chimeric PCC composed of the PCC α-subunit from *Ruegeria pomeroyi* and the PCC β-subunit from *Roseobacter denitrificans*. The structure solution of this chimeric PCC was reportedly preceded by several technical hurdles like difficulties in native complex reconstitution, crystals with poor diffraction properties and/or perfect twinning (Huang*, et al.*, 2010), suggesting that these enzymes, in general, are problematic candidates for crystallography.

Members of the YCC family are generally conserved, with moderate to high levels of sequence and presumable structural homology, among mycobacterial species (Cole*, et al.*, 1998, Kapopoulou*, et al.*, 2011). As a direct consequence, samples of the *M. tuberculosis* PCC components were 'contaminated' with the endogenously expressed *M. smegmatis* counterparts of their interacting partners. Hence, crystallisation/electron microscopy analysis of the complex was further complicated by biochemical heterogeneity stemming from the usage of *M. smegmatis* as the expression host. The possibility of tackling this problem by means of using *M. smegmatis* knockout versions of the *accA3/accD5/accA3-accD5* genes as expression hosts was ruled out, as these genes are essential for mycobacterial viability (Sassetti*, et al.*, 2003), Elke Noens, personal communication).

Expression in *E. coli*, purification and mixing of proteins did not lead to effective reconstitution of the AccA3-AccD5 complex, primarily because of the inherent tendency of AccA3 to exist in multiple oligomeric forms. It is known that the

assembly of a carboxylase complex in a stable, monodisperse form could be highly dependent on pH (Ehebauer, *et al*., unpublished). Various standardization trials could not decisively identify conditions for the preparation of full-length AccA3 in a satisfactory form for crystallisation experiments. The core BC domain of the α-subunit might form a compact structure, but the BCCP domain is expected to be conformationally flexible as its role is to transport carboxybiotin and biotin to and back from the active site of the β-subunit. The structure of the *R. pomeroyi – R. denitrificans* PCC revealed that the distance between the active sites of the α- and β-subunits is approximately 55 Å, necessitating the translocation of the entire BCCP domain during catalysis (Huang*, et al.*, 2010). In an endeavour to obtain AccA3 with potentially restricted conformational flexibility, AccA3 ΔBCCP versions were constructed; however, conditions for overexpression in soluble forms remained elusive.

The third component of PCC - the ε-subunit - is unique to acyl-CoA carboxylases of actinobacteria, suggesting that this subunit was present in the common ancestor of all actinobacteria. The *M. tuberculosis* genome codes for only a single ε-subunit (AccE5), which is known to be essential for survival (Cole*, et al.*, 1998, Sassetti*, et al.*, 2003). The corresponding open reading frame has been placed in the list of genes (Marmiesse*, et al.*, 2004) that have been conserved by all mycobacterial species, including *M. leprae*, a subject of severe reductive evolution. Though these studies emphasize that AccE5 is responsible for an indispensable function, very little is known about the function and biophysical aspects of AccE5, a 'conserved hypothetical protein'. A couple of studies have shown that AccE5 is involved in the enhancement of the activity of the PCC complex (Gago*, et al.*, 2006, Oh*, et al.*, 2006). But the production of the protein, the stoichiometry of binding to AccA3 and AccD5, and the mechanism of activity enhancement have not been unambiguously described. In this study, considerable effort was invested in the overexpression and purification of AccE5. GlobPlot analysis suggested that over 70% of AccE5 could be intrinsically disordered. Furthermore, based on sequence, the instability index of AccE5 was calculated to be above 45.0 (Guruprasad*, et al.*, 1990) indicating a high level of *in vitro* instability. CD spectroscopy experiments confirmed the sequence-based prediction that AccE5 is an intrinsically disordered protein. Preliminary attempts to

characterise AccE5 by nuclear magnetic resonance (NMR) spectroscopy (in collaboration with Bernd Simon, EMBL Heidelberg) failed. This was because AccE5 required the presence of at least 5% (v/v) glycerol for its short-lived stability and such a buffer environment is inappropriate for NMR measurements. Initiatives were taken to overexpress AccE5 as a fusion protein with a green fluorescent protein (GFP) tag in *M. smegmatis*; but under standard conditions of culture density and at various concentrations of tetracycline inducer, overexpression of the fusion protein could not be achieved. In conclusion, detailed characterisation of AccE5 was hampered by serious difficulties in the production and handling of the protein.

3.2 Characterisation of the carboxyltransferase subunit of the long-chain acyl-CoA carboxylase of *Mycobacterium tuberculosis*

Mycolic acids, long-chain 2-alkyl, 3-hydroxy fatty acids present as esters of either trehalose (role in virulence) or arabinogalactan (role in lowering cell wall permeability) in the mycobacterial cell wall, are synthesized by the concerted action of two multi-component enzyme systems: fatty acid synthase I and fatty acid synthase II (Barry*, et al.*, 1998, Asselineau*, et al.*, 2002). Mycolic acids are of significant scientific interest, because they are unique to mycobacteria and mycolic acid metabolism is the only well-known target of the first-line antitubercular drug isoniazid (Barry*, et al.*, 1998, Daffe & Draper, 1998). Despite the importance of these complex biomolecules in virulence and antibiotic resistance and extensive biochemical and molecular genetic studies, the pathways of mycolic acid biosynthesis, maturation and transport are not completely understood. Based on genomic location and studies on homologous proteins from *M. smegmatis* and *Corynebacterium glutamicum*, the protein encoded by the *accD4* gene (*Rv3799c*) of *M. tuberculosis* is believed to be the carboxyltransferase subunit of an acyl-CoA carboxylase crucial for mycolic acid biosynthesis in *M. tuberculosis* (Trivedi*, et al.*, 2004, Portevin*, et al.*, 2005, Gago*, et al.*, 2011). This section of the chapter reports the characterisation of AccD4, the less-studied carboxyltransferase subunit of the *M. tuberculosis* long-chain acyl-CoA carboxylase. Efforts towards the reconstitution of the AccA3-AccD4 complex and towards the structural characterisation of AccD4 by X-ray crystallography and small angle scattering are described.

3.2.1 Results

3.2.1.1 Reconstitution of the AccA3-AccD4 complex

3.2.1.1.1 *In vitro* reconstitution

Using constructs M02 and M06 [Table 2-2], the proteins AccA3 and AccD4 could be overexpressed individually in expression strains of *E. coli* (BL21 star (DE3) pRare2

and BL21 (DE3) respectively). Typical overexpression experiments resulted in yields of 5 – 10 mg of AccA3 and 12 – 15 mg of AccD4 per litre of expression culture. During SEC experiments, AccA3 eluted as a mixture of oligomers, hexamers likely being the principal species, as was described in section 3.1.1.7. Surprisingly, AccD4 eluted as a homotrimer unlike other homohexameric actinobacterial CT subunits, including AccD5, described so far. Static light scattering analyses indicated a molecular weight of approximately 160 kDa [Figure 3-15B] that corresponds to a trimeric form of AccD4 (MW of a monomer: 56 kDa). The incubation of AccA3 and AccD4 together in equal stoichiometry resulted in the formation of a high-molecular weight AccA3-AccD4 complex that eluted as a broad, asymmetrical peak from Superose™ 6 10/300 GL in an SEC experiment [Figure 3-15]; the complex was unstable and hence could not be forwarded to crystallisation trials.

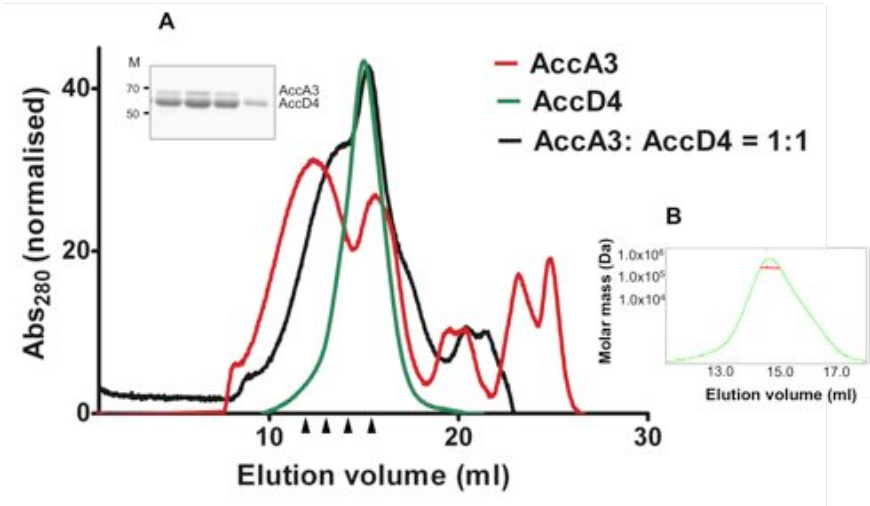

Figure 3-15. *In vitro* reconstitution of AccA3-AccD4. Size exclusion chromatography profiles of *M. tuberculosis* AccA3 and AccD4 (expressed individually in *E. coli*) overlaid with that of the AccA3-AccD4 complex obtained by mixing the subunits in equimolar concentrations. Molecular weight standards and elution volumes (ml) on Superose™ 6 10/300 GL in brackets: 670 kDa (12.1), 158 kDa (15.5). **Inset A:** SDS-PAGE analysis of fractions of the AccA3-AccD4 complex indicated by black arrows along the x-axis (in the same left-to-right order). **Inset B:** Mass distribution (molar mass as a function of elution volume) of AccD4 peak fractions, as determined by SLS.

3.2.1.1.2 Co-expression of AccA3-AccD4

Construct pairs M06-M17 and M02-M21 [Table 2-2], based on the compatible pETM11-pCDF13 vector backbones, were used for co-expression of AccA3-AccD4. In each pair, one of the members was expressed in tandem with a polyhistidine

affinity tag while the interacting partner was expressed in a non-tagged form, to enable imidazole-based elution of the complex putatively assembled *in vivo*. In either trial, AccD4 was overexpressed at high levels in the insoluble pellet fractions whereas AccA3 was expressed only at low levels in both cell supernatant and cell pellet (confirmed by peptide mass fingerprinting) [Figures 3-16A, B]. Upon scale-up of the overexpression and purification, the formation of a binary complex could not be observed. Another version of co-expressed AccA3-AccD4 material (AccA3 expressed together with a *Strep*-tag® and AccD4 expressed together with a polyhistidine tag) was prepared in collaboration with Lionex GmbH. However, in this preparation, the AccA3-AccD4 complex existed in several oligomeric states in dynamic equilibrium with each other [Figure 3-16C] and hence was not subjected to crystallisation screens.

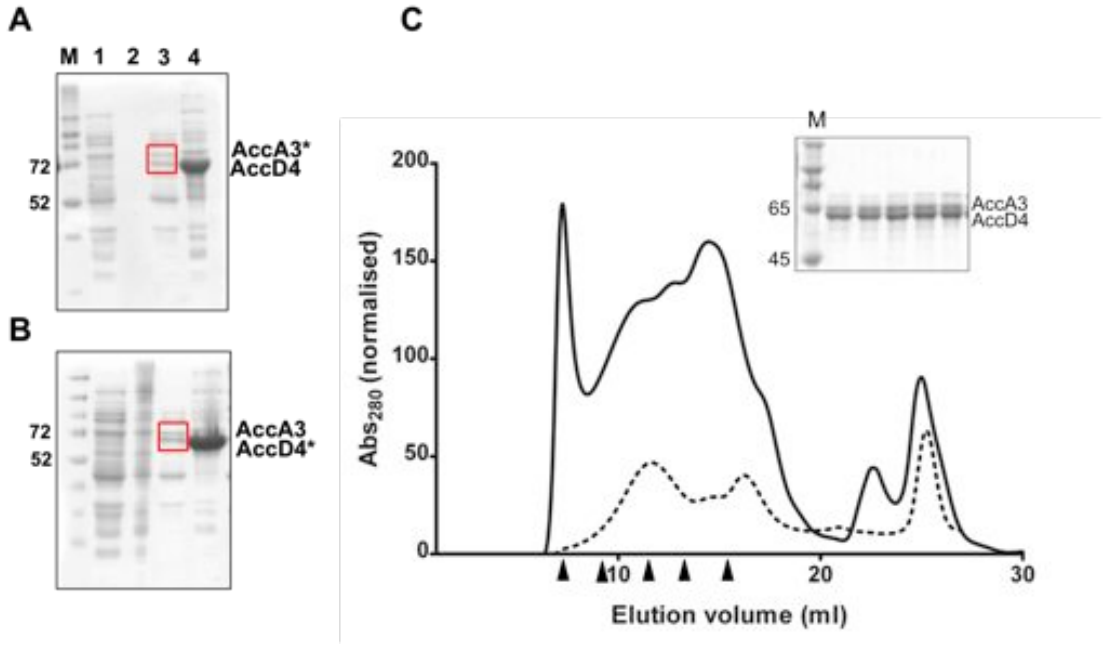

Figure 3-16. Co-expression of AccA3-AccD4. In gel images **(A)** and **(B)**, lanes (1) and (2): cell supernatant and cell pellet respectively from non-induced cultures; lanes (3) and (4): cell supernatant and cell pellet respectively from induced cultures; lane M: protein marker with standard molecular weights (in kDa). Protein bands of the expected molecular weights and whose identities were confirmed by peptide mass fingerprinting have been marked with red boxes. The polyhistidine-tagged protein of each pair has been marked with a (*). **(C)** Solid curve: SEC profile of the AccA3-AccD4 complex produced by co-expression (in collaboration with Lionex GmbH). Dashed curve: SEC profile of the 10-12 ml peak (from above) re-analysed on Superose™ 6 10/300 GL, indicating the existence of dynamic equilibrium among the different oligomeric states. Molecular weight standards and elution volumes (ml) in brackets: 670 kDa (12.1), 158 kDa (15.5); Void volume: 8.0 ml. **Inset:** SDS-PAGE analysis of fractions of the AccA3-AccD4 complex indicated by black arrows along the x-axis (in the same left-to-right order). Theoretical MW: AccD4: 56.6 kDa, AccD4+His tag: 59.9 kDa, AccA3: 63.7 kDa, AccA3+His tag: 67 kDa, AccA3+Strep tag: 65 kDa.

3.2.1.2 Carboxylase activity of AccA3-AccD4

The carboxylase activity of the AccA3-AccD4 reconstituted *in vitro* was tested on several acyl-CoA substrates of different acyl-chain lengths (butyryl-CoA [C4], lauroyl-CoA [C12], myristoyl-CoA [C14], and palmitoyl-CoA [C16]) by a coupled-enzymatic reaction-based assay that probes ATP hydrolysis spectrophotometrically at 340 nm (Diacovich, *et al*., 2002). The activity of the enzyme at different concentrations (approximately 0.2 μM – 2 μM) on a given concentration of butyryl-CoA (0.5 mM) is depicted in Figure 3-17; an enzyme concentration-dependent consumption of the substrate was evident.

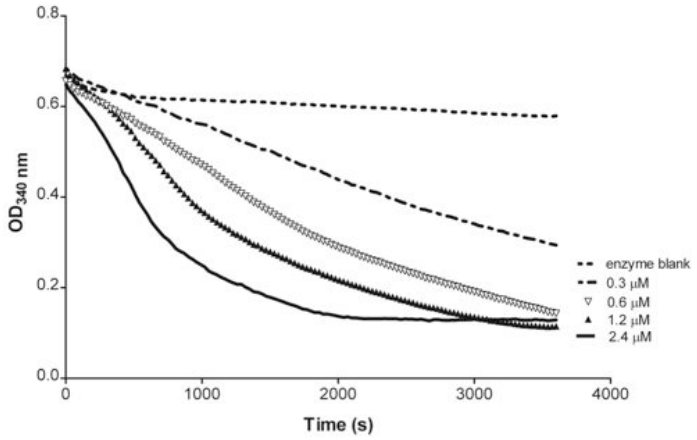

Figure 3-17. AccA3-AccD4 is an active carboxylase *in vitro*. The $OD_{340\ nm}$ of the carboxylase assay mixture, an indicator of the consumption of NADH (and hence of the acyl-CoA substrate, i.e., butyryl-CoA), plotted as a function of time. Concentration of butyryl-CoA in each reaction: 0.5 mM. Concentrations of AccA3-AccD4 enzyme used are mentioned in the legend above.

Similar enzyme-concentration dependent carboxylase effects were observed with other substrates (lauroyl-CoA, myristoyl-CoA and palmitoyl-CoA) of different acyl-CoA chain lengths. Therefore, these preliminary kinetic analyses, aimed at checking if the AccA3-AccD4 assembled *in vitro* is a catalytically active acyl-CoA carboxylase, were indeed affirmative.

3.2.1.3 AccD4: Crystallisation trials

Stable AccA3-AccD4 material suitable for crystallisation trials could not be obtained by the co-purification and co-expression options tested. However, given that stable and monodisperse AccD4 protein could be produced and it was of interest to generate high-resolution structural data on individual β-subunits to probe into the molecular

determinants of substrate specificity, AccD4 was subjected to screening for crystallisation hits. Extensive efforts were undertaken to accomplish the crystallisation of AccD4 for structural studies. The crystallisation parametric space was explored by varying a number of parameters like protein concentration, crystallisation drop volume, and the presence of biotin and/or CoA (a substrate analog) in various stoichiometric proportions. Routine crystal screening and expansion experiments resulted in crystals in a few conditions; unfortunately, these crystals were not of diffraction quality [Figure 3-18D].

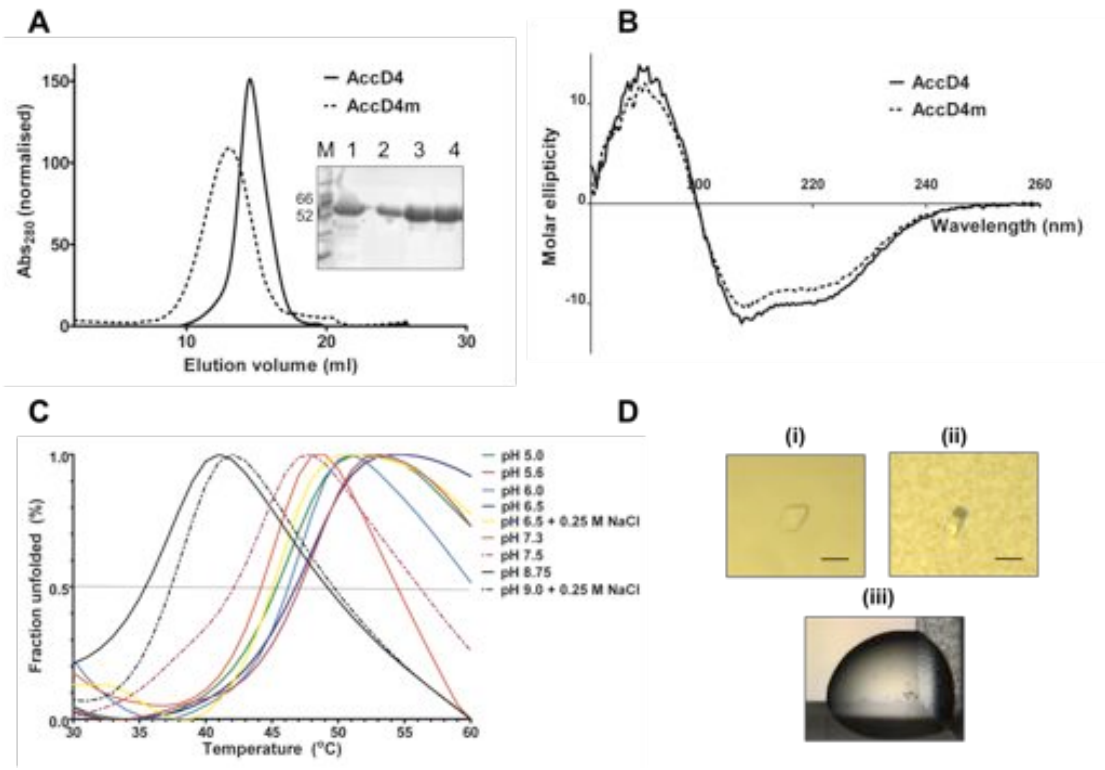

Figure 3-18. AccD4 crystallisation trials. **(A)** SEC profiles of AccD4 and methylated AccD4 on Superose™ 6 10/300 GL. **Inset:** SDS-PAGE analysis of peak fractions of purified AccD4 (lanes 1 and 2) and methylated AccD4 (lanes 3 and 4); lane M: protein marker with standard molecular weights (in kDa). **(B)** CD spectra (molar ellipticity in M deg cm^2 $dmol^{-1}$ plotted against wavelength in nm) of AccD4 and methylated AccD4 measured between 180 and 260 nm wavelength with settings as follows: 0.1 cm cell length, 1 nm band width, 0.5 nm step size and 0.5 sec response duration. **(C)** Thermofluor assay on AccD4 pointed to a number of potentially favourable acidic buffer conditions for increased thermostability: citric acid (pH 5.0); succinic acid/phosphate/glycine (pH 5.6); potassium phosphate (pH 6.0); bis-tris (pH 6.5); bis-tris (pH 6.5) + 0.25 M sodium chloride; thermostability was relatively lower at conditions of basic pH: ammonium acetate (pH 7.3); HEPES (pH 7.5); citric acid/CHES/HEPES (pH 8.75); bicine (pH 9.0) + 0.25 M sodium chloride. **(D)** (i) Bipyrimidal and (ii) Short-rod shaped crystals of AccD4 obtained using 0.2 M magnesium acetate, 0.1 M sodium cacodylate (pH 6.5), 30% (v/v) MPD. The black bar corresponds to a length of 50 µm. (iii) Microcrystals of AccD4 (with CoA, 1:5) obtained with 0.1 M bicine (pH 9.0), 5% (w/v) PEG 6000.

Data from thermal denaturation experiments on AccD4 suggested that the thermostability (an usually critical criterion for successful crystallisation) of AccD4 could be enhanced in an acidic environment [Figure 3-18C]. However, attempts to purify or 'exchange' the protein into such buffer conditions were not successful because of the problem of excessive precipitation.

It has been known that protein motions due to solvent-exposed, flexible amino acid side chains can be disruptive to an ordered three-dimensional crystal lattice. In such cases, surface entropy reduction mutagenesis techniques like lysine methylation have been used as rescue strategies for effective protein crystallisation (Walter*, et al.*, 2006). The primary structure of AccD4 contains about 4% lysine, an amino acid with a flexible side chain. In an attempt to restrict overall protein mobility, AccD4 was modified by a standard methylation protocol; complete methylation of all Lys residues was ensured by MALDI-TOF (0.88 kDa increase in molecular weight, corresponding to the trimethylation of 21 Lys). In SEC experiments, methylated AccD4 eluted as a higher oligomeric form than AccD4 [Figure 3-18A]; based on static light scattering data, methylated AccD4 could be a 9-mer in solution. The secondary structural content of methylated AccD4, however, was identical to that of AccD4, as verified by CD spectroscopy [Figure 3-18B]. Initial crystallisation trials with methylated AccD4 (of unusual oligomeric status) were unsuccessful and hence, work with this protein was discontinued. Effort was invested in the characterisation of AccD4 by small angle X-ray scattering, a biophysical method widely considered to be complementary to X-ray crystallography.

3.2.1.4 SAXS analysis

AccD4 was subjected to preliminary structural investigation by small angle X-ray scattering in collaboration with Anne Tuukkanen, Svergun group, EMBL Hamburg. Scattering data of usable quality could not be obtained at protein concentrations higher than 1 mg/ml because of the formation of high-molecular weight aggregates at these concentrations. Molecular dimensions obtained from the investigation of samples at concentrations lower than 1 mg/ml supported the possibility of the trimeric quaternary structure of AccD4 [Table 3-4]. To enable comparative analysis, AccD5

and AccD6 samples were also included in the SAXS investigation; molecular dimensions obtained from these experiments verified the homohexameric state of AccD5 and the homodimeric existence of AccD6 (discussed in sub-chapter 3.3).

Table 3-4. Dimensions of YCC β-subunits from SAXS analysis

Protein	R_g (nm)	D_{max} (nm)	Volume (nm^3)	MW (kDa)	Oligomeric state	Quality (%)
AccD4	5.08 ± 3%	17.78	522	170	Trimer	84
AccD5	5.31 ± 1%	18.2	720	319	Hexamer	94
AccD6	4.98 ± 1%	17.5	355	129	Dimer	89

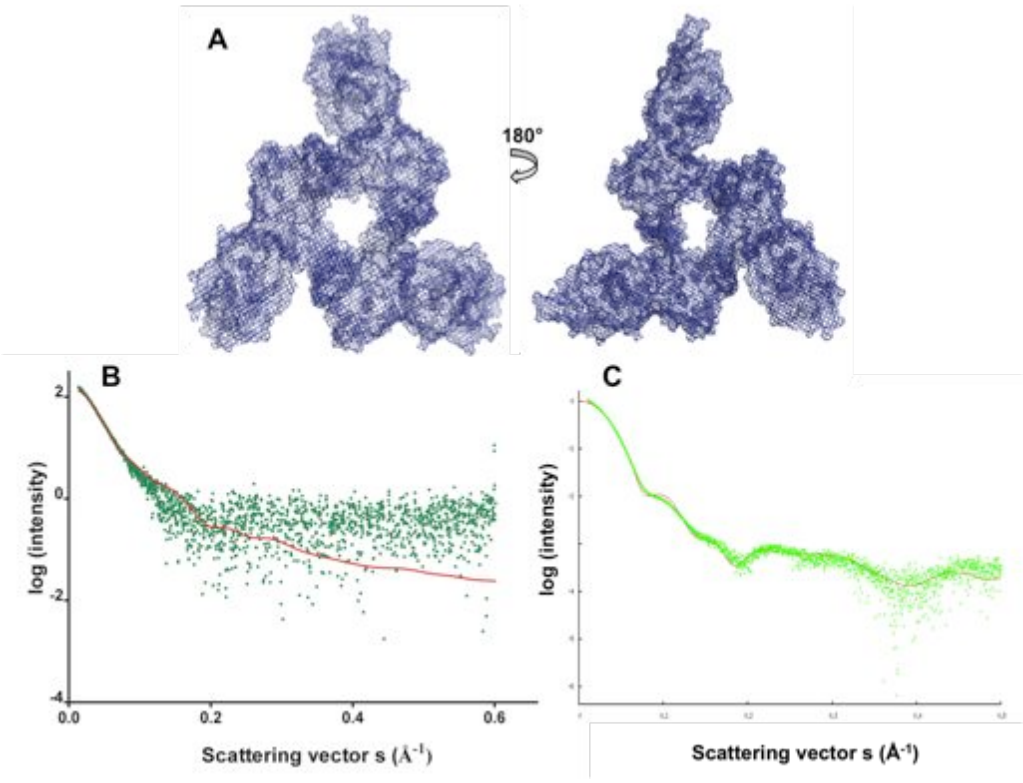

Figure 3-19. SAXS analysis of AccD4. **(A)** Mesh representation of the rigid body model of AccD4 generated by *SASREF*. **(B)** Experimental scattering data (green) from AccD4 and the calculated scattering curve (red) from the rigid body model plotted as the logarithm of scattered intensity against the momentum transfer. **(C)** For comparison, experimental scattering data (green) from AccD5 and the scattering curve (red) calculated from the crystal structure of AccD5, plotted as the logarithm of scattered intensity against the momentum transfer.

Rigid body modeling of AccD4 based on SAXS measurements generated a model with an unexpected arrangement of AccD4 protomers; the model that best fits (goodness of fit: approximately 2.0; Figure 3-19B) the observed data supports the arrangement of three AccD4 monomers in a fan-like fashion around a central axis [Figure 3-19A], unlike the compact trimeric ring organization found in other actinobacterial YCC β-subunits (Diacovich*, et al.*, 2004, Holton*, et al.*, 2006). In the SAXS-generated model, that measures approximately 180 Å at the longest dimension, the C-terminal subdomains of AccD4 form interactions at the central core of the homotrimer while the N-terminal subdomains are located at the periphery of the ring, away from the N-terminal or C-terminal subdomains of adjacent protomers. Efforts to confirm this model, by means of collecting data of finer quality so as to accomplish better goodness of fit, are underway.

3.2.2 Discussion

In this part of the study, the structural and biophysical characterisation of AccD4, and its reconstitution in the form of a holo-complex together with its interacting partner AccA3, were attempted. The YCC components AccA3 and AccD4 did not assemble into a stable complex appropriate for crystallographic studies. As with the case of AccA3-AccD5 discussed in the previous sub-chapter, the inability of AccA3-AccD4 to form a stable complex can be attributed to the likely absence of strong interactions between the two subunits. It is also possible that the subunits AccA3 and AccD4 might assemble into a stable entity only in the presence of/by the action of a third (or more) interacting partner(s). On the basis of co-immunoprecipitation studies using *M. smegmatis* AccD4, it has earlier been proposed that the proteins AccA3 and AccD5 are the other subunits of the AccD4-containing acyl-CoA carboxylase involved in mycolic acid biosynthesis (Portevin*, et al.*, 2005), but the existence of such a ternary complex has not been conclusively proven. The observation that AccA3 and AccD5 co-immunoprecipitate with AccD4-myc (Portevin*, et al.*, 2005), is not direct evidence that the three components arrange in the form of a multi-subunit complex, but could be a result of AccA3 being the common biotin-carboxylating partner of the β-subunits AccD4 and AccD5 individually. Furthermore, each structurally characterised holo-YCC complex has been shown to contain a core catalytic β-subunit composed of

protomers of only one protein (Huang*, et al.*, 2010, Huang*, et al.*, 2011). Based on existing literature, it is unlikely that an acyl-CoA carboxylase complex would contain a heteromeric β-subunit capable of carboxylating more than one acyl-CoA substrate.

The function of the AccD4-containing carboxylase (i.e., the carboxylation of the fatty acid that will serve as the α-chain of the mycolic acid, whose meromycolate part will be activated and provided by FadD32) was deduced based on gene knockout and lipidomic studies in *C. glutamicum* (Portevin*, et al.*, 2005). The understanding that AccD4 is responsible for the carboxylation of C_{16} fatty acids was on the basis of two observations: (1) the presence of an alkylmalonate species (whose mass spectrum corresponded to that of an activated C_{16} fatty acid derivative) in wild-type *C. glutamicum*, but not in the *ΔaccD4* mutant; (2) the accumulation of tetradecylmalonic acid in the *ΔfadD32* mutant, but not in the *ΔaccD4* mutant. Nevertheless, a comprehensive understanding of the enzyme kinetics of the AccD4-containing YCC is lacking. A surprising finding from the current study has been the fact that AccD4 exists as a trimer in solution, unlike the other hexameric β-subunits found in literature. Preliminary kinetic analyses, aimed at checking if the AccD4 trimer is a catalytically active carboxyltransferase, were indeed affirmative, hence suggesting that a hexameric arrangement of protomers is not necessary for carboxylase activity.

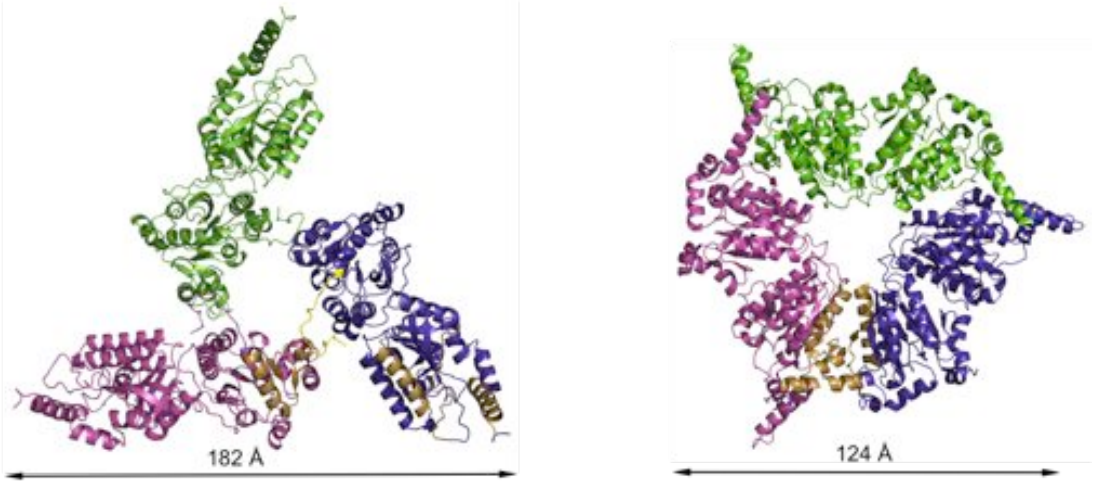

Figure 3-20. SAXS model of AccD4. The rigid-body model of AccD4 (left) based on SAXS measurements compared with the crystal structure of AccD5 (right). For clarity, only one trimeric ring of the AccD5 hexamer is shown. The monomers of each model are coloured green, magenta and purple. The structural elements important for hexameric arrangement and catalytic activity in AccD5 and the corresponding features mapped on the AccD4 model are shown in golden colour.

Preliminary SAXS experiments also indicated molecular dimensions of AccD4 that correspond to a trimeric quaternary arrangement. Rigid body modeling based on SAXS measurements generated a preliminary model that describes relative positioning of the AccD4 protomers in a fashion different from that of the other YCC β-subunits described so far. It has been known that conserved didomain dimeric interactions are crucial for carboxylase activity; in all the CT subunits characterised so far, it has been shown that the juxtaposition of members of the dimeric pair gives rise to conserved oxyanion holes, strategically located for the stabilization of the acyl-CoA substrate, the carboxybiotin (translocated to the active site of the β-subunit by the BCCP domain of the α-subunit) and the reaction intermediates. The possibility of such an arrangement is not supported by the current model of AccD4 [Figure 3-20]; therefore, it remains to be verified if such an arrangement is biologically relevant. The unsatisfactory goodness of fit between the observed data and the expected model could be arising due to possible oligomeric heterogeneity; the implementation of measures (such as in-line purification and biophysical characterisation steps coupled to small angle scattering measurements) potentially capable of overcoming the problem of sample heterogeneity will be required to validate/verify the current model and to thereby derive better insights into the structure and function of AccD4.

3.3 Structural study of the carboxyltransferase subunit of acetyl-CoA carboxylase

Acetyl-CoA carboxylase is one of the most fundamental metabolic enzymes in the majority of life forms as it catalyses the generation of malonyl-CoA, a key lipid metabolite. The two fatty acid synthase systems (FAS-I and FAS-II) of *Mycobacterium* species rely on the supply of malonyl-CoA as substrate for the biosynthesis of the complex lipids that constitute the cell wall, such as the phthiocerol dimycocerosates (Trivedi*, et al.*, 2005), sulpholipids (Jackson*, et al.*, 2007), and long-chain α-alkyl, β-hydroxymycolic acids (Bhatt*, et al.*, 2007). An acyl-CoA carboxylase of *M. tuberculosis*, reconstituted from components AccA3 and AccD6, was shown to be active on acetyl-CoA and propionyl-CoA with comparable efficiency (Daniel*, et al.*, 2007). On the basis of the location of *accD6 (Rv2247)* in a genetic locus together with members of the FAS-II system and the characterisation of an *accD6 M. smegmatis* conditional mutant (Kurth*, et al.*, 2009), the physiological role of *M. tuberculosis* AccA3-AccD6 is known to be the carboxylation of acetyl-CoA. Although the function of AccD6 is known, its biophysical properties and the structural basis of its catalytic activity and substrate preference are not completely understood. Sub-chapter 3.3 reports the crystallisation, structure determination and analysis of AccD6.

3.3.1 Results

3.3.1.1 Purification of AccD6

Using construct M27 [Table 2-2], the protein AccD6 could be overexpressed in the BL21 (DE3) expression strain of *E. coli*. Typical overexpression and purification experiments resulted in a final yield of 10 – 15 mg of AccD6 protein per litre of culture. During SEC experiments using Superose™ 6 10/300 GL, AccD6 eluted as a single symmetrical peak whose elution volume corresponded to the dimeric state. Static light scattering experiments indicated a molecular mass of approximately 123 kDa (monomer MW: 53 kDa), confirming the existence of AccD6 as a homodimer.

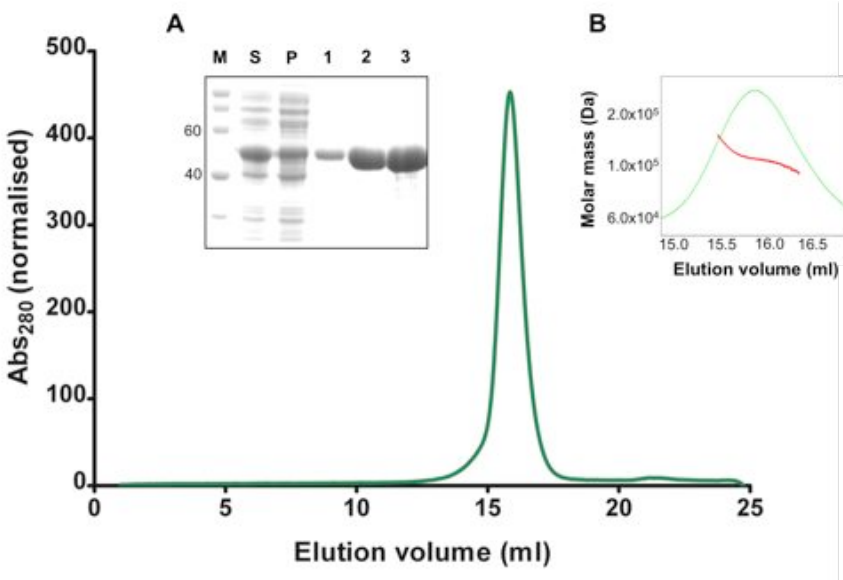

Figure 3-21. Purification of AccD6. SEC elution profile of AccD6 on Superose™ 6 10/300 GL. Molecular weight standards used and respective elution volumes (ml) in brackets: 670 kDa (12.1), 158 kDa (15.5). **Inset A:** SDS-PAGE analysis of supernatant (S) and insoluble pellet (P) of cell lysate, wash (1) and eluate (2) of IMAC purification, and peak fraction (3) of SEC purification; lane M: protein marker with standard molecular weights (in kDa). **Inset B:** Mass distribution (molar mass as a function of elution volume) of SEC peak fractions, as determined by SLS.

3.3.1.2 Crystallisation of AccD6

AccD6, in apo-form and in combination with CoA (substrate analog) and/or biotin (cofactor analog) at varying stoichiometries, was used in crystallisation trials by the sitting-drop and hanging-drop vapour diffusion methods. Crystals of two different morphologies were obtained [Figure 3-22]. The crystallisation conditions corresponding to both crystal forms were similar to those observed in a structural genomics consortium study conducted elsewhere during the course of this work (Niu, *et al.*, 2011). Crystals of type I were roughly rod-shaped and generally, either possessed weak diffraction properties or resulted in X-ray diffraction data marred by very high degrees of mosaicity; the more robust type II crystals, on the other hand, had a trigonal bipyramidal morphology and generally diffracted X-rays to high resolution. Crystals were treated with a suitable cryoprotectant (10 – 15% (v/v) glycerol, in addition to the mother liquor) and flash-cooled in a stream of liquid nitrogen immediately prior to exposure to X-rays; the diffraction experiments were performed at the P14 beamline of PETRA III, DESY.

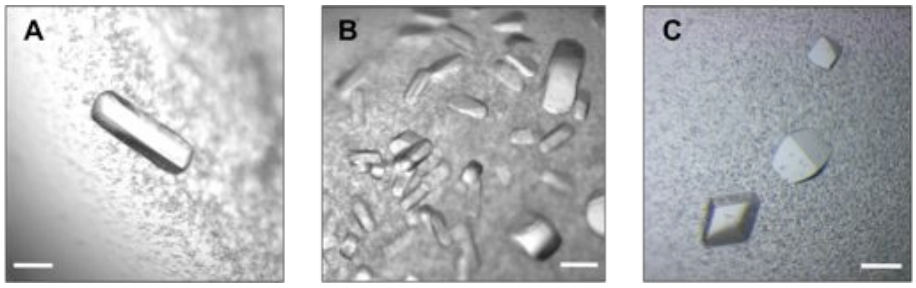

Figure 3-22. Crystallisation of AccD6. (A) Type I crystals of AccD6 obtained with 2.6 M sodium formate (pH 7.0). (B) Type I crystals of AccD6 + biotin + CoA (molar ratio 1:3:3) obtained with 2.75 M sodium formate (pH 7.5). (C) Type II crystals of AccD6 obtained with 27% (w/v) PEG 3350, 0.1 M Tris (pH 8.5) and 0.3 M ammonium sulphate. The white bars represent a length of 100 μm.

3.3.1.3 X-ray data collection, structure solution and refinement

Diffraction data from a crystal of type II were collected to a maximum resolution of 1.95 Å at a detector distance of 270 mm; the data were indexed, scaled and merged using *XDS* (Kabsch, 2010) and *SCALA* (Evans, 2006). The crystal belonged to the common orthorhombic space group $P2_12_12_1$, with a space group confidence of 0.973, Laue group confidence of 0.993 and systematic absence probability of 0.983, as estimated by the program *POINTLESS* (Collaborative Computational Project, 1994). Solvent content analysis based on Matthew's coefficient indicated the presence of two monomers in the asymmetric unit. The structure of AccD6 was solved by molecular replacement with *Phaser* (McCoy*, et al.*, 2007) using the crystallographic structure of a monomer of AccD5 as the search model. The parameters R_{conv} and R_{free} and adherence to allowed peptide bond geometry were used to monitor the progress of the alternating, iterative cycles of manual model building with *COOT* (Emsley & Cowtan, 2004) and maximum-likelihood refinement with REFMAC (Murshudov*, et al.*, 1997). The crystallographic data collection and refinement statistics were as in Table 3-5.

Table 3-5. Crystallographic statistics of AccD6

Crystallographic parameter	**Value**
Data collection and processing	
Wavelength [Å]	1.24
Resolution range [Å]	76.94 – 1.95 (2.05 – 1.95)
Space group	$P2_12_12_1$
Unit cell dimensions [Å]	76.94, 90.31, 154.56
Average mosaicity [°]	0.01
Number of measured reflections	1021845 (145681)
Number of unique reflections	79385 (11309)
Multiplicity	12.9 (12.9)
Mean $I/\sigma(I)$	34.4 (5.5)
Completeness [%]	99.8 (98.5)
[1] R_{sym}	0.04 (0.50)
Values in parentheses correspond to the highest resolution shell.	
Refinement	
[2] R_{conv} / [3] R_{free}	0.181 / 0.241
Average B-factor [Å^2]	30.17
Root mean square deviations	
Bond lengths [Å]	0.019
Bond angles [°]	1.970
Ramachandran plot analysis	
Favoured [%]	97.1
Additionally allowed [%]	2.9

[1] $Rsym = \sum_h \sum_j \left| I_{h,j} - \langle I_h \rangle \right| \Big/ \sum_h \sum_j I_{h,j}$ where $I_{h,j}$ is the intensity of the j^{th} observation of unique reflection h.

[2] $Rconv = \sum_h \left\| Fo_h \right| - \left| Fc_h \right\| \Big/ \sum_h \left| Fo_h \right|$ where Fo_h and Fc_h are the observed and calculated structure factor amplitudes for reflection h.

[3] R_{free} is equivalent to R_{conv}, but is calculated using a 5% disjoint set of reflections excluded from the maximum likelihood refinement stages.

3.3.1.4 Overall architecture of AccD6

Consistent with size exclusion chromatographic and static light scattering data, the crystal structure of AccD6 is a homodimer [Figure 3-23A]; each of the two monomers (A and B) is made of two structural subdomains - the N- and C-terminal subdomains - and assumes a variant of the conventional crotonase fold. The monomers are structurally similar, as evidenced by an RMSD value of 0.475 Å for an alignment of about 400 equivalent main-chain carbon atoms. Residues 328 to 333 and 400 to 430 of monomer A and residues 156 to 187 of monomer B could not be modeled with confidence, presumably due to high conformational flexibility. Residues 400 to 430 of monomer B could be modeled on the basis unbiased electron density features, though

the corresponding region could not be modeled in monomer A; this was most probably because of relatively reduced conformational flexibility of this region (in monomer B) due to close crystal contacts (generated by the effect of molecular packing) with a neighbouring symmetrical partner in the unit cell. This region contains a few residues that are putatively involved in substrate-binding and will be discussed in section 3.3.1.7.

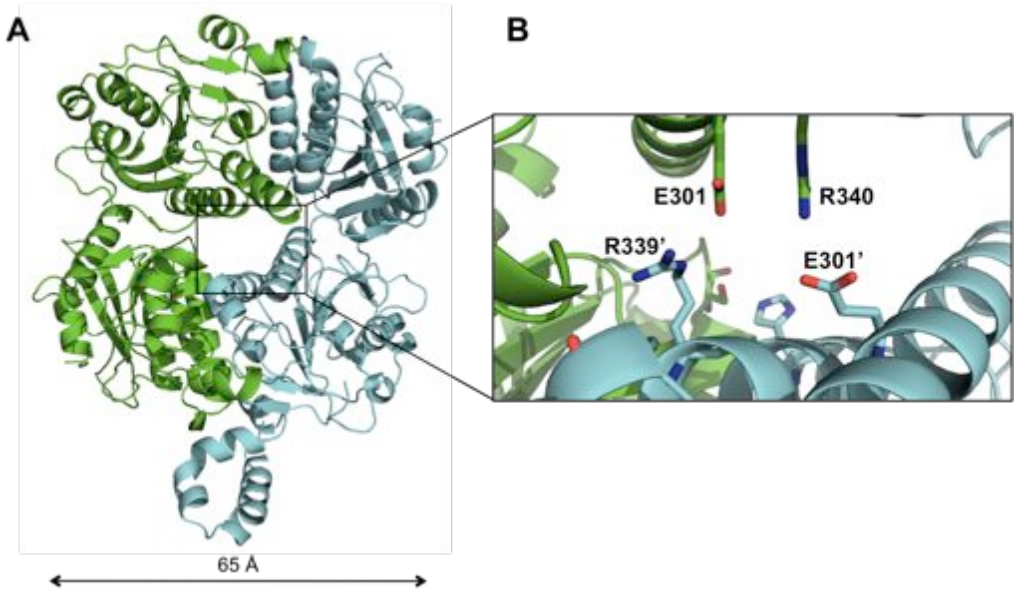

Figure 3-23. The crystal structure of AccD6. (A) The two monomers of AccD6 are coloured green and cyan respectively. (B) A reoriented, closer view of the dimeric interface at the central core of the homodimer. Residues that make strong interactions at the dimeric interface are shown in 'stick' mode with carbon, nitrogen and oxygen atoms coloured cyan/green, blue and red respectively.

As in the case of the A-D dimeric pair in the crystal structure of AccD5 (section 3.1.1.5.1), the N-terminal subdomain of one monomer of AccD6 makes extensive atomic interactions with the C-terminal subdomain of its counterpart. About 20% (5350 $Å^2$) of the total surface area of the protein (29800 $Å^2$) is buried and solvent-inaccessible, implying the formation of a very strong and compact dimeric interface. Several salt bridges and hydrogen bond-mediated interactions [Table 3-6] between fully- or functionally-conserved residues of the two monomers contribute to the maintenance of this intact dimeric interface. The details of these interactions were verified by interface and assembly analyses using the *PDBePISA* tool (Krissinel & Henrick, 2007).

Table 3-6. Interactions at the AccD6 dimeric interface

Monomer A	Monomer B	Distance (Å)
Salt bridges		
Glu 301 (OE1)	Arg 339 (NH2)	3.88
Glu 301 (OE2)	Arg 339 (NE)	3.31
Glu 301 (OE2)	Arg 339 (NH2)	3.28
Arg 340 (NH2)	Glu 301 (OE1)	3.45
Arg 340 (NH2)	Glu 301 (OE2)	3.12
Hydrogen bonds		
Ser 122 (OG)	His 346 (NE2)	3.17
Ser 177 (O)	Trp 334 (NE1)	2.96
Ser 188 (O)	Arg 339 (NH1)	2.80
Glu 301 (OE2)	Arg 339 (NE)	3.31
Glu 301 (OE2)	Arg 339 (NH2)	3.28
Val 337 (O)	Tyr 141 (OH)	3.51
His 346 (NE2)	Ser 122 (OG)	3.18
Arg 340 (NH2)	Glu 301 (OE2)	3.12
Val 159 (N)	Glu 333 (OE2)	2.96
Thr 160 (N)	Glu 333 (OE2)	3.74

Table 3-7. Top structural homologs of AccD6

	Target (PDB entry)			Scoring		Structural/sequence similarity		
	PDB ID	**Organism**	**Protein**	**Q**	**Z**	**Sequence identity (%)**	**rmsd (Å)**	**N_{align}**
1	3N6R	*Roseobacter denitrificans*	PCC β-subunit	0.64	21.1	44	1.26	397
2	2BZR	*Mycobacterium tuberculosis*	PCC β-subunit (AccD5)	0.64	21.7	45	1.31	397
3	1ON9	*Propionibacterium freudenreichii subsp. shermanii*	Transcarboxylase 12S domain	0.63	21.4	41	1.27	392
4	1XNY	*Streptomyces coelicolor*	PCC β-subunit	0.62	21.9	42	1.28	395
5	1X0U	*Sulfolobus tokodaii*	Putative PCC β-subunit	0.61	20.8	42	1.35	396

The three-dimensional structure of AccD6 shares high degrees of structural homology with several carboxyltransferase proteins [Table 3-7], as identified by *PDBeFold* analysis (Krissinel & Henrick, 2007). In the above table, PDB entries that were identified as 'hits' more than once are listed only once. For these structures, values for the chain with the closest structural homology to AccD6 are shown. Entries have been sorted by the Q-score, which is an indicator of the quality function of the Cα alignment, as maximized by the SSM alignment algorithm (Krissinel & Henrick, 2007). The Z-score represents the statistical significance of a match in terms of Gaussian statistics; N_{align} is the length of alignment, or number of matched residues, calculated at best 3D superposition of the query and target structures.

3.3.1.5 Dimer *versus* hexamer

Unlike the *M. tuberculosis* PCC β-subunit AccD5 (this study, Holton, *et al*., 2006, Lin, *et al*., 2006), the *S. coelicolor* PccB (Diacovich, *et al.*, 2004), and the 12S domain of *P. shermanii* transcarboxylase (Hall, *et al.*, 2003) that have been investigated structurally, the carboxyltransferase AccD6 of *M. tuberculosis* ACC assumes a dimeric quaternary arrangement. The AccD6 homodimer is structurally homologous to and superposes well with the A-D (or B-E / C-F) dimer of AccD5 [Figure 3-24A], as evidenced by an RMSD value as low as 1.3 Å for an alignment of approximately 400 equivalent Cα atoms.

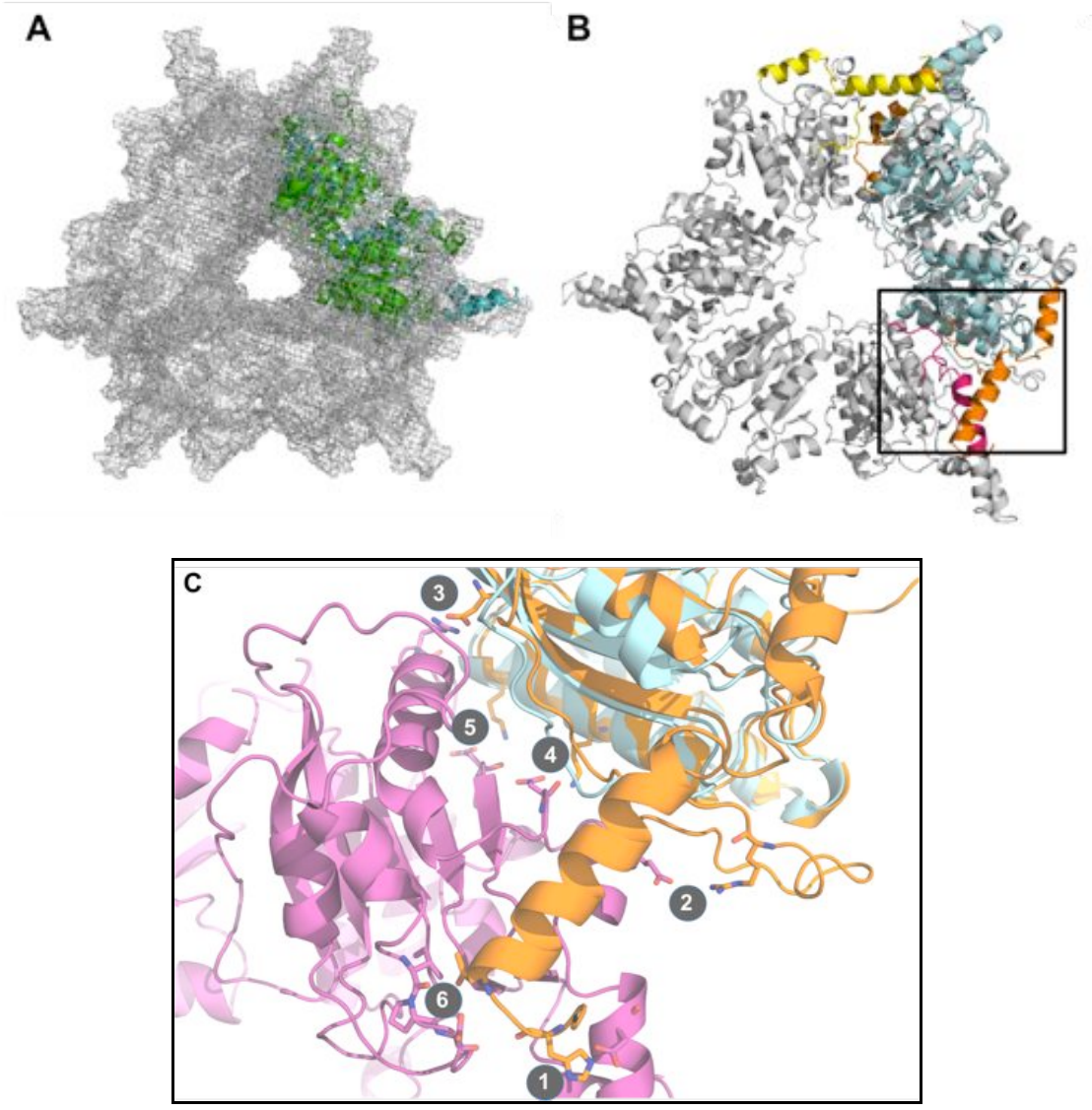

Figure 3-24. AccD6 lacks structural features that favour hexamer formation. (A) Superposition of the crystal structure of AccD6 (monomers in cyan and green) with that of AccD5 represented as a grey mesh. (B) Structural superposition as in (A), but only one layer of each structure (one trimeric ring of AccD5 and one monomer of AccD6) are shown, for

clarity. The secondary structural elements that predominantly favour hexameric arrangement in AccD5 are coloured orange; the corresponding elements of adjacent monomers on either side are coloured yellow and magenta respectively. (C) Closer view of the region marked by the box in (B). The monomers of AccD5 are coloured orange and magenta respectively. Residues involved in interactions that most likely establish hexameric arrangement are represented in the 'stick mode' with carbon, nitrogen and oxygen atoms coloured orange/magenta, blue, and red, respectively. The details of the interactions labeled 1 to 6 are listed in Table 3-8.

Comparative structural and sequence analyses revealed that AccD6 lacks certain secondary structural elements that are crucial for interactions between protomers in order to result in a hexameric ring arrangement [Figure 3-24C, 3-25]. A β-subunit that forms homohexamers, AccD5 for instance, possesses a short loop and an approximately 15-residue-long helix at the N-terminal of one protomer (say, monomer A). These elements are predominantly responsible for the formation of atomic interactions [Table 3-8], in the form of salt bridges and hydrogen bonds, with regions close to the C-terminal of a neighbouring protomer (say, monomer B) in the trimeric ring of the protein structure. These interactions are identical between the pairs B-C and C-A as well, and thus contribute to the maintenance of the trimeric ring structure. The existence of these interactions, in addition to the strong dimeric interactions between pairs A-D, B-E, and C-F, forms the basis of the hexameric architecture of most β-subunits.

Table 3-8. Interactions that favour hexamer formation in AccD5

No.	Monomer A	Monomer B	Distance (Å)
	Salt bridges		
1	His 23 (NE2)	Asp 497 (OD1)	3.61
	His 23 (NE2)	Asp 497 (OD2)	3.22
	His 23 (ND1)	Asp 497 (OD2)	3.91
2	Arg 92 (NH2)	Glu 507 (OE1)	3.75
3	Asp 109 (OD1)	Arg 533 (NH2)	3.56
4	Lys 134 (NZ)	Asp 512 (OD1)	2.66
5	Lys 141 (NZ)	Asp 447 (OD1)	3.32
	Hydrogen bonds		
6	Thr 25 (OG1)	Ile 304 (O)	2.61
	Thr 25 (OG1)	Pro 305 (O)	2.74
	Thr 25 (N)	Asp 306 (OD2)	2.95

It is most likely that AccD6 forms neither homotrimers nor homohexamers, because of the absence of such interactions. The residues His 23 and Thr 25 of AccD5 are part

of the N-terminal loop and α-helix respectively and facilitate a strong molecular 'hand shake' of the N-terminal subdomain of one monomer with the C-terminal subdomain of the neighbouring monomer. These two amino acids do not have equivalent residues in the AccD6 primary structure [Figure 3-25], thus ruling out the possibility of such an anchoring set of interactions in AccD6.

In the AccD5 structure, Arg 92 (that forms electrostatic interactions with Glu 507) is part of a 30-residue-long loop region between the third α-helix and the first β-strand of the canonical crotonase topology. The corresponding loop in the AccD6 structure is significantly shorter (17 residues long) and is in close proximity to the N-terminal end of the protein due to the absence of the N-terminal loop and α-helix, as mentioned earlier [Figure 3-24C]. The positively charged residues in this short loop that might contribute to a potential salt bridge are Arg 37 and Arg 39, but the side chains of these residues point farther away, in the opposite direction of the Arg 92 side chain of AccD5. Additionally, the conserved Glu 507 (AccD5 numbering) of hexameric β-subunits has been replaced by the shorter Asp 437 in AccD6, causing a further reduction in the possibility of establishment of a charge-based interaction in this region of the protomers.

Asp 109 and Arg 533 of AccD5 make up another pair of interacting residues and lie close to the central core of the hexameric ring. These residues have been substituted in AccD6 by the oppositely charged Arg 53 and the non-polar Ala 463, respectively, thereby precluding the establishment of a charge-based interaction, in this region, between protomers. Similarly, the conserved residues Lys 134 and Lys 141 (AccD5 numbering), which are involved in key electrostatic interactions along the intratrimeric interface in hexameric β-subunits, have been replaced in AccD6 by the negatively charged Asn 78 and Glu 85, respectively [Figure 3-25]. The structural data indicates that it is highly probable that all these amino acid differences together form the basis of the nonexistence of homotrimeric or homohexameric arrangement in AccD6.

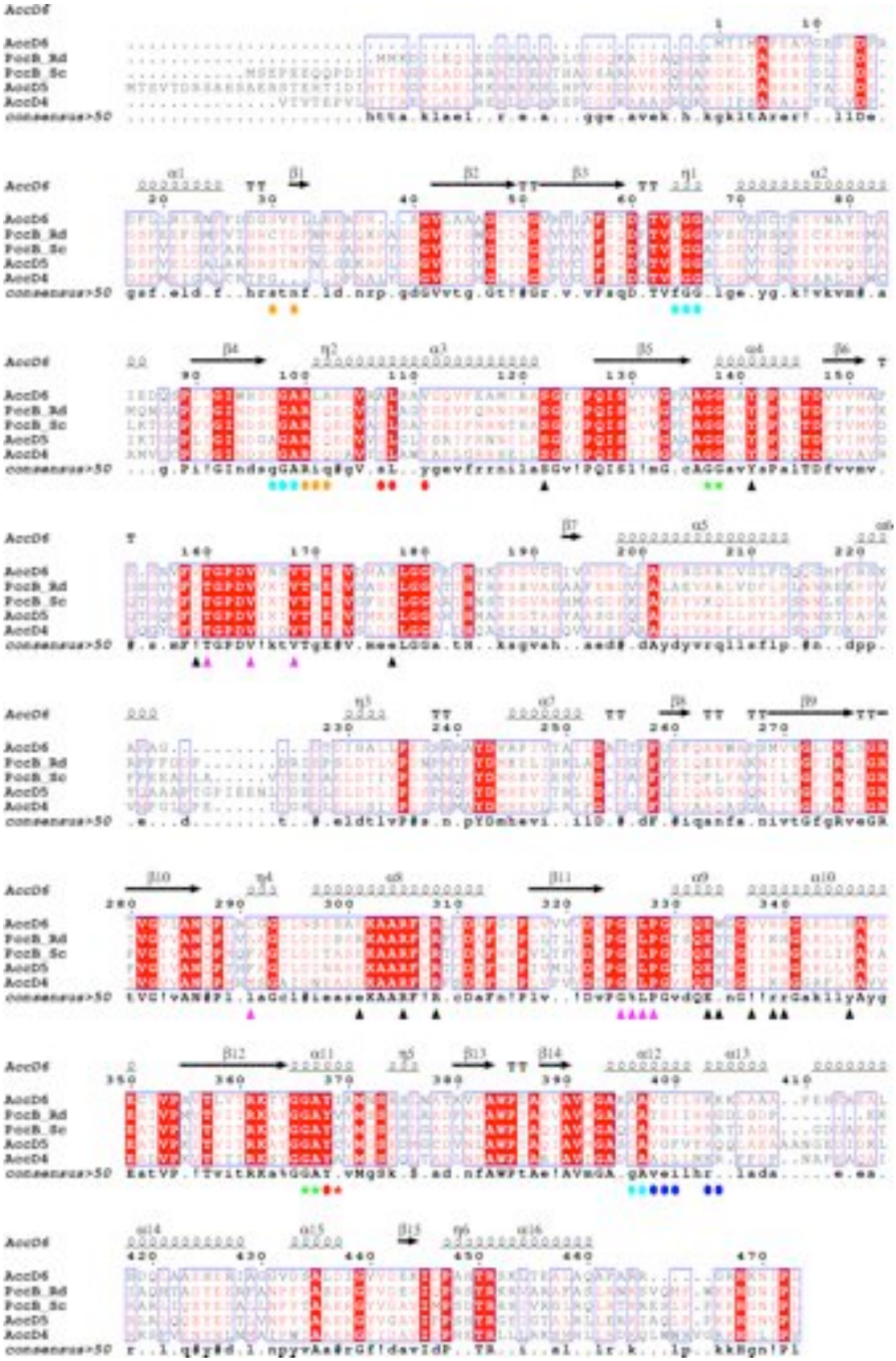

Figure 3-25. Sequence alignment of AccD6 with structural relatives. Sequence alignment of the *M. tuberculosis* carboxyltransferases AccD6, AccD5, and AccD4, the PCC

β-subunit of *R. denitificans* (PccB_Rd), and the PCC β-subunit of *S. coelicolor* (PccB_Sc). Residue numbering and secondary structure assignment (arrow marks for strands and helical symbols for helices) shown on the first row of each block of the alignment are based on the AccD6 structure. Conserved blocks of residues are shown in blue rectangular frames, with partially- and fully-conserved residues represented in a white and red background respectively. Legend: red star, putative substrate specificity determinant residue; green star, conserved active site residues that favour oxyanion hole formation; black triangle, residues contributing to the dimeric interface; pink triangle, putative carboxybiotin-binding residues; circle, putative substrate-binding residues that might interact with the adenine and biphosphate (dark blue), ribose (cyan), pantotheine (orange), and acetyl (red) moieties respectively; $ and !, conserved hydrophobic amino acids; %, conserved aromatic residues; #, conserved hydrophilic residues. The figure was generated using the *ESPript 2.2* program (Gouet*, et al.*, 2003).

Some non-actinobacterial β-subunits have been reported to exist and function as homodimers (Wendt*, et al.*, 2003, Zhang*, et al.*, 2003). The CT subunit of yeast ACC is known to feature dimeric didomain interactions that are ubiquitous to all carboxyltransferase proteins (Zhang*, et al.*, 2003). The unique feature of the homodimeric yeast protein, however, is the presence of structural attributes (a large β-insertion at its N-terminal subdomain, one β-insertion and one α-helical insertion at its C-terminal subdomain, and an additional four-helix stretch at its C-terminus) that have been implicated in the prevention of hexamer formation (Zhang*, et al.*, 2003). Interestingly, AccD6 does not form hexamers even in the absence of such additional structural features.

3.3.1.6 Putative cofactor-binding site

Attempts to obtain the crystal structure of AccD6 in complex with a cofactor analog and/or a substrate analog by means of co-crystallisation were unsuccessful. The absence of electron density corresponding to the cofactor or substrate analog could be a result of the hydrolysis of the analogs in the course of the crystallisation process or could reflect a lack of specific contacts between the protein and the analogs under the crystallisation conditions tested. Nevertheless, the apo-AccD6 structure provided snapshots of the putative cofactor- and substrate-binding sites of the protein. Based on structural homology to previously characterised β-subunits, the substrate (acetyl-CoA) and cofactor (carboxybiotin) are expected to bind at the interface between the N-terminal subdomain of one AccD6 monomer and the C-terminal subdomain of its dimeric partner [Figure 3-26].

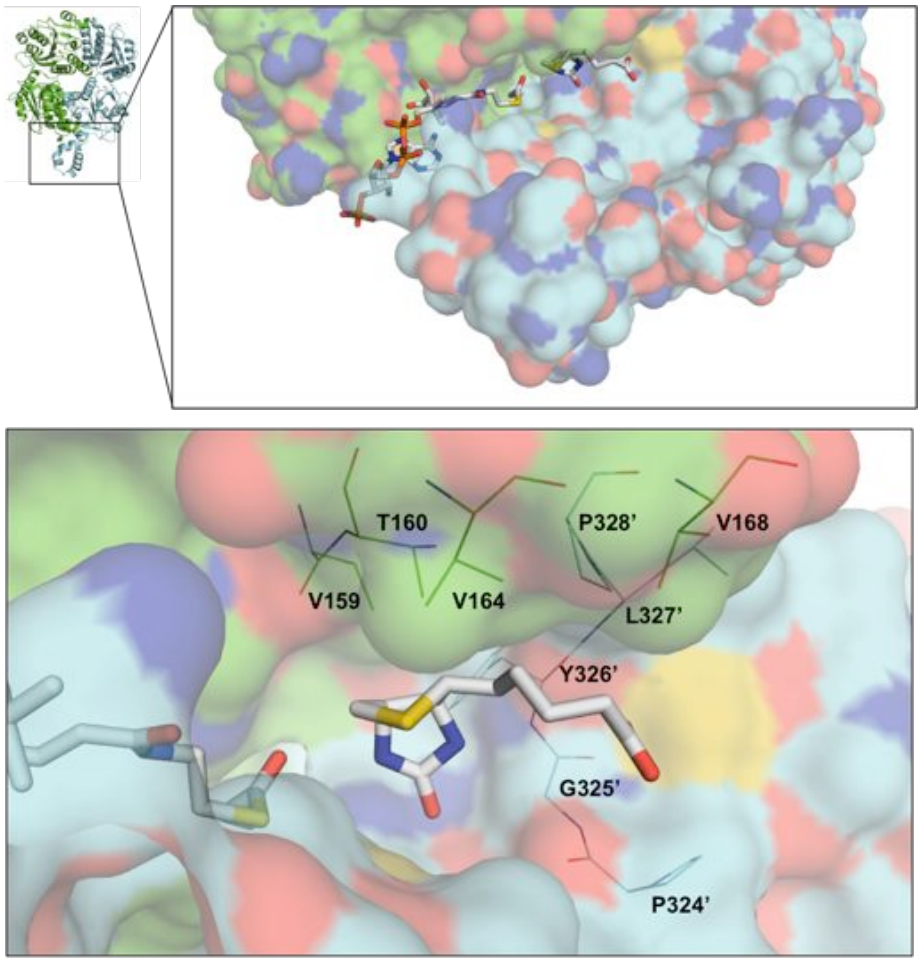

Figure 3-26. The putative cofactor-binding site. Top panel: The putative active site of AccD6 lies at the dimeric di-domain interface; the two monomers are coloured green and cyan respectively. Models of biotin and propionyl-CoA (based on their location in the crystal structure of *S. coelicolor* PccB; PDB ID: 1XNY) are shown in 'stick mode' with carbon, nitrogen, oxygen, phosphorus, and sulphur atoms coloured grey, blue, red, orange, and yellow, respectively. In the surface representation of AccD6, carbon, nitrogen, oxygen, and sulphur atoms have been coloured green/cyan, blue, red, and yellow, respectively; the surfaces of the monomers have been rendered partially transparent for the sake of clarity. **Bottom panel:** Closer view of the putative cofactor-binding site. Residues that could potentially be involved in cofactor binding are shown as 'lines' with colouring scheme as in the top panel.

The putative pocket for cofactor-binding is defined by the short loops between N-β5 and N-α4 of monomer A, C-β12' and C-α11', and the relatively longer loop between C-β11' and C-α9' of monomer B [Figure 3-25]. The biotin-streptavidin interaction is known to be predominantly van der Waals and hydrophobic in nature, involving the participation of many aromatic residues (Lindqvist & Schneider, 1996). The crystal structure of AccD6 revealed that the potential carboxybiotin-AccD6 interaction would also be predominantly hydrophobic, carried out by fully or functionally conserved residues that line the carboxybiotin-binding pocket. Residues Val 159, Val 164, and

Val 168 of monomer A and residues Leu 291', Pro 324', Gly 325', Tyr 326', Leu 327', and Pro 328' contribute to the hydrophobic environment suitable for cofactor-binding [Figure 3-26]. Furthermore, the putative cofactor-binding pocket of AccD6 is appropriately shaped to allow the ureido moiety of the cofactor to point to two possible oxyanion-holes that, in homologs, have been implicated to be essential for catalysis (Diacovich*, et al.*, 2004).

3.3.1.7 Putative substrate-binding pocket

The probable pocket for substrate-binding is located perpendicular to the cofactor-binding pocket and the two pockets intersect at the junction of the L-shaped active site [Figure 3-27]. The crystal structure of AccD6 bears a high degree of homology to that of *S. coelicolor* PccB [Table 3-7]; it is hence reasonable to believe that the substrate of AccD6 would bind in a conformation similar to that of propionyl-CoA in the PccB-propionyl-CoA-biotin cocrystal structure (Diacovich*, et al.*, 2004). The substrate analog modeled, in such a conformation, into the active site of AccD6 does not face clashes due to steric effects. This conformation is indeed very similar to those observed for the 12S domain of transcarboxylase (Hall*, et al.*, 2003) and the CT subunit of yeast ACC (Zhang*, et al.*, 2003).

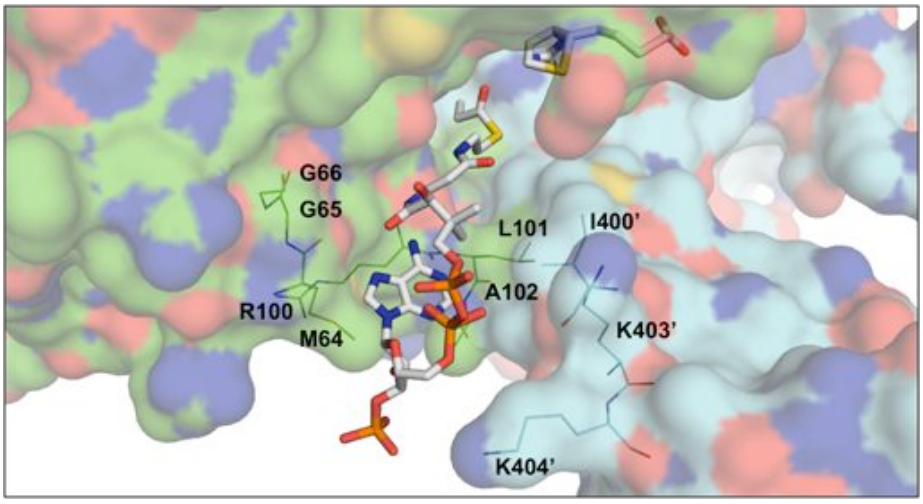

Figure 3-27. The putative substrate-binding pocket. The putative substrate-binding pocket of AccD6 lies perpendicular to the cofactor-binding site and is lined by fully- or partially-conserved residues placed strategically to interact with the different chemical groups of the substrate. In the surface representation of AccD6, carbon, nitrogen, oxygen, and sulphur atoms have been coloured green/cyan, blue, red, and yellow respectively; the surfaces of the monomers have been rendered partially transparent for the sake of clarity. Residues that could potentially be responsible for substrate-binding are shown as 'lines' with the same

colouring scheme. Models of biotin and propionyl-CoA (based on their location in the crystal structure of *S. coelicolor* PccB; PDB ID: 1XNY) are shown in 'stick mode' with carbon, nitrogen, oxygen, phosphorus, and sulphur atoms coloured grey, blue, red, orange, and yellow respectively.

Substrate-binding could be brought about by the concerted action of both the monomers of AccD6, with the N-terminal subdomain contributing most interactions. The partially-conserved residues Val 398', Gly 399', and Ile 400' of the C-terminal subdomain are placed suitably to generate a hydrophobic patch to accommodate the adenine ring of the substrate molecule. The partially conserved, positively charged residues Lys 403' and Lys 404' might be involved in stabilizing the biphosphate moiety of the substrate [Figure 3-27]. These residues are respectively Arg and Gln in AccD5 and Arg and Arg in the *S. coelicolor* homolog PccB [Figure 3-25]. Although these positions are functionally conserved and may contribute equally to the charge-based anchorage of the biphosphate, the differing sizes of the side chains indicate a possibility of subtle variations in the recognition and stabilization of this part of the substrate. As explained in section 3.3.1.4, this region (residues 400 to 430) of the AccD6 protein could not be modeled with confidence in one of the two monomers, presumably because of a high degree of conformational flexibility. This conformational flexibility could potentially be restricted in the presence of the substrate or an analog as well as by the establishment of crystal contacts with adjacent symmetry-related molecules.

The AccD6 structure contains a flexible, conserved patch of mainly non-polar residues that could accommodate the ribose part of the substrate; this patch is composed of the residues Gly 65, Gly 66, Gly 97, Gly 98, Ala 99, Ala 396', and Ala 397' [Figure 3-25]. The pantotheine group of the substrate could be stabilized by the residues Arg 100, Leu 101, and Ala 102. Thus, our structural data suggests that for the regions of the substrate described above, the acyl-CoA binding mode will be conserved.

3.3.1.8 Substrate specificity

To understand the molecular basis of substrate specificity, the active site of AccD6 was compared with those of the propionyl-CoA carboxylase β-subunits *M.*

tuberculosis AccD5 and *S. coelicolor* PccB (Diacovich*, et al.*, 2004). Most residues in the acyl-moiety binding region are conserved, except residue Ile 369 of AccD6 at the end of the substrate-binding pocket. For carboxyltransferases that have a preference for propionyl-CoA as the substrate, this position is occupied by small residues such as a Cys (in AccD5), Asp (in *S. coelicolor* PccB) or Val (in *R. denitrificans* PccB) [Figure 3-25]. The presence of an Ile, a larger and more hydrophobic residue, in AccD6 suggests that the property of the residue at this position may play a major role in defining the shape of the acyl-CoA binding pocket. Interestingly, the corresponding position in the AccD4 sequence is occupied by an Ala residue [Figure 3-25], which will presumably favour the binding of longer, hydrophobic acyl chains, in line with the observation that the homologs of AccD4 carboxylate acyl-chains at least as long as C_{16} (Portevin*, et al.*, 2005).

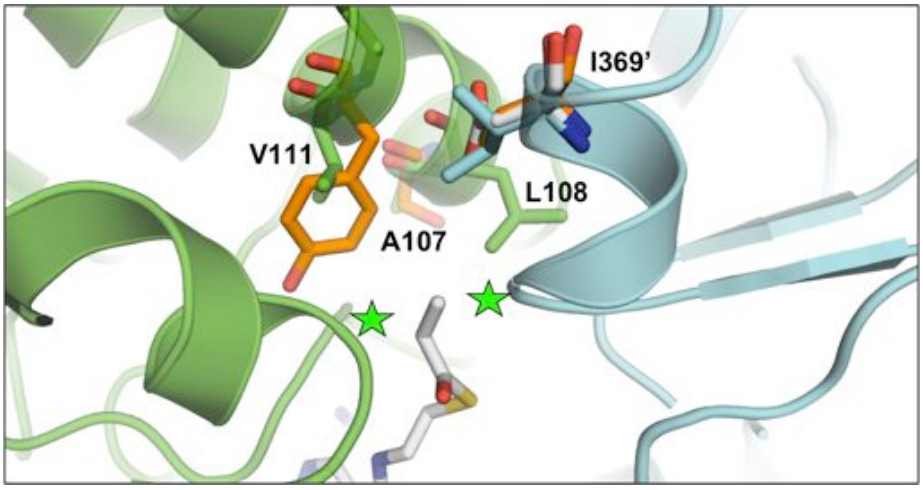

Figure 3-28. Site of substrate specificity determination. The putative determinant of AccD6 substrate specificity, Ile 369', is located at the end of the substrate-binding pocket. Important residues of AccD6 at the acyl-moiety binding region are shown as 'sticks' with carbon, nitrogen, and oxygen atoms coloured green/cyan, blue, and red, respectively. Residues of the AccD5 and *S. coelicolor* PccB structures are represented with the carbon atoms coloured orange and grey, respectively. The position of propionyl-CoA (as in the *S. coelicolor* PccB structure) is indicated, with colouring scheme as in Figure 3-27. The green stars represent the location of the conserved oxyanion holes implicated to be essential for catalysis.

The size, shape and chemical nature of the acyl-moiety-binding region of the putative substrate pocket are defined by a set of conserved residues (such as Leu 108 and Tyr 368') from both the monomers, further emphasizing the importance of the dimeric, di-domain arrangement for substrate recognition, binding, and catalysis. However, a couple of non-canonical features were identified in the proximity of the AccD6 active

site: the otherwise fully conserved residues Ser 163 and Tyr 167 (AccD5 numbering) close to the active site have been replaced by the non-polar Ala 107 and the hydrophobic Val 111 respectively, in the AccD6 structure [Figures 3-25 and 3-28]. These residues are suitably positioned to dictate the chemical environment of the substrate acyl-moiety binding site and therefore, might play direct and important roles in the choice of substrate acyl-chain length.

The entrance to the active site of the enzyme is marked by the presence of two thoroughly conserved oxyanion-stabilizing holes [Figures 3-25 and 3-28], generated by the backbone amide groups of two conserved pairs of residues (Gly 137 – Gly 138 and Gly 366' – Ala 367'). It has been proposed that these oxyanion holes help in the stabilization of the (1) carboxyl group released from the carboxybiotin (brought into the β-subunit active site by the BCCP arm of the α-subunit), (2) ureido enolate resulting from the loss of the carboxyl group from carboxybiotin, (3) the acyl enolate resulting from the deprotonation of the α-carbon of the substrate acyl-CoA (Diacovich*, et al.*, 2004), and hence are absolutely essential for the carboxyltransferase activity.

3.3.2 Discussion

Given the indispensability of fatty acids and mycolic acids in the integrity of the mycobacterial cell wall, and hence in virulence and resistance to drugs and host immunity, enzymes that catalyse the biosynthesis of these cell wall components are of significant biological and therapeutic interest. *M. tuberculosis* acetyl-CoA carboxylase is among the most important of these enzymes as it executes the first committed, rate-limiting reaction in the biosynthetic pathways of cell wall lipids. Surprisingly, full understanding of the three-dimensional structure and of other biophysical properties of this key enzyme has been lacking. In this part of the study, a high-resolution crystal structure of AccD6 was determined and a comparative analysis with structural homologs was carried out.

The crystal structure of AccD6 has shed light on the molecular basis of its unusual quaternary structure. Unlike other actinobacterial carboxyltransferases studied to date, AccD6 was found to exist as a homodimer, both in solution and in the crystal

structure. The oligomeric state of an YCC β-subunit may be related to its function and metabolic regulation (Dakshinamurti & Chauhan, 1988); it was therefore interesting to compare the quaternary structure of AccD6 with the hexameric structures of *M. tuberculosis* AccD5, *S. coelicolor* PccB (Diacovich, *et al*., 2004), and the 12S domain of *P. shermanii* transcarboxylase (Hall, *et al*., 2003), and the dimeric structures of yeast ACC β-subunit (Zhang, *et al*., 2003) and glutaconyl-CoA decarboxylase (Wendt, *et al*., 2003). AccD6 bears a higher degree of sequence and structural homology to the hexameric β-subunits than to the dimeric ones. Although the dimeric, di-domain arrangement is completely conserved in all of these carboxyltransferases, the yeast ACC β-subunit possesses unique structural features that are believed to prevent hexamer formation (Zhang, *et al*., 2003). These additional structural elements at the N- and C-terminal ends of the protein apparently render it unable to form higher oligomers. AccD6 does not contain any homology, sequence or structural, to these additional insertion elements of the yeast ACC β-subunit. Nevertheless, the protein does not hexamerize; the inability of AccD6 to form the canonical hexameric β-subunit ring structure can be attributed to the absence of certain secondary structural elements close to its N- and C-terminal ends, as confirmed by sequence and structural analyses.

Besides reiterating the importance of the dimeric, di-domain interactions and arrangement for carboxyltransferase activity, the crystal structure of AccD6 has shed light on the molecular details of the putative cofactor-binding site and substrate-binding pocket. The structure has also offered insights into the molecular determinants of substrate specificity, i.e., what causes the enzymes to differentiate acetyl-CoA from propionyl-CoA or long-chain acyl-CoA. AccD6 shares high levels of sequence identity with AccD5 and AccD4, but the subunits have differing affinities for different substrates. AccD5 acts on both acetyl-CoA and propionyl-CoA *in vitro*, but has a preference for propionyl-CoA that is many-fold higher than that for acetyl-CoA (Gago, *et al*., 2006). Based on the genomic location and on genetic and biochemical studies in homologs, AccD4 is believed to be the carboxyltransferase acting on long-chain acyl-CoA substrates (Portevin, *et al*., 2005). AccD6 exerts comparable levels of activity on acetyl-CoA and propionyl-CoA *in vitro* (Daniel, *et al*., 2007), but its genomic location and data from knock out studies performed in *M.*

smegmatis have led to the belief that the physiological substrate of AccD6 is acetyl-CoA (Kurth, *et al*., 2009). A comparative analysis of the AccD6 structure highlighted the potential importance of, among other non-canonical features, the residue Ile 369 that likely discriminates acetyl-CoA from longer-chain acyl-CoA substrates. Sequence and structural alignments reveal that the propionyl-CoA carboxylases have less hydrophobic residues at this position, whereas AccD4 has an Ala residue, which accommodates longer, more hydrophobic molecules. Systematic mutational analyses of the *S. coelicolor* PccB have proved that alterations at this position result in changes in the size, shape, and electrostatic property of the acyl-CoA binding pocket (Arabolaza, *et al*., 2010). It has further been shown that alterations at this key position can cause long-range conformational changes that influence protein-CoA phosphate binding, a phenomenon critical for the activity of CoA-binding enzymes (Arabolaza, *et al*., 2010). It is hence justifiable to believe that Ile 369 is the major determinant of AccD6 substrate specificity.

The fact that AccD6 is the only actinobacterial β-subunit, so far, to be characterised as a homodimer, might have implications for the possible molecular evolution of these enzymes. The active site architecture and probable mechanism of action are conserved in all carboxyltransferases, but the quaternary structures have been subjected to variation. Although there is a reasonable understanding of the carboxyltransferase mechanism of action, the possibility of enzyme cooperativity in the catalytic activity of carboxyltransferases has not been investigated so far; the molecular rationale behind assembly into a 400-kDa hexamer remains to be understood. It is indeed intriguing that certain members of an enzyme family assemble into a large oligomeric structure while another member inherently lacks the ability to form such oligomers and yet retains comparable levels of catalytic activity. It would be of biological and biochemical interest to further explore the molecular basis for these differences.

4 Conclusions and Future Perspectives

The main objective of this study was to gain insights into the three-dimensional architecture of *Mycobacterium tuberculosis* acyl-CoA carboxylases, potential drug targets. The propionyl-CoA carboxylase AccA3-AccD5 of *M. tuberculosis* was reconstituted *in vitro* by expression in the near-native host *M. smegmatis*. However, the carboxyltransferase subunit AccD5 crystallised preferentially over AccA3-AccD5 / AccA3, indicating possible instability of the complex. To overcome the problem of biochemical heterogeneity arising from the interaction of *M. tuberculosis* YCC subunits with the highly homologous endogenous *M. smegmatis* counterparts, *E. coli* was resorted to as an expression host; various co-expression and co-purification strategies were attempted for the reconstitution of the YCC complexes AccA3-AccD5 and AccA3-AccD4. Unfortunately, structural data on a holo-YCC complex could not be acquired during the study. AccE5, the purported modulator of PCC activity, was found to be an intrinsically disordered protein; its association with its binding partners AccA3 and AccD5 could not be probed into due to technical constraints.

Another aim of the work was to understand the molecular basis of the differences in substrate specificity of the individual carboxyltransferase subunits of *M. tuberculosis.* Biophysical investigation of these proteins revealed unexpected diversity in carboxyltransferase subunit quaternary structure. AccD4, the β-subunit that is implicated in long-chain acyl-CoA carboxylation, was found to exist as a homotrimer; preliminary small angle scattering analysis suggested arrangement of AccD4 protomers in a fashion different from those observed in the homohexameric β-subunits characterised so far. Efforts to confirm and validate the SAXS-generated model are ongoing.

The high-resolution crystal structure of AccD6, the carboxyltransferase of *M. tuberculosis* acetyl-CoA carboxylase, offered snapshots of the putative active site, cofactor-binding site, and substrate-binding pocket of the enzyme. The structural data helped elucidate the molecular basis of the unusual homodimeric arrangement of AccD6. Comparative analysis with structural homologs identified residues in the proximity of the active site that could be responsible for determining variations in

substrate preference. The crystal structure(s) of AccD6, in complex with analogs of substrate and/or cofactor, would be useful in further understanding the modes of binding and the mechanism of carboxyltransferase activity. High-resolution structural data on AccD4 would complement such an understanding, besides providing a molecular explanation as to how these closely related enzymes utilize their three-dimensional structures to discriminate between substrates of different chemical natures.

5 Appendix

Appendix I: Materials for molecular cloning

Table A-1. Bacterial expression vectors used in the study

Vector	Promoter	Selection	Tag	Protease cleavage site	Origin of replication	Source	Description
pCDF-11	T7/lacO	Spec	N-ter His	TEV	CDF	A. Geerlof	Vector for co-expression
pCDF-13	T7/lacO	Spec	None	None	CDF	A. Geerlof	Vector for co-expression
pET-28a	T7/lac	Kan	N-ter His	Thrombin	pBR322	Novagen	Expression vector
pnEK	T7/lac	Kan	None	None	pET-28b	C. Romier	Vector for co-expression
pET-M11	T7/lac	Kan	N-ter His	TEV	pBR322	G. Stier	Contains MAD insert in modified pET-24d
pET-M11/LIC	T7/lac	Kan	N-ter His	TEV	pBR322	A. Geerlof	pET-M11 modified for LIC
pET SUMO/LIC	T7/lac	Kan	N-ter His N-ter SUMO	TEV	pBR322	H. Meyerhofer, A. Geerlof	pET-24d modified for LIC
pET Z2-1a/LIC	T7/lac	Kan	N-ter His N-ter Z tag2	TEV	pBR322	A. Geerlof	pET-24d modified for LIC
pMyNT	Acetamidase	Hyg	N-ter His	TEV	pSD24	A. Geerlof	Mycobacterial shuttle vector

Table A-2. Bacterial strains used in the study

Strain	Description	Application	Antibiotic resistance	Source
E. coli DH5α-T1R	T1 phage-resistant	Non-expression host for plasmid propagation	None	Life Technologies
E. coli BL21 (DE3)	Deficient in *Ion* and *ompT* proteases	General purpose expression host	None	Novagen
E. coli BL21 (DE3) CodonPlus-RP	Deficient in *Ion* and *ompT* proteases; contains plasmid encoding *argU* and *proL*	Expression of genes encoding tRNAs for rare Arg and Pro codons	Cam	Stratagene
E. coli BL21 star (DE3)	RNaseE (*rne131*) mutant	General purpose expression host with reduced mRNA degradation	None	Invitrogen
E. coli BL21 star (DE3) pRare2	RNaseE mutant; contains plasmid encoding *argU, argW, argX, glyT, ileX, leuW, metT, proL, thrT, thrU* and *tyrU*	General purpose expression host with reduced mRNA degradation; expression of genes encoding tRNAs for several rare codons	Cam	A. Geerlof
M. smegmatis groEL1ΔC	Deficient in His-rich coding sequence at C-terminus	Efficient expression and purification of His-tagged mycobacterial proteins	None	(Noens, *et al.*, 2011)

Table A-3. List of primers

Primer ID	**Primer sequence (5' → 3')**	**Purpose**
MTE_accD5-E5_forward	GATCTCATGAGTATGACAGCGTTACCG	Cloning of *accD5-accE* into pMyNT (between *NcoI* and *HindIII* sites)
MTE_accD5-E5_reverse	GTCAAAGCTTTTATCGGCGCATGTGCG	
MTE_accA3_forward	GATCCCATGGGTATGGCTAGTCACGCC	Cloning of *accA3* into pMyNT/ pCDF-11/ pCDF-13 (between *NcoI* and *HindIII* sites)
MTE_accA3_reverse	GTCAAAGCTTTTACTTGATCTCGGCGAGC	
MA_E5_LIC_F	CAGGGCGCCATGATGGGAACGTGC C	LIC of *accE* into various vectors for expression in *E. coli*
MA_E5_LIC_R	GACCCGACGCGGTAATCATCGGCGCAT	
MA_D5_LIC_F	CAGGGCGCCATGATGACAAGCGTTACCGA	LIC of *accD5* into various vectors for expression in *E. coli*
MA_D5_LIC_R	GACCCGACGCGGTAATCACAGGGGCACGT	
MAP07_D5_pCDF11F	GATCACATGTAATGACAAGCGTTACCG	Cloning of *accD5* into pCDF-11/ pCDF-13 (between *PciI* and *HindIII* sites)
MAP02_D5_pCDF13R	CATCAAGCTTTCACAGGGGCACGTT	
MAP08_E5_pCDF11F	AAAACCATGGGATGGGAACGTGCCCC	Cloning of *accE* into pCDF-11/ pCDF-13 (between *NcoI* and *HindIII* sites)
MAP04_E5_pCDF13R	GATCAAGCTTTCATCGGCGCATGTGC	
MAP09_D4_pCDF11F	CGATCCATGGTATGGTGACCGTCACCG	Cloning of *accD4* into pCDF-11/ pCDF-13 (between *NcoI* and *HindIII* sites)
MAP06_D4_pCDF13R	GACCAAGCTTCTAGACCGGGATCAGGCC	
MAP14_A3dBCCP1_R	ATATAAGCTTCTAGATCCAGCGGG	Together with primer *MTE_accA3_forward*, cloning of *accA3 ΔBCCP* versions into pET-M11/ pCDF-13 (between *NcoI* and *HindIII* sites)
MAP15_A3dBCCP2_R	ATATAAGCTTCTAGAGAGGTTCGCCG	
MAP16_A3dBCCP3_R	ATATAAGCTTCTACTTCTGACGCGGC	
MAP17_A3d_LICF	CAGGGCGCCATGGTGGCTAGTCACG	

MAP18_A3d1_LICR	GACCCGACGCGGTAAGATCCAGCGGGTAT	LIC of *accA3 ΔBCCP* versions into various vectors for expression in *E. coli*
MAP19_A3d2_LICR	GACCCGACGCGGTAAGAGAGGTTCGCCGT	
MAP20_A3d3_LICR	GACCCGACGCGGTAACTTCTGACGCGG	
MAP41_D5_MCN_F	GATCCATATGCAATGACAAGCGTTACCGACCGC	Cloning of *accD5* into pnEK (between *NdeI* and *AflII* sites)
MAP42_D5_MCN_R	GATCCTTAAGTTATCACAGGGGCACGTTCCCATGC	
MAP45_D4_MCN_F	CTAGCATATGCAATGGTGACCGTCACCGAGCCGG	Cloning of *accD4* into pnEK (between *NdeI* and *AflII* sites)
MAP46_D4_MCN_R	GTACCTTAAGTTACTAGACCGGGATCAGGCCGTGC	
MAP47_D6_MCN_F	CGGACATATGCAATGACAATCATGGCCCCCGAGGCGG	Cloning of *accD6* into pnEK (between *NdeI* and *AflII* sites)
MAP48_D6_MCN_R	CATGCTTAAGCTACAGCGGGATGTTCTTGTGGCG	

Appendix II: Microbial growth media and electrophoresis reagents

7H9 complete medium

4.7 g Middlebrook 7H9 medium powder was dissolved in 900 ml water and autoclaved (121 °C at 2 bar for 20 min). 100 ml ADS solution (13.9 mM NaCl, 0.5% (w/v) BSA, 0.2% (w/v) glucose), 4 ml of 50% (v/v) glycerol and 2.5 ml of 20% (v/v) Tween 80 were added. The complete medium was filter-sterilized and stored at 4°C.

7H9 expression medium

4.7 g Middlebrook 7H9 medium powder was dissolved in 1000 ml water and autoclaved. Just before use, 10 ml of 20% (w/v) glucose, 4 ml of 50% (v/v) glycerol and 2.5 ml of 20% (v/v) Tween 80 were added.

7H10 agar

19 g Middlebrook 7H10 agar was dissolved in 900 ml water and autoclaved and cooled. 100 ml ADS solution (13.9 mM NaCl, 0.5% (w/v) BSA, 0.2% (w/v) glucose), 10 ml of 50% (v/v) glycerol, 2.5 ml of 20% (v/v) Tween 80 and 1 ml of hygromycin (50 mg/ml commercial available stock) were added just before pouring the medium into plastic Petri plates.

Acetamide solution (220X)

20 g acetamide was dissolved in 31 ml water, filter sterilized and stored at 4 °C.

Agarose gel electrophoresis-running buffer (5X)

Tris base	54 g
Boric acid	27.5 g
0.5 mM EDTA (pH 8.0)	20 ml
ddH_2O	to 1000 ml

Antibiotics

1000X Kanamycin	50 mg/ml in water
1000X Chloramphenicol	34 mg/ml in ethanol

The above stock solutions were filter-sterilized and stored in aliquots at -20 °C.

Coomassie blue safe stain

To about 500 ml ddH_2O, 0.8 g Brilliant Blau G-250 was added and dissolved by stirring overnight. 34 ml of 32% (v/v) HCl was added and the volume was made up to 1000 ml with ddH_2O. The staining solution was stored at room temperature, protected from light.

DNA sample buffer (6X)

0.5 M EDTA (pH 8.0)	5 ml
Glycerol	15 ml
Bromophenol blue	125 mg
ddH_2O	to 50 ml

The buffer was filtered and frozen until use.

Luria broth (LB)

20 g LB granulated powder was dissolved in distilled water to a final volume of 1000 ml and sterilized by autoclaving.

Luria broth-agar (LB agar)

37 g LB agar granulated powder was dissolved in distilled water to a final volume of 1000 ml and sterilized by autoclaving.

Super optimal broth (SOB)

Tryptone	20 g
Yeast extract	5 g
NaCl	0.59 g
KCl	0.19 g

The above were dissolved in 980 ml ddH_2O, autoclaved and added with 20 ml of sterile 1 M $MgCl_2$ after cooling.

Super optimal catabolite repression medium (SOC)

Yeast extract	2.5 g
Tryptone	10 g
5 M NaCl	1 ml

3 M KCl	416 µl
1 M $MgCl_2$	5 ml
1 M $MgSO_4$	5 ml
ddH_2O	to 495 ml

The above solution was autoclaved, cooled, added with 5 ml of filter-sterilized 1 M glucose and stored at 4°C.

SDS-PAGE gel recipes (for Bio-Rad minigels)

Resolving gels

Component	7.5%	12%	15%
1.5 M Tris-HCl (pH 8.8)	2.5 ml	2.5 ml	2.5 ml
ddH_2O	5.34 ml	4.14 ml	3.34 ml
10% SDS	0.1 ml	0.1 ml	0.1 ml
40% acrylamide	2 ml	3.2 ml	4 ml
20% APS	0.05 ml	0.05 ml	0.05 ml
TEMED	0.01 ml	0.01 ml	0.01 ml

Stacking gels

Component	4%
0.5 M Tris-HCl (pH 6.8)	1.25 ml
ddH_2O	3.14 ml
10% SDS	0.05 ml
40% acrylamide	0.53 ml
20% APS	0.025 ml
TEMED	0.005 ml

SDS-PAGE running buffer (10X)

Tris base	151.4 g
Glycine	720 g
SDS	50 g
ddH_2O	to 5000 ml

SDS-PAGE sample buffer (4X)

NuPAGE® LDS sample buffer (Invitrogen) supplemented with 10% (v/v) β-ME.

Terrific broth (TB[#])

47.6 g TB granulated powder and 4 ml glycerol were dissolved in distilled water to a final volume of 1000 ml and sterilized by autoclaving.

Transformation buffer (TB[§])

PIPES	0.34 g
$CaCl_2.6H_2O$	0.33 g
$KCl.4H_2O$	1.83 g

The above were dissolved in 85 ml ddH_2O and pH was adjusted to 6.7 with KOH; the solution was added with 5.5 ml of 1 M $MnCl_2.4H_2O$ and the volume was adjusted to 100 ml with ddH_2O; the buffer was filter-sterilized and cooled until use.

Appendix III: Buffers for protein purification

All buffers were prepared using Milli-Q grade water deionized to a resistance of 18.2 mW cm^{-1} by the Milli-Q system (Millipore). The pH of all buffers was checked using a S20-K SevenEasy™ pH meter (Mettler Toledo). Each SEC buffer was filtered through a 0.22 μm cellulose membrane filter and degassed before use.

Stock solutions

(i) Buffers and salts

HEPES (1 M)

238.30 g HEPES was dissolved and made up to a volume of 1000 ml in ddH_2O, after adjustment to the required pH.

Imidazole (2 M)

136.15 g imidazole was dissolved in ddH_2O to a final volume of 1000 ml.

NaCl (3 M)

175.32 g sodium chloride was dissolved in ddH_2O to a final volume of 1000 ml.

KCl (3 M)

223.65 g potassium chloride was dissolved in ddH_2O to a final volume of 1000 ml.

Tris-HCl (1 M)

121.13 g Tris base was dissolved and made up to a volume of 1000 ml in ddH_2O, after adjustment to the required pH.

(ii) Other additives

Glycerol, 1-thioglycerol, NP-40 and β-ME were diluted as per requirement from commercially available stock concentrations (99% (v/v), 99% (v/v), 100% (v/v), 14.3 M respectively). DTT and TCEP powders were added to buffers immediately prior to filtration and usage.

Table A-4. Optimal expression conditions and composition of buffers used for protein purification

Protein	Expression host	Expression conditions	Buffers for purification
AccD5 (Construct M09)	*M. smegmatis groEL1ΔC*	1X Acetamide 37°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 0.2% (v/v) NP-40, 0.02% (v/v) 1-thioglycerol (TG)[#] **Binding:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **Wash:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 50 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **Elution:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 300 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **SEC:** 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP[#]
AccA3 (Construct M03)	*M. smegmatis groEL1ΔC*	1X Acetamide 37°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 0.2% (v/v) NP-40, 0.02% (v/v) 1-thioglycerol[#] **Binding:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **Wash:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 50 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **Elution:** 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 300 mM imidazole, 0.01% (v/v) 1-TG[#], 20% (v/v) glycerol **SEC:** 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP[#]

AccE5 (Construct M11)	*E. coli* BL21 (DE3) / BL21 (DE3) CodonPlus-RP	0.5 mM IPTG 21°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.2% (v/v) NP-40, 1 mM β-ME[#] **Binding:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, 1 mM β-ME[#] **Wash:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 50/75/100 mM imidazole, 10% (v/v) glycerol, 1 mM β-ME[#] **Elution:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol, 1 mM β-ME[#] **SEC:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5% (v/v) glycerol, 1 mM β-ME [#]
AccD4 (Construct M06)	*E. coli* BL21 (DE3)	0.5 mM IPTG 21°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.2% (v/v) NP-40 **Binding:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole, 10% glycerol **Wash:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 50/75 mM imidazole, 10% glycerol **Elution:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 250 mM imidazole, 10% glycerol **Dialysis prior to methylation:** 50 mM HEPES (pH 7.5), 150 mM NaCl **SEC:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5% glycerol, 2 mM DTT [#]
AccD5 (Construct M08)	*E. coli* BL21 star (DE3) pRare2	0.25 mM IPTG 18°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.2% (v/v) NP-40, 0.02% (v/v) 1-TG[#] **Binding:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole, 0.01% (v/v) 1-TG[#] **Wash:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 50 mM imidazole, 0.01% (v/v) 1-TG[#] **Elution:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 300 mM imidazole, 0.01% (v/v) 1-TG[#] **SEC:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 2% (v/v) glycerol, 2 mM DTT[#]

AccA3 (Construct M02)	*E. coli* BL21 star (DE3) pRare2	0.3 mM IPTG 30°C 16 hours	**Lysis:** 50 mM HEPES (pH 7.5), 300 mM KCl, 0.2% (v/v) NP-40, 5% (v/v) glycerol, 1 mM β-ME[#] **Binding:** 50 mM HEPES (pH 7.5), 300 mM KCl, 20 mM imidazole, 5% (v/v) glycerol, 1 mM β-ME[#] **Wash:** 50 mM HEPES (pH 7.5), 300 mM KCl, 50/75 mM imidazole, 5% (v/v) glycerol, 1 mM β-ME[#] **Elution:** 50 mM HEPES (pH 7.5), 300 mM KCl, 300 mM imidazole, 5% (v/v) glycerol, 1 mM β-ME[#] **SEC:** 50 mM HEPES (pH 7.5), 300 mM KCl, 5% (v/v) glycerol, 1 mM TCEP[#]
AccD6 (Construct M27)	*E. coli* BL21 (DE3)	0.25 mM IPTG 18°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM imidazole, 0.02% (v/v) 1-TG[#] **Binding:** 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole, 0.01% (v/v) 1-TG[#] **Wash:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 50 mM imidazole, 0.01% (v/v) 1-TG[#] **Elution:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 300 mM imidazole, 0.01% (v/v) 1-TG[#] **SEC:** 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 2% glycerol, 2 mM DTT [#]
AccA3-AccD5 (co-expressed) (Constructs M02-M15, M17-M08)	*E. coli* BL21 star (DE3)	0.5 mM IPTG 21°C 16 hours	**Lysis:** 25 mM HEPES (pH 7.0), 100 mM KCl, 0.2% (v/v) NP-40, 2 mM β-ME[#] **Binding:** 25 mM HEPES (pH 7.0), 100 mM KCl, 20 mM imidazole, 2 mM β-ME[#] **Wash:** 25 mM HEPES (pH 7.0), 250 mM KCl, 50 mM imidazole, 2 mM β-ME[#] **Elution:** 25 mM HEPES (pH 7.0), 200 mM KCl, 250 mM imidazole, 5% (v/v) glycerol,

			2 mM β-ME[#] **SEC:** 25 mM HEPES (pH 7.0), 250 mM KCl, 5% (v/v) glycerol, 2 mM DTT[#]
AccA3-AccD4 (co-expressed) (In collaboration with Lionex GmbH)	*E. coli* BL21 (DE3)	0.5 mM IPTG 21°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 7.2), 150 mM NaCl, 10 mM imidazole, 0.2% (v/v) NP-40, 0.01% TG[#] **Binding:** 50 mM Tris-HCl (pH 7.2), 300 mM NaCl, 20 mM imidazole, 0.01% TG[#] **Wash:** 50 mM Tris-HCl (pH 7.2), 300 mM NaCl, 50 mM imidazole, 0.01% TG[#] **Elution:** 50 mM Tris-HCl (pH 7.2), 150 mM NaCl, 300 mM imidazole, 0.01% TG[#] **SEC:** 50 mM Tris-HCl (pH 7.2), 100 mM NaCl, 2 mM DTT[#]
AccA3-AccD6 (co-expressed) (In collaboration with Lionex GmbH)	*E. coli* BL21 (DE3)	0.5 mM IPTG 21°C 16 hours	**Lysis:** 50 mM Tris-HCl (pH 7.2), 150 mM NaCl, 10 mM imidazole, 0.2% (v/v) NP-40, 0.01% thioglycerol **Binding:** 50 mM Tris-HCl (pH 7.2), 300 mM NaCl, 20 mM imidazole, 0.01% TG[#] **Wash:** 50 mM Tris-HCl (pH 7.2), 300 mM NaCl, 50 mM imidazole, 0.01% TG[#] **Elution:** 50 mM Tris-HCl (pH 7.2), 150 mM NaCl, 300 mM imidazole, 0.01% TG[#] **SEC:** 50 mM Tris-HCl (pH 7.2), 100 mM NaCl, 2 mM DTT[#]

[#] Added just before use

Appendix IV: Structural characterisation of a D-isomer specific 2-hydroxyacid dehyrogenase from *Lactobacillus delbrueckii* ssp. *bulgaricus*

Background

Hydroxyacid dehydrogenases (HDHs) catalyse the stereospecific conversion of 2-keto acids to 2-hydroxyacids and are useful enzymes in the biotechnology industry (Hummel, 1999); chiral hydroxyacids are useful as synthons, as precursors for valuable chiral products and as flavour enhancing agents in dairy products. HDHs belong to either of two evolutionarily distinct subclasses: the D- and L-isomer specific HDHs, whose products possess similar biochemical properties but differ in chirality (Kochhar*, et al.*, 1992). The genome of *Lactobacillus delbrueckii* ssp. *bulgaricus*, an economically important lactic acid producing bacterium, codes for several of these enzymes whose structural and functional characteristics remain poorly understood (van de Guchte*, et al.*, 2006). Given their high economic potential, it is of significant scientific interest to gain in-depth knowledge of the substrate specificities, reaction kinetics and molecular mechanisms of action of D-isomer specific 2-HDHs.

Author's contribution

Alongside work related to the main topic of the thesis, effort was invested in the structural characterisation of a D-isomer specific 2-HDH (encoded by *Ldb1010*) of *L. bulgaricus*, both in its apo form and in complex with the cofactor NAD^+. Contribution made by the author of the thesis to this characterisation was in the form of literature review, crystallographic model refinement, structural analysis, and manuscript preparation. This work (reprint attached) is the first report of the three-dimensional structures of *L. bulgaricus* D2-HDH, in apo- and cofactor-bound forms. The asymmetric unit contained a dimer of dimers of D2-HDH, related by non-crystallographic symmetry. The structures revealed the presence of conventional substrate- and cofactor-binding domains, conserved cofactor-binding sites and in addition, non-canonical features of the putative substrate-binding site that might dictate substrate-specificity. Results of the study add to existing knowledge in the largely unexplored structural proteome of *L. bulgaricus*.

Journal of Structural Biology 181 (2013) 179–184

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Structure Report

Structural characterization of a D-isomer specific 2-hydroxyacid dehydrogenase from *Lactobacillus delbrueckii* ssp. *bulgaricus*

Simon J. Holton *, Madhankumar Anandhakrishnan, Arie Geerlof [1], Matthias Wilmanns

EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

ARTICLE INFO

Article history:
Received 19 July 2012
Received in revised form 29 September 2012
Accepted 1 October 2012
Available online 27 October 2012

Keywords:
Lactobacillus delbrueckii ssp. *bulgaricus*
Lactic acid
NAD^+-dependent dehydrogenase
X-ray crystallography

ABSTRACT

Hydroxyacid dehydrogenases, responsible for the stereospecific conversion of 2-keto acids to 2-hydroxy-acids in lactic acid producing bacteria, have a range of biotechnology applications including antibiotic synthesis, flavor development in dairy products and the production of valuable synthons. The genome of *Lactobacillus delbrueckii* ssp. *bulgaricus*, a member of the heterogeneous group of lactic acid bacteria, encodes multiple hydroxyacid dehydrogenases whose structural and functional properties remain poorly characterized. Here, we report the apo and coenzyme NAD^+ complexed crystal structures of the *L. bulgaricus* D-isomer specific 2-hydroxyacid dehydrogenase, D2-HDH. Comparison with closely related members of the NAD-dependent dehydrogenase family reveals that whilst the D2-HDH core fold is structurally conserved, the substrate-binding site has a number of non-canonical features that may influence substrate selection and thus dictate the physiological function of the enzyme.

1. Introduction

NAD-dependent hydroxyacid dehydrogenases, key players in carbohydrate metabolism, catalyse the reversible conversion of keto acids into chiral hydroxy acids, concomitant with the oxidation of NADH to NAD^+ (Hummel, 1999). These enzymes belong to either of two evolutionary distinct subfamilies: the D- and L-isomer specific hydroxyacid dehydrogenases, whose products differ only in their chirality (Kochhar et al., 1992b). The genome of *Lactobacillus delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*), a member of the heterogeneous group of lactic-acid producing bacteria, harbors five genes (*Ldb0101*, *Ldb1010*, *Ldb0813*, *Ldb1147* and *Ldb2021*) that are predicted to encode different D-isomer specific 2-hydroxyacid dehydrogenases (van de Guchte et al., 2006). The protein sequences share only low sequence identities (27–37%). An additional two genes encoding putative L-isomer specific 2-hydroxyacid dehydrogenases (*Ldb0120, Ldb0094*) have also been reported (van de Guchte et al., 2006).

In this contribution we have characterized the *L. bulgaricus* (strain ATCC 11842) *Ldb1010* gene product. The resulting 333-residue protein, D2-HDH, is annotated to encode a D-isomer specific 2-hydroxyacid dehydrogenase (van de Guchte et al., 2006). Although the D2-HDH protein has not directly been characterized beyond this automatic functional annotation, a 99% homologous enzyme has been sequenced, cloned and its substrate specificity characterized (Bernard et al., 1994). The protein, 2-hydroxyisocaproate dehydrogenase (HO-HxoDH), is virtually identical to the D2-HDH, with only three amino-acid differences between the two proteins, all at sites not known to be biologically relevant (Suppl. Fig. 1). We therefore assume that functional and structural conclusions are the same for both enzymes.

Given the potential use of chiral hydroxy acids as valuable synthons and as precursors for useful chiral products in the biotechnology industry (Hummel, 1999) it is of significant biological interest to gain in-depth insights into the substrate specificities, biochemical kinetics and molecular mechanisms of action of D-isomer specific 2-hydroxyacid dehydrogenase enzymes. As a step towards this goal we report in this study the recombinant expression, purification and subsequent three-dimensional structural characterization of the *L. bulgaricus* (strain ATCC 11842) D-isomer specific 2-hydroxyacid dehydrogenase (D2-HDH, encoded by *Ldb1010*), both in its apo form and in complex with the cofactor NAD^+. A comparison of the D2-HDH structural features with those of related proteins is presented.

Abbreviations: D-HicDH, D-2-hydroxyisocaproate dehydrogenase from *L. casei*; D2-HDH, D-isomer specific 2-hydroxyacid dehydrogenase; FDH, formate dehydrogenase ; HGDH, (R)-2-hydroxyglutarate dehydrogenase from *Acidaminococcus fermentans*; HO-HxoDH, 2-hydroxyisocaproate (i.e. 2-hydroxy-4-methyl-pentanoate) dehydrogenase from *L. bulgaricus*; IPTG, Isopropyl β-D-1-thiogalactopyranoside; *L. bulgaricus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*; LDH, lactate dehydrogenase; NAD, nicotinamide adenine dinucleotide; PCR, Polymerase Chain Reaction; PGDH, D-3-phosphoglycerate dehydrogenase from *E. coli*.

* Corresponding author. Present Address: Bayer Pharmaceuticals, Structural Biology/Lead Discovery, D-13342 Berlin, Germany. Fax: +49 30 468 993724.
E-mail address: sjholton@googlemail.com (S.J. Holton).
[1] Present address: Helmholtz Zentrum Muenchen, Institute of Structural Biology, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany.

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2012.10.009

2. Purification and enzymatic activity of D2-HDH

The *Ldb1010* gene coding the full-length D2-HDH (Gene ID: 4084222) was PCR-amplified from *L. bulgaricus* (strain ATCC 11842) chromosomal DNA and ligated into the pETM-13/LIC vector, which harbors a non-cleavable C-terminal His_6-tag, via a ligase-independent cloning strategy. His-tagged D2-HDH was then expressed heterologously in *E. coli* BL21 (DE3) pLysS cells cultured at 37 °C in Terrific Broth (TB) medium supplemented with 30 μg ml^{-1} kanamycin and 10 μg ml^{-1} chloramphenicol. Expression of D2-HDH was induced at OD_{600nm} of 0.9 by the addition of a final concentration of 0.4 mM IPTG. The cells were incubated for 16 h at 25 °C and then harvested by centrifugation, resuspended in lysis buffer (40 mM HEPES pH 7.4, 300 mM NaCl, 20 mM imidazole, 0.02% (v/v) monothioglycerol (MTG)) supplemented with protease inhibitors and stored at −80 °C. Cells were lysed by freeze–thawing, DNaseI treatment and pulse sonication. Cell debris was removed by centrifugation and the soluble cellular fraction was loaded onto Ni–NTA resin pre-equilibrated with lysis buffer. After extensive washing of the resin with lysis buffer the protein was eluted using elution buffer (40 mM HEPES pH 7.4, 300 mM NaCl, 300 mM imidazole, 0.02% (v/v) MTG). The eluate was concentrated using a Vivaspin 20 (10 kDa cutoff, GE Healthcare) and loaded onto a Superdex200 HiLoad (26/60) column (Pharmacia) pre-equilibrated with gel filtration buffer (40 mM HEPES pH 7.4, 300 mM NaCl, 0.02% (v/v) MTG). The recombinantly prepared D2-HDH protein, which includes a non-native C-terminal hexahistidine tag and a 4-residue linker (Thr Ala Ser Gly linker), is enzymatically active (Suppl. Fig. 2).

The 2-hydroxyacid dehydrogenase activity of D2-HDH on the substrate phenylpyruvate follows Michaelis–Menten kinetics, with a K_m of 59.1 ± 14.4 μM and a k_{cat} of 8.48 ± 0.81 s^{-1}. These results are in the same order of magnitude as those published for HO-HxoDH (Bernard et al., 1994) supporting our expectation that the high sequence identity shared between the two proteins translates to similar biochemical properties. Interestingly, the activity of D2-HDH also seemed to be inhibited by high concentrations of the substrate (Suppl. Fig. 2) with a dissociation constant K_i of 1.21 ± 0.26 mM. Substrate inhibition has also been reported for HO-HxoDH (Bernard et al., 1994).

3. Crystallization, data collection and structure determination of apo-D2-HDH and D2-HDH-NAD$^+$ complex

Diffraction-quality crystals of apo-D2-HDH were obtained using the hanging-drop vapor-diffusion method by mixing equal volumes (400 nl) of protein (10 mg/ml) and reservoir solution (25% PEG 3350, 200 mM $MgCl_2$, 100 mM HEPES pH7.5), equilibrated against 100 μl reservoir solution at 19 °C. D2-HDH-NAD$^+$ crystals were generated by soaking apo crystals overnight with mother liquor supplemented with 10 mM NAD$^+$ (Li-salt). Crystals were briefly transferred to a cryo-protectant solution (mother liquor supplemented with 30% ethylene glycol) before being flash-frozen at 100 K. The apo-D2-HDH and NAD$^+$-D2-HDH complex diffraction datasets were collected at the European Synchrotron Radiation Facility beamline ID23-1 at 100 K. The data were integrated, scaled and merged using the programs *MOSFLM* and *SCALA* (CCP4, 1994). Data collection statistics are shown in Table 1.

The structure of D2-HDH was solved by molecular replacement using the program *Phaser* (McCoy, 2007). The monomer structure of D-2-hydroxyisocaproate dehydrogenase (D-HicDH) from *Lactobacillus casei* (PDB ID: 1DXY (Dengler et al., 1997)) was used as the search model. Unbiased electron density features indicated a correct molecular replacement solution. An initial model was built using the automated model rebuilding protocols in the program *Phenix* (Adams et al., 2010). Thereafter, alternative rounds of manual rebuilding using *COOT* (Emsley et al., 2010) and maximum-likelihood refinement (*REFMAC5* (Murshudov et al., 1997)) were implemented. Non-crystallographic symmetry (NCS) restraints were applied during restrained refinement and, in the later stages, TLS parameters were also used. Crystallographic refinement and characteristics of the final models are summarized in Table 1. The electron density for residues 92 and 93 in protomer A, 91–93 in protomer B, 91–93, 325 and 326 in protomer D was weak and so these residues were not modeled. The refined apo-D2-HDH and NAD$^+$-complexed D2-HDH structures have been deposited in the Protein Data Bank with the accession codes **2yq4** and **2yq5,** respectively.

Table 1
Crystallographic statistics.

	Native	Ligand-bound
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	92.21, 108.37, 147.18	92.72, 110.01, 147.74
Wavelength (Å)	0.9334	0.8726
Overall resolution range (Å)	46.72–3.45	49.24–2.75
Highest resolution range (Å)	3.64–3.45	2.90–2.75
Number of unique reflections	19722 (2872)[d]	39722 (5742)[d]
Multiplicity	3.1 (3.2)	4.7 (4.8)
Mean I/σ(I)	8.0 (2.2)	12.7 (2.6)
Completeness (%)	98.9 (99.7)	99.5 (99.9)
R_{sym}[a]	0.167 (0.537)	0.105 (0.556)
Mosaicity (°)	0.84	0.51
Refinement		
Protein atoms	10154	10162
Other atoms	34	207
R_{conv}[b]/R_{free}[c]	0.1828/0.2496	0.1952/0.2634
Rms deviations		
Bond lengths (Å)	0.020	0.015
Bond angles (°)	1.836	1.626

[a] $R_{sym} = \sum_h \sum_j |I_{h,j} - \langle I_h \rangle| / \sum_h \sum_j I_{h,j}$ where $I_{h,j}$ is the intensity of the jth observation of unique reflection h.

[b] $R_{conv} = \sum_h ||Fo_h| - |Fc_h|| / \sum_h |Fo_h|$ where Fo_h and Fc_h are the observed and calculated structure factor amplitudes for reflection h.

[c] R_{free} is equivalent to R_{conv}, but is calculated using a 5% disjoint set of reflections excluded from the maximum likelihood refinement stages.

[d] Values in parentheses correspond to the highest resolution shell.

4. Overall structure of apo-D2-HDH

The overall monomeric structure of the *L. bulgaricus* D2-HDH can be described in terms of two compact domains separated by a deep cleft (Fig. 1). This separation of the substrate binding and the co-factor binding domains is conserved in other D2-hydroxycarboxylate dehydrogenases (Razeto et al., 2002). The asymmetric unit contains two homo D2-HDH dimers and is consistent with our observation that the protein can form a dimeric species in solution (Suppl. Fig. 3). The two protomers adopt an elongated structure with dimensions of 109 Å × 56 Å × 48 Å. The 2740 Å^2 interface between the two protomers is formed from residues that mainly originate from the co-factor domain and comprises of predominantly hydrophobic interactions with an additional number of hydrogen-bonding interactions. Solution studies also identified a tetrameric D2-HDH solution species that may correspond to a dimer of dimers (Suppl. Fig. 3).

The co-factor binding domain, formed from residues 109–303, has a central β-sheet containing six parallel strands that are flanked on both sides by α-helices. This topology is a variant of the conventional dinucleotide-binding Rossmann fold (βαβ) that is widely conserved among various members of the NAD-dependent dehydrogenase family (Rao and Rossmann, 1973). The smaller substrate-binding domain (residues 1–108 and 304–331)

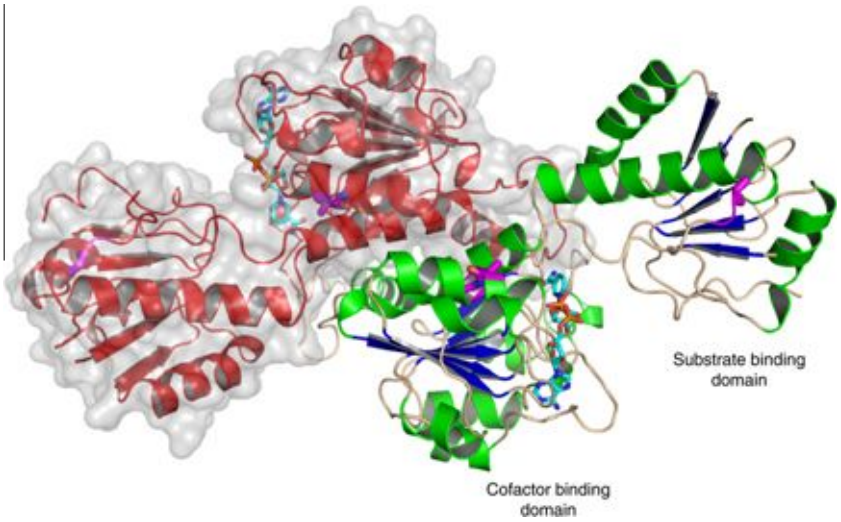

Fig.1. X-ray crystal structure of D2-HDH. Combined cartoon/surface representation of the D2-HDH dimer. One protomer is colored according to secondary structure content (helices, sheets and loops are colored green, blue and cream respectively) and the second protomer is colored red with its molecular surface rendered in gray. The NAD^+ cofactor that binds in a deep cleft at the base of the interdomain region, is shown in stick representation with C, N, O and P atoms colored cyan, blue, red and orange respectively. Residues that differ between the *L. bulgaricus* D2-HDH (strain ATCC 11842) and HO-HxoDH (strain NCIB1178) are highlighted in magenta. The figure was generated using PyMOL (http://www.pymol.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is also characterized by a modified Rossmann fold topology composed of a five-stranded parallel β-sheet flanked by five helices. The substrate-binding domain has higher average temperature factors than those from the cofactor-binding domain indicating a greater inherent flexibility of this domain (Suppl. Fig. 4). A similar observation in D-LDHb is attributed to a role in domain closure during catalysis, proposed as an essential feature of the substrate-binding domain (Razeto et al., 2002).

Tertiary structure similarity searches of the PDB database using PDBeFold (Krissinel and Henrick, 2007) revealed the closest structural homolog of D2-HDH to be the D-2-hydroxyisocaproate dehydrogenase (D-HicDH) from *Lactobacillus casei* (Dengler et al., 1997) (Suppl. Table S1). The two structures share an r.m.s.d. of 1.4 $Å^2$ over 309 equivalent Cα atoms and share 34% sequence identity. Both the core fold and the dimeric assembly are structurally conserved (Fig. 2). Structurally the largest differences between the two

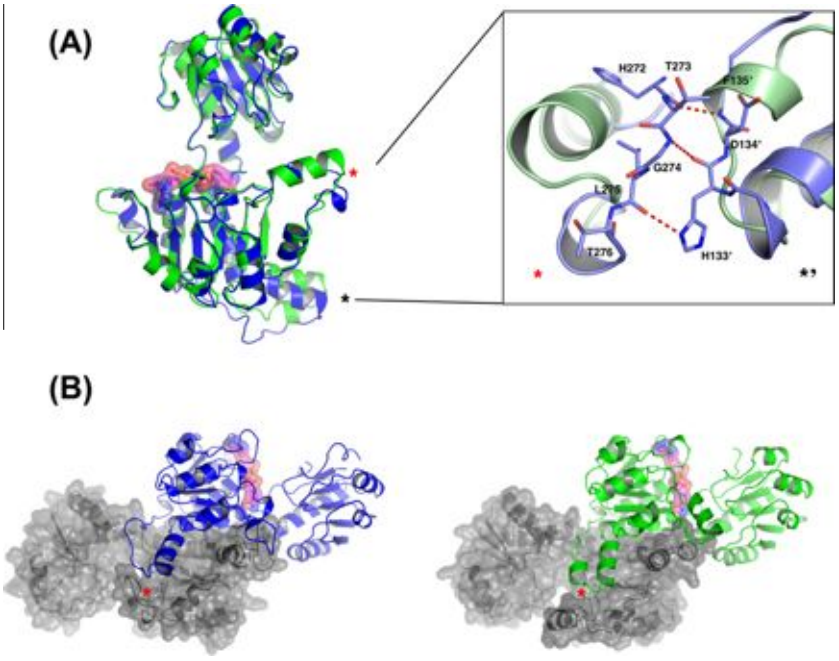

Fig.2. Structural comparison between D2-HDH and D2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*. (A) Structural superposition of the D2-HDH protein (blue) and D2-hydroxy-isocaproate dehydrogenase (D-HicDH) from *Lactobacillus casei* (green). The surface of the bound NAD^+ cofactor is shown in red. The two regions that display the most significant structural divergence are highlighted; ⁎residues H272-T276 and ⁎ residues H133-F135. In the right hand panel, the molecular interactions with the dimeric protomer at the ⁎ site are shown in detail. D2-HDH residues are shown in ball and stick representation with carbon, nitrogen and oxygen atoms colored light blue, royal blue and red, respectively. Hydrogen bonds to the symmetry copy (residues labeled as X') are shown as red dotted lines. Superimposed in green is the *L. casei* structure. Due to the internal symmetry of the D2-HDH dimer the interactions at the ⁎ site are identical. (B) Dimeric complex formation of D2-HDH (blue) and D-HicDH (green). In each case the dimeric partner protein secondary structure and surface are shown in gray. The complexes are viewed from identical orientations. The location of the ⁎ site, described in (A), is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proteins are found at residues 267–280 and 132–138 (Fig. 2A). The respective sequences are poorly conserved in both of these regions (Suppl. Fig. 1). As a result of the 2-fold internal symmetry within the D2-HDH homodimer, each of these two regions in fact interacts with the other region from the dimeric partner via a series of hydrogen-bonds (Fig. 2A). Although both of the regions are located at the dimeric interface, the dimeric assembly in both D2-HDH and D-HicDH remains otherwise conserved (Fig. 2B). The close structural homology shared between D2-HDH and D-HicDH also reveals that the D2-HDH substrate-binding domain is in the "open" conformation (Razeto et al., 2002). This conformation is catalytically inactive and consistent with the fact that no substrate is bound at the active site (Razeto et al., 2002). Sequence conservation between D2-HDH and D-HicDH is particularly high for the remainder of the dimer interface (Suppl. Fig. 5). Many residues located at the core of the protein fold are also highly conserved. The potential role of specific active site residue differences between the two proteins is discussed later.

5. Binding of the NAD$^+$ cofactor to D2-HDH

We have also determined the X-ray structure of D2-HDH in complex with the cofactor NAD$^+$ (Fig. 3). Cofactor binding does not trigger any major conformational changes to the structure of the enzyme; the apo-structure protomers share r.m.s.d's with those of the NAD$^+$-bound structural equivalents in the range 0.2 Å–0.3 Å. The NAD$^+$ molecule binds in an extended conformation where the equiplanar adenine ribose and nicotinamide rings are approximately perpendicular to the adenine ring. The NAD$^+$ conformation is stabilized along its entire length via an extensive network of hydrogen-bonding, hydrophobic and hydrophilic interactions with the co-factor binding domain. The NAD$^+$ adenine group binds in a predominantly hydrophobic pocket that is lined by the side chains of residues Leu206, Thr208, Phe211, Met217, Val179, Ile154, Tyr177, Pro209 and Leu240. The NAD$^+$ adenine ring additionally forms a polar interaction with the Thr214 side-chain. The ribose ring (adjacent to the adenine) forms two hydrogen bonds with the carboxylate side chain of Asp178, a residue that is highly conserved between different dehydrogenases. In addition to directly binding and stabilizing NAD$^+$, it is important in the discrimination between NAD$^+$ and the bulkier NADP$^+$ (Bernard et al., 1995). The NAD$^+$ pyrophosphate atoms form hydrogen bonds with the main-chain nitrogen atom from Ile159 as well as with the main-chain nitrogen and side-chain imidazole ring of His158. Collectively residues in this region of the active site are part of the NAD$^+$-binding fingerprint motif GxGxxGx(17)D, which extends between Gly155 and Asp178 in D2-HDH (Wierenga et al., 1986). The nicotinamide ribose ring does not form any interactions with the protein. The NAD$^+$ dihydronicotinamide group is flanked by a cluster of hydrophobic residues (Ile107, Ile159, Ala236 and Ala300). The ability of NAD-dependent dehydrogenases to bind NADH more tightly than NAD$^+$ has been attributed to such hydrophobic clusters (Razeto et al., 2002). Lastly, the carboxamide oxygen is held in the cis-orientation with respect to the C4 N of NAD$^+$ through hydrogen-bonding interactions with Asp261 OD2, Cys235 O and Phe301 N atoms. The combination of hydrophobic and polar interactions with the nicotinamide group ensures the activation of C4 N atom for the hydride transfer required for the dehydrogenase reaction (Li and Goldstein, 1992). D2-HDH residues involved in the binding of NAD$^+$ are either absolutely, or conservatively conserved in related proteins (Suppl. Fig. 1) and is in keeping with the general observation that NAD$^+$ binding sites are highly conserved between different HDH proteins (Lesk, 1995).

6. The D2-HDH active site: conserved and non-conserved features

Our attempts to obtain D2-HDH crystals in complex with both co-factor and a substrate were unsuccessful. To identify the features that regulate D2-HDH substrate specificity and to gain an insight into potential substrate preferences we therefore compared the D2-HDH structure with related proteins for which both struc-

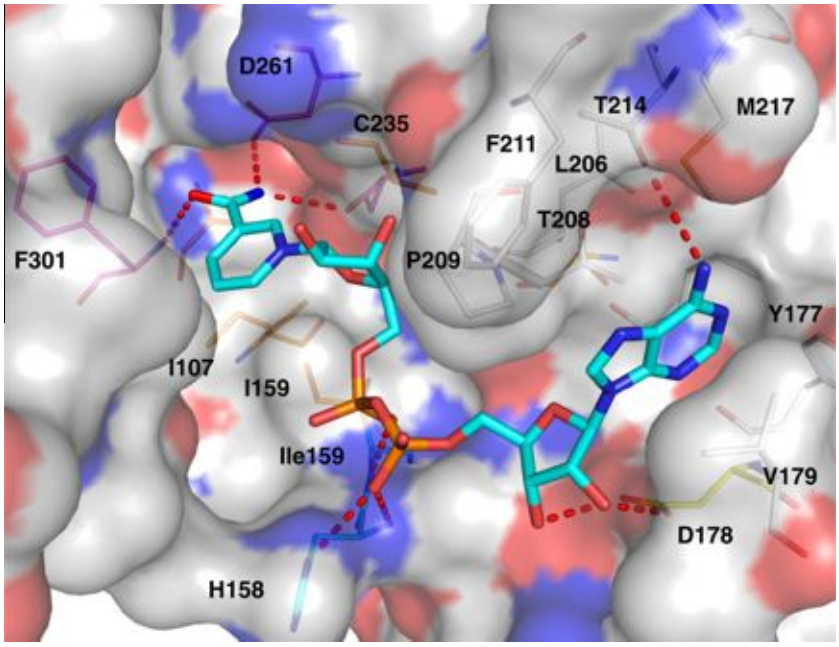

Fig.3. Binding mode of the cofactor NAD$^+$ to D2-HDH. Interactions formed between the NAD$^+$ cofactor and D2-HDH. D2-HDH residues that interact with the cofactor are shown in stick representation. Protein carbon atoms are colored according to the region of the cofactor with which they interact (white – adenine, yellow – adenine-ribose, cyan – pyrophosphate, orange – nicotinamide or magenta – carboxamide). The NAD$^+$ carbon atoms are colored cyan whilst N, O and P atoms are colored blue, red and orange, respectively. Hydrogen-bonding interactions are depicted in red. The transparent surface of the D2-HDH protein is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

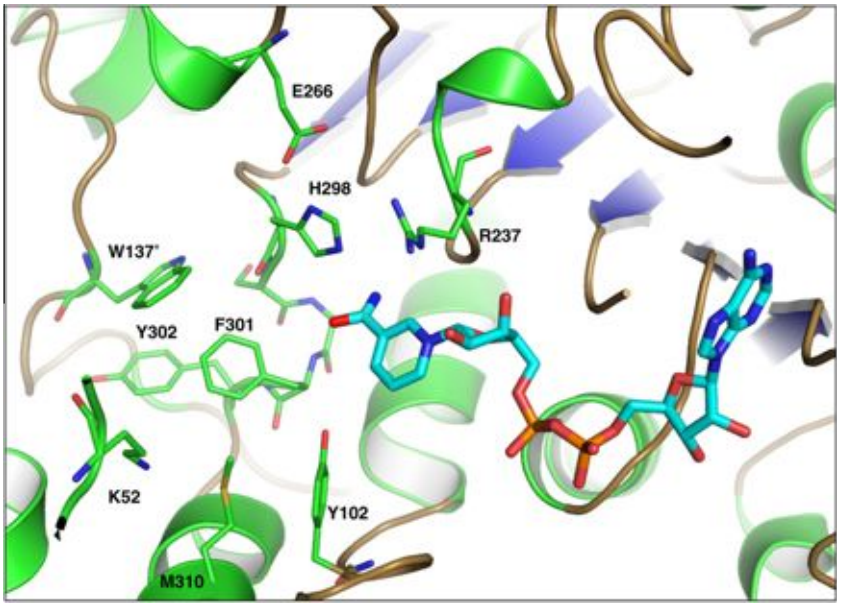

Fig.4. D2-HDH substrate-binding pocket. Residues lining the D2-HDH substrate-binding pocket are shown in stick representation with carbon, nitrogen, oxygen and sulfur atoms colored green, blue, red and yellow, respectively. The protein secondary structural elements and the NAD^+ cofactor are colored as in Fig. 1. For clarity some adjacent secondary structural elements are not depicted. Residues originating from the dimeric copy are labeled e.g. W137'. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tural and biochemical data are available. This analysis confirmed that the signature triad of active site residues, His298, Glu266 and Arg237, is conserved within the D2-HDH active site (Fig. 4) and is consistent with our observation that D2-HDH is enzymatically active (Suppl. Fig. 2). By analogy to other members of the D2-hydroxyacid dehydrogenase family, His298 is expected to function as the internal acid–base catalyst in the execution of the dehydrogenase mechanism, Glu266 to stabilize the protonated form of the catalytic His298 and lastly Arg237 to stabilize the bound substrate (Dengler et al., 1997; Kochhar et al., 1992a; Razeto et al., 2002; Taguchi and Ohta, 1993, 1994).

Residues lining the D2-HDH substrate-binding site differ from related proteins at several key positions that are known to influence substrate specificity (Suppl. Fig. 1). Two of these residues, Lys52 and Phe301 in D2-HDH, are located on opposing sides of the pocket and together dictate both the size and the electrostatic character of the substrate-binding site (Figs. 3 and 4). Considering D2-HDH Lys52, a tyrosine and a leucine are found in the equivalent positions of the *L. delbruekii* D-lactate dehydrogenase (D-LDHb) and *L. casei* D-HicDH proteins, respectively (Suppl. Fig. 1). The size of the residue at this position plays a steric role in substrate selectivity: the relatively large size of D-LDHb tyrosine prevents substrates larger than the native pyruvate from binding whilst the relatively small leucine contributes towards the broad substrate specificity of D-HicDH (Dengler et al., 1997; Razeto et al., 2002). Furthermore, in PGDH, an arginine at this position stabilizes the negative charge from the phosphoglycerate substrate (Dey et al., 2007). Despite its larger size, D2-HDH-Lys52 is not appropriately positioned within the pocket to act as a size filter analogous to the role of Tyr52 in D-LDHb.

The second site of sequence divergence, Phe301 in D2-HDH, is highly isosteric with the tyrosine found at the same position in D-HicDH, and is therefore expected to impose similar constraints on the substrate specificity, i.e. a preference for aliphatic substrates. In contrast, polar residues at this position are found in enzymes with more hydrophilic substrates, for example PGDH and HGDH (Dey et al., 2007; Martins et al., 2005). Additionally, residues located at the same position as Trp137' (originating from the dimeric protein partner) have also been observed to influence enzyme specificity. A smaller alanine residue at this position in D-HicDH correlates with broad substrate specificity, whilst in PGDH (substrate: phosphoglycerate) a positively charged lysine is found at this position. In LDH, substrate-induced domain closure allows the equivalent tryptophan to shield the substrate from the solvent (Razeto et al., 2002). This alternative conformation is stabilized by a hydrogen bond with Pro295 (Pro297 in D2-HDH).

7. Conclusion

In summary, this study is the first to describe the crystal structures of *L. bulgaricus* D-2-hydroxyacid dehydrogenase in both its apo and coenzyme NAD^+ complexed form. D2-HDH has a conserved NAD^+ binding site into which the cofactor binds. In contrast the substrate-binding site contains structural features that may influence the physiological function and substrate specificity of D2-HDH. Results of the study add to existing knowledge in the largely unexplored structural proteome of the economically important bacterium *L. bulgaricus*.

Author contributions

S.H. designed and carried out most of the experiments. S.H. and M.A. jointly carried out crystallographic model refinement, structural analysis and manuscript preparation. A.G. carried out the kinetic studies. M.W. coordinated the project and supported the work by S.H., M.A. and A.G.

Acknowledgment

We thank Dr. Maarten van de Guchte for the *Lactobacillus delbrueckii* ssp. *bulgaricus* (strain ATCC 11842) chromosomal DNA.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2012.10.009.

References

Adams, P.D. et al., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221.

Bernard, N. et al., 1994. NAD(+)-dependent D-2-hydroxyisocaproate dehydrogenase of Lactobacillus delbrueckii subsp. bulgaricus. Gene cloning and enzyme characterization. Eur. J. Biochem. 224, 439–446.

Bernard, N. et al., 1995. D175 discriminates between NADH and NADPH in the coenzyme binding site of Lactobacillus delbrueckii subsp. bulgaricus D-lactate dehydrogenase. Biochem. Biophys. Res. Commun. 208, 895–900.

CCP4. 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763.

Dengler, U. et al., 1997. Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, NAD+ and 2-oxoisocaproate at 1.9 A resolution. J. Mol. Biol. 267, 640–660.

Dey, S. et al., 2007. The effect of hinge mutations on effector binding and domain rotation in Escherichia coli D-3-phosphoglycerate dehydrogenase. J. Biol. Chem. 282, 18418–18426.

Emsley, P. et al., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501.

Hummel, W., 1999. Large-scale applications of NAD(P)-dependent oxidoreductases: recent developments. Trends Biotechnol. 17, 487–492.

Kochhar, S., Chuard, N., Hottinger, H., 1992a. Glutamate 264 modulates the pH dependence of the NAD(+)-dependent D-lactate dehydrogenase. J. Biol. Chem. 267, 20298–20301.

Kochhar, S. et al., 1992b. Evolutionary relationship of NAD(+)-dependent D-lactate dehydrogenase: comparison of primary structure of 2-hydroxy acid dehydrogenases. Biochem. Biophys. Res. Commun. 184, 60–66.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Lesk, A.M., 1995. NAD-binding domains of dehydrogenases. Curr. Opin. Struct. Biol. 5, 775–783.

Li, H., Goldstein, B.M., 1992. Carboxamide group conformation in the nicotinamide and thiazole-4-carboxamide rings: implications for enzyme binding. J. Med. Chem. 35, 3560–3567.

Martins, B.M. et al., 2005. Structural basis for stereo-specific catalysis in NAD(+)-dependent (R)-2-hydroxyglutarate dehydrogenase from Acidaminococcus fermentans. FEBS J. 272, 269–281.

McCoy, A.J., 2007. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. 63, 32–41.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255.

Rao, S.T., Rossmann, M.G., 1973. Comparison of super-secondary structures in proteins. J. Mol. Biol. 76, 241–256.

Razeto, A. et al., 2002. Domain closure, substrate specificity and catalysis of D-lactate dehydrogenase from Lactobacillus bulgaricus. J. Mol. Biol. 318, 109–119.

Taguchi, H., Ohta, T., 1993. Histidine 296 is essential for the catalysis in Lactobacillus plantarum D-lactate dehydrogenase. J. Biol. Chem. 268, 18030–18034.

Taguchi, H., Ohta, T., 1994. Essential role of arginine 235 in the substrate-binding of Lactobacillus plantarum D-lactate dehydrogenase. J. Biochem. 115, 930–936.

van de Guchte, M. et al., 2006. The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. Proc. Natl. Acad. Sci. USA 103, 9274–9279.

Wierenga, R.K., Terpstra, P., Hol, W.G., 1986. Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. J. Mol. Biol. 187, 101–107.

BMC Biotechnology

RESEARCH ARTICLE Open Access

Improved mycobacterial protein production using a *Mycobacterium smegmatis groEL1ΔC* expression strain

Elke E Noens[1*], Chris Williams[1], Madhankumar Anandhakrishnan[1], Christian Poulsen[2], Matthias T Ehebauer[1] and Matthias Wilmanns[1]

Abstract

Background: The non-pathogenic bacterium *Mycobacterium smegmatis* is widely used as a near-native expression host for the purification of *Mycobacterium tuberculosis* proteins. Unfortunately, the Hsp60 chaperone GroEL1, which is relatively highly expressed, is often co-purified with polyhistidine-tagged recombinant proteins as a major contaminant when using this expression system. This is likely due to a histidine-rich C-terminus in GroEL1.

Results: In order to improve purification efficiency and yield of polyhistidine-tagged mycobacterial target proteins, we created a mutant version of GroEL1 by removing the coding sequence for the histidine-rich C-terminus, termed GroEL1ΔC. GroEL1ΔC, which is a functional protein, is no longer able to bind nickel affinity beads. Using a selection of challenging test proteins, we show that GroEL1ΔC is no longer present in protein samples purified from the *groEL1ΔC* expression strain and demonstrate the feasibility and advantages of purifying and characterising proteins produced using this strain.

Conclusions: This novel *Mycobacterium smegmatis* expression strain allows efficient expression and purification of mycobacterial proteins while concomitantly removing the troublesome contaminant GroEL1 and consequently increasing the speed and efficiency of protein purification.

Background

Heterologous expression of recombinant proteins in *Escherichia coli* can result in the production of insoluble inclusion bodies. Recent statistics show that less than half of the *M. tuberculosis* (Mtb) proteins expressed in *E. coli* are soluble [1]. Therefore, the non-pathogenic bacterium *Mycobacterium smegmatis* is often used as an alternative, more closely related host for the expression of mycobacterial proteins. Furthermore, *M. smegmatis* may also provide mycobacterium-specific chaperones, which can help correct folding of Mtb proteins [1].

During nickel affinity purification, it has been observed that a protein of 56 kDa is co-purified with polyhistidine-tagged recombinant proteins while using *M. smegmatis* as an expression system. This contaminant was previously identified as the Hsp60 chaperone GroEL1 of *M. smegmatis* [1-3]. The protein sequence of GroEL1 shows a histidine-rich C-terminus (7 out of 11 amino acids are histidines), which is likely to be the reason for the observed nickel sepharose binding [1,2].

Unlike most other bacteria, mycobacteria possess two Hsp60 chaperone *groEL* genes, one of which is arranged in the bicistronic *groESL* operon [4]. *M. smegmatis* also encodes a third Hsp60 protein (Msmeg1978), which is more distantly related to GroEL1 (Msmeg1583) and GroEL2 (Msmeg0880) [3]. Although *groEL1* of *M. smegmatis* can be found in the same operon as *groES*, an arrangement indispensable for the chaperone function in bacteria, its histidine-rich tail is distinct from the more typical glycine-methionine-rich C-terminal region found in GroEL2 [3]. Furthermore, *groEL2* is an essential gene and exists in all actinobacteria, in contrast to *groEL1* [3,5]. Recently, it has been shown that *groEL2* and *groES* are expressed more strongly than *groEL1*, which might have arisen from a difference in stability of the predicted post-transcriptionally cleaved mRNAs for

* Correspondence: e.noens@embl-hamburg.de
[1]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Building 25a, Notkestraße 85, 22603 Hamburg, Germany
Full list of author information is available at the end of the article

groES and *groEL1* [5]. Consistent with the current chaperone model in mycobacteria, one chaperone, here GroEL2, would act as the main house keeping chaperone in *M. smegmatis*, with the other chaperones (GroEL1 and Msmeg1978) adopting more specialised functions. Indeed, GroEL1 of *M. tuberculosis* was recently identified as being associated with nucleotides, suggesting a role as a DNA chaperone, while GroEL1 of *M. smegmatis* was found to have a role in mycolic acid biosynthesis during biofilm formation [5,6,3].

The co-purification of GroEL1 with histidine-tagged recombinant proteins can be particularly problematic since native GroEL1 is expressed at relatively high levels, meaning that in the case of a low yield of recombinant protein, GroEL1 may well compete with the protein of interest for binding sites on nickel affinity beads. Minimal sample manipulation is recommended during protein purification to improve efficiency. Therefore, additional steps required to remove GroEL1 can result in a significant loss of the protein of interest.

In this article, we describe an *M. smegmatis* expression strain containing a mutant version of GroEL1, termed GroEL1ΔC, which consists of a *groEL1* gene without a coding sequence for the histidine-rich C-terminal tail. We show that GroEL1ΔC is a functional protein, which no longer co-purifies when using nickel affinity purification and we provide evidence that proteins purified from this strain are correctly folded, active and that they behave identically to those purified from the original expression strain. Taken together, our data demonstrate that *M. smegmatis groEL1ΔC* is a competent protein expression strain, which allows the efficient removal of the troublesome contaminant GroEL1 without the requirement of additional purification steps.

Methods

Bacterial strains and media

The *E. coli* strains DH5α (Invitrogen) and HB101 (Promega) were used for cloning of expression constructs and the target substrate to generate the mutant version of *groEL1* using standard procedures [7]. Transformants were selected in Luria Broth containing the appropriate antibiotics.

M. smegmatis mc^2155 was used as the parent (wild type) strain for the *groEL1ΔC* strain. Both *M. smegmatis* strains were maintained in Middlebrook 7H9 or 7H10 medium supplemented with 0.2% (v/v) glycerol, 10% ADC, 0.05% (v/v) tween-80 and the appropriate antibiotics.

For biofilm formation, 10 ml of biofilm media was inoculated with 10 μl of saturated culture and incubated at 30°C without disturbance [3,8].

For the expression of the recombination proteins in *M. smegmatis* in order to create the mutant form of *groEL1*, 0.2% succinate (w/v) was added as a carbon source to 7H9 medium supplemented with 0.2% (v/v) glycerol, 0.05% (v/v) tween and the appropriate antibiotics. Expression of his-tagged recombinant proteins in *M. smegmatis* was performed in 7H9 medium supplemented with 0.2% (w/v) glucose as carbon source. Acetamide was added to a final concentration of 0.2% (w/v) at 0.5 OD_{600} and at 2.5 OD_{600} for the expression of the recombination proteins and his-tagged recombinant proteins, respectively.

Plasmids, constructs and oligonucleotides

All plasmids and constructs are summarised in Table 1 and oligonucleotides are listed in Table 2. pJV53 was used to express the recombination proteins [9]. pYUB854 was used for the preparation of the target substrate to create the *groEL1ΔC* strain [10]. pGH542, harbouring a δγ resolvase, was used to generate an unmarked deletion [11]. Using the primer pairs Msmeg1583-F1.2 & Msmeg1583-R1 and Msmeg1583-F2 & Msmeg1583-R2.1, two 500 bp fragments, homologous to the fragments +1067/+1587 and +1621/+2176 relative to the translational start of Msmeg1583, were amplified

Table 1 Plasmids and constructs used in this study

Plasmid/ construct	Description	Reference
pJV53	Che9c recombination proteins under control of the acetamidase promoter in pLAM12	[9]
pYUB854	Hyg^R cassette flanked by γδ-res sites and 2 MCSs	[10]
pGH542	Expressing an γδ resolvase and tetracycline resistant	[11]
pEN15	pYUB854 with a 520 bp fragment harbouring *groEL1* (+1067/+1587, relative to *groEL1*) inserted upstream of the Hyg^R cassette and a 555 bp fragment downstream of *groEL1* including the STOP codon of *groEL1*, inserted downstream of the Hyg^R cassette	This paper
pMyNT	Mycobacterial overexpression vector	Geerlof *et al.*, unpublished data
pMyNT/ PrcA-B	Rv2109-2110 in pMYNT, Rv2110 is N-terminally his-tagged	[12]
pMyNT/ AccD5E5	Rv3280-3281 in pMYNT. Only his-tagged Rv3280 seems to express using this construct.	This paper
pMyNT/ AccA3	Rv3285 in pMyNT	This paper
pMyNT/ CFP10-ESAT6	Rv3874-3875 in pMYNT, Rv3874 is N-terminally his-tagged	[12]
pMyNT/ ACPS	Rv2523 in pMYNT	This paper

Table 2 Primers used in this study

Primer	Sequence (5'-3')	Location 5'	Relative to
Msmeg1583-F1.2	GCGC**CTTAAG**CGACTGGGATCGCGAGAAGCTGC	+1067	Msmeg1583
Msmeg1583-R1	GCGC**TCTAGA**CTCGTCCTCGTCGGCCGGCTTG	+1587	Msmeg1583
Msmeg1583-F2	GCGC**AAGCTT**GATCCATTTCACGCGACACCCCC	+1620	Msmeg1583
Msmeg1583-R2.1	GCGC**ACTAGT**GGTGTTCGATCGTCTGGCCGATG	+2176	Msmeg1583
accD5E5-F	GATC**TCATGA**GTATGACAAGCGTTACC G	+1	Rv3280
accD5E5-R	GTCA**AAGCTT**TTATCGGCGCATGTGCG	+2161	Rv3280
accA3-F	GATC**CCATGG**GTATGGCTAGTCACGCC	+2	Rv3285
accA3-R	GTCA**AAGCTT**TTACTTGATCTCGGCGAGC	+1803	Rv3285
Rv2523-F	CATG**CCATGG**GCATCGTCGGTGTGGGG	+1	Rv2523
Rv2523-R	CCC**AAGCTT**ACGGGGCCTCCAGGATGGC	+391	Rv2523

Restriction sites are presented in bold face. CTTAAG = *Eco*RI, TCTAGA = *Xba*I, CCATGG = NcoI, TCATGA = *Bsp*HI.
AAGCTT = *Hin*dIII, ACTAGT = *Spe*I.

and subsequently ligated *Afl*II-*Xba*I (F1.2-R1) and *Hin*-*d*III-*Spe*I (F2-R2.1) into pYUB854, creating pEN15.

For the expression of *M. tuberculosis* proteins in *M. smegmatis*, the pMyNT expression vector was used [Geerlof *et al.*, unpublished data]. pMyNT/ACPS, pMyNT/AccA3 and pMyNT/AccD5 were made as follows: PCR was performed with primer pair Rv2523-F & Rv2523-R for ACPS, accA3-F & accA3-R for AccA3 and accD5E5-F & accD5E5-R for AccD5 and the resulting fragments were digested with *Nco*I-*Hin*dIII and inserted into *Nco*I-*Hin*dIII digested pMyNT.

Creation of the *groEL1ΔC* mutant

The *groEL1ΔC* mutant was created using the mycobacterial recombineering method [9]. pEN15 was digested with *Afl*II and *Spe*I to create the linear target substrate, which was introduced into mc^2155 electrocompetent cells, expressing the recombinase genes on pJV53 and in this way creating hygromycin-resistant transformants. The hygromycin-resistance cassette was removed using δγ resolvase, expressed on pGH542, generating an unmarked deletion [11].

Southern blot analysis

Genomic DNA (5ug) was isolated as described [9], digested with the appropriate enzymes, separated on a 0.9% agarose gel and transferred to a positively charged nylon membrane (Roche). For DNA probe labelling, hybridisation and detection, the DIG high prime DNA labelling and detection starter kit 1 (Roche) was used.

Growth curves

Bacterial growth was followed by measuring the optical densities at a wavelength of 600 nm as a function of time. Cultures were prepared with 7H9 expression medium (0.2% (w/v) glucose as carbon source) in identical triplicates for each strain. Duplicate samples were taken every 4 hours for 40 hours. When the optical density at 600 nm exceeded 1.5, samples were diluted in order to remain within the linear range of the detector.

Protein expression and purification

All methods related to protein expression in *M. smegmatis* were carried out as described [12,13]. Protein-protein complexes from operon-encoded proteins were expressed using the native operon structure [9]. In brief, pellets from 500 ml cultures were dissolved in 30 ml lysis buffer containing 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.5 M urea with protease inhibitor cocktail (Sigma) and 1 mg/ml DNase I (Serva). Resuspended cells were sonicated four times, each for 5 min (with a 0.3 s pulse and 0.7 s rest) at 5 min intervals to prevent overheating, using a Bandelin VW3200 probe at 45% amplitude. The supernatant was collected after centrifugation (30,000 × g) for 1 h at 4°C, filtered through a 0.44 μm filter and loaded onto a nickel affinity sepharose (NiAC) column. After washing with 10 column volumes of 50 mM Tris-HCl pH 8.0, 300 mM NaCl and 20 mM imidazole, proteins were eluted in 50 mM Tris-HCl, 100-150 mM NaCl and 250-500 mM imidazole and subjected to size exclusion chromatography using either a Superdex 75 (16/60) column (GE Healthcare) or, for large protein complexes, a Superose 6 (10/300) (GE Healthcare) with 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM DTT as buffer. The collected protein samples were analysed by SDS-PAGE and concentrated accordingly.

Circular Dichroism (CD) spectrum analysis

CD measurements were performed on a Jasco J-810 spectropolarimeter. Prior to measurement, samples were dialysed into 10 mM potassium phosphate, 150 mM NaCl, pH 7.4. Spectra were recorded between 182 and 260 nm in a 2 mm cuvette with machine settings as

follows: 1 nm bandwidth, 1 sec response, 1 nm data pitch, 100 nm/min scan speed, cell length of 0.1 cm. Each curve presented is the average of three separate measurements.

Coupled enzyme assay

Enzymatic activity of the AccD5-AccA3 complex was estimated by a coupled enzyme assay that follows the rate of ATP hydrolysis spectrophotometrically [14]. The production of ADP during the reaction was coupled to pyruvate kinase and lactate dehydrogenase, and the oxidation of NADH was probed at 340 nm. The assay mixture contained 7 units of pyruvate kinase, 10 units of lactate dehydrogenase, 50 mM $NaHCO_3$, 3 mM ATP, 0.5 mM phosphoenol pyruvate, 0.2 mM NADH, 0.3 mg/ml BSA, 100 mM K_2HPO_4 pH 7.6 and 5 mM $MgCl_2$ and varying concentrations of propionyl-coenzyme A. Reactions were initiated by the addition of enzyme to the assay mixture and were maintained at 30°C. Data were acquired using a Tecan infinite M1000 microplate reader. The kinetic parameters K_m and V_{max} were determined by fitting the mean velocities *versus* the substrate concentration to the Michaelis-Menten equation of enzyme kinetics using nonlinear regression analysis, executed by the program Prism 5 (GraphPad Software™).

Results and Discussion

Creation of the *groEL1ΔC* strain

Currently, the role of GroEL1 in protein folding is uncertain. A closer look at the structure of *E. coli* GroEL [15] indicates that, although the C-terminal region of the protein is not easily accessible, pointing towards the central cavity of the wheel-like structure adopted by oligomeric GroEL, the extreme C-terminal 20 amino acids are absent from the model. Similarly, the GroEL structure of *Paracoccus denitrificans* also lacks these residues [16]. These observations suggest that the C-terminal region of GroEL is highly flexible and could reach out of the central cavity, allowing in this way *M. smegmatis* GroEL1 to bind nickel affinity beads. Additionally, as native GroEL1 from *M. tuberculosis* is oligomeric [17], nickel binding would require only one accessible histidine-rich region. Therefore, we decided to change only the last eleven amino acids of the protein, rather than to make a full knock out strain, in order to minimise changes to the expression strain. A precise chromosomal deletion of fragment 1588-1620, relative to the translational start of *groEL1* (Msmeg1583), was created using the mycobacterial recombineering technique [9] (Figure 1A). Southern hybridisation (Figure 1B) was used to verify that correct homologous recombination had taken place in the hygromycin-resistant and the unmarked deletion strain (Figure 1B). The latter strain, in which the hygromycin resistance cassette has been removed, has the C-terminal eleven residues of GroEL1, containing seven histidines, replaced by six non-histidine residues, which are part of the "scar" sequence left behind after removal of the resistance cassette (Figure 1A). The stop codon of this recombinant version of GroEL1 is TAA, which although rare, is recognised in high G+C *mycobacteria* [18]. This unmarked deletion strain, referred to as *M. smegmatis groEL1ΔC*, is used in all further experiments.

Ojha *et al.* reported that the last 18 amino acids of GroEL1 are essential for the formation of mature biofilms [3]. Therefore, to test the functionality of the GroEL1ΔC protein, we compared biofilm formation in both the wild type and g*roEL1ΔC* strains. Both strains were able to form mature biofilms after an incubation time of 7 days at 30°C, indicating that GroEL1ΔC is indeed fully functional (Figure 2A). Taking into account the data from Ojha *et al.*, our results could suggest that either the amino acids important for biofilm formation are upstream of those removed in the GroEL1ΔC protein, or that removal of the last 18 residues may affect the folding of at least a part of GroEL1. Additionally, as this newly created strain was constructed for the overexpression of mycobacterial proteins, its growth in 7H9 expression medium was compared to the original expression strain *M. smegmatis* mc^2155 (Figure 2B). We observed no significant differences in growth between the two strains, with both reaching an OD_{600} of between 2.5-3.0 after approximately 18 hours, at which time expression is usually induced.

GroEL1ΔC is absent during nickel affinity purification of proteins expressed in *M. smegmatis groEL1ΔC*

To demonstrate the absence of GroEL1ΔC as a contaminant when using the *M. smegmatis groEL1ΔC* expression strain, we determined the expression and purification efficiency of our strain in comparison to the wild type strain using five different constructs, representing a variety of different protein molecules, including the mycobacterial proteasome, the CFP10-ESAT6 complex, the AccD5-AccA3 dodecameric acyl-CoA carboxylase complex and the holo-acyl-carrier protein synthase (for details, see Table 3). Additionally, we also used the empty pMyNT vector, to check for GroEL1 binding in the absence of a his-tagged protein. All constructs were transformed into both *M. smegmatis* mc^2155 and *groEL1ΔC* and the resulting transformants were cultured in 7H9 expression medium and induced by the addition of acetamide to a final concentration of 35 mM. Eighteen hours after induction, the cells were collected by centrifugation, lysed and the soluble protein fraction was passed over a nickel affinity column, with the elution fraction being analysed by SDS-PAGE (Figure 3). While GroEL1 was visible in samples purified

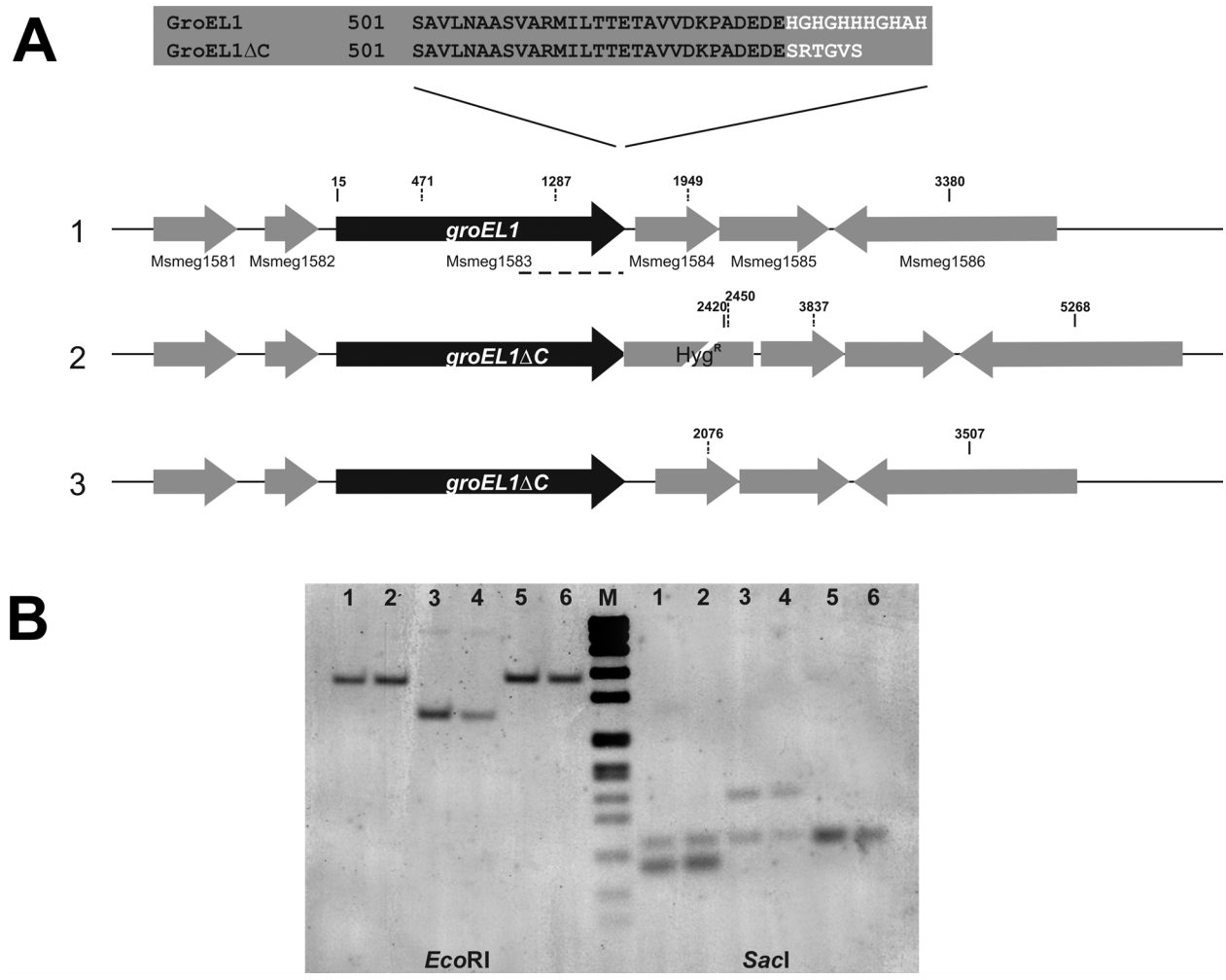

Figure 1 Construction of the *groEL1ΔC* strain. (A) A schematic representation of the genomic organisation of *groEL1* (Msmeg1583) in *M. smegmatis* mc^{2}155 (1), the hygromycin-resistant *groEL1ΔC* mutant (2) and the unmarked *groEL1ΔC* deletion strain (3). Msmeg1582 is *groES* and Msmeg1581, 1584, 1585 and 1586 encode for proteins of unknown function. Zoom sequence of the C-terminal 40 amino acids of the *groEL1* gene product, showing the histidine-rich C-terminal region. In the *groEL1ΔC* strain, the eleven C-terminal amino acids (white) were replaced by six different residues. Hyg^R = hygromycin-resistance cassette. (B) Southern blot analysis performed with genomic DNA of *M. smegmatis* mc^{2}155 (1), mc^{2}155 carrying pJV53 (2), two correct hygromycin-resistant *groEL1ΔC* mutants (3-4) and two correct unmarked *groEL1ΔC* deletions (5-6). The genomic DNA was digested with *Eco*RI or *Sac*I. The positions of the restriction sites *Eco*RI and *Sac*I, relative to the start of *groEL1*, are presented as a full and dotted vertical line in A, respectively. The PCR product of primer pair Msmeg1583-F1.2 & Msmeg1583-R1 was used as a probe, shown as a dashed line under the genes in A1. (M) DNA marker VII, DIG labelled (Roche).

from *M. smegmatis* mc^{2}155 (Figure 3, lanes a), the protein was noticeably absent in five out of six protein samples isolated from the *groEL1ΔC* strain (Figure 3, lanes b). Due to the fact that AccD5 has a similar size to GroEL1, we were unable to determine its presence or absence in samples of the purified acyl-CoA carboxylase complex by SDS-PAGE. Therefore, samples isolated from gel (Figure 3) were analyzed by mass spectrometry (Additional file 1). While numerous peptides from both GroEL1 and AccD5 could be identified from gel slices deriving from the mc^{2}155 strain, only AccD5 peptides could be detected in the sample obtained from the *groEL1ΔC* strain (Additional file 1). Likewise, MALDI-TOF mass spectrometry was performed on the other protein samples, verifying the absence of GroEL1 peptides in the protein samples derived from *M. smegmatis groEL1ΔC* (data not shown).

Proteins purified from *M. smegmatis groEL1ΔC* behave identically to those purified from the wild type strain

M. smegmatis encodes three forms of the Hsp60 chaperone GroEL: Msmeg1583 (GroEL1), Msmeg0880 (GroEL2) and Msmeg1978. However, the precise molecular function of each protein remains unclear.

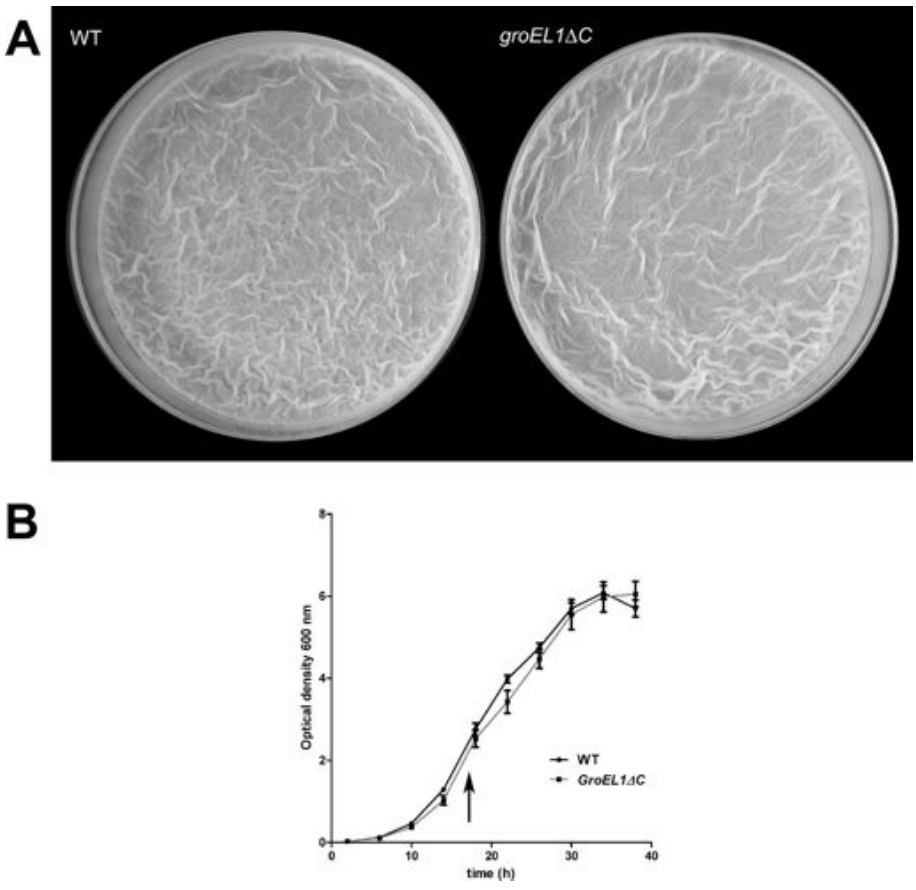

Figure 2 Biofilm formation and growth rates of *M. smegmatis* mc^2 155 and *M. smegmatis groEL1ΔC* are comparable. (A) Both *M. smegmatis* mc^2 155 (WT) and *M. smegmatis groEL1ΔC* strains are able to form biofilms after an incubation time of 7 days at 30°C. (B) Growth curve of *M. smegmatis* mc^{2}155 (WT = black) and *M. smegmatis groEL1ΔC* strains (grey) in 7H9 expression medium. The arrow represents the typical time of induction in *M. smegmatis*.

Changing the last 18 amino acids of GroEL1 does not alter growth but does result in a strong defect in biofilm formation [3]. To confirm that the newly created recombinant version of GroEL1 has no effect on the correct folding and, ultimately, the function of the proteins expressed in *M. smegmatis groEL1ΔC*, a number of different proteins and protein complexes have been expressed and analysed.

In the previous section, we have shown that it is possible to express and purify potentially challenging protein complexes, such as the proteasome complex PrcA-B and the CFP10-ESAT6 complex, from the recombinant *groEL1ΔC* strain. These data imply that the proteins isolated from the *groEL1ΔC* strain are correctly folded, since we were able to observe all components after purification. In both examples, complex formation

Table 3 List of test proteins used to validate the *groEL1ΔC* expression strain

ORF	Annotation	Description	Expressed ...	Mol. Mass (kDa)
Rv2109c Rv2110c	PrcA PrcB	α- and β-subunit of the mycobacterial proteasome ($\alpha_7\beta_7\beta_7\alpha_7$ subunit organisation)	Using native operon content, producing a 730 kDa multimeric complex	26.8 30.3
Rv3285 Rv3280	AccA3 AccD5	α- and β-subunit from acyl-CoA carboxylase AccD5-AccA3 complex ($\alpha_3\beta_3\beta_3\alpha_3$ subunit organisation)	As monomeric proteins, mixed to form a acyl-CoA carboxylase complex of 740 kDa	63.8 59.4
Rv3874 Rv3875	CFP10 ESAT6	Potential virulence factor CFP10-ESAT6 complex	Using native operon content, producing a heterodimeric (1:1) complex	10.8 9.9
Rv2523c	ACPS	Holo-acyl-carrier protein synthase	As monomeric protein	14

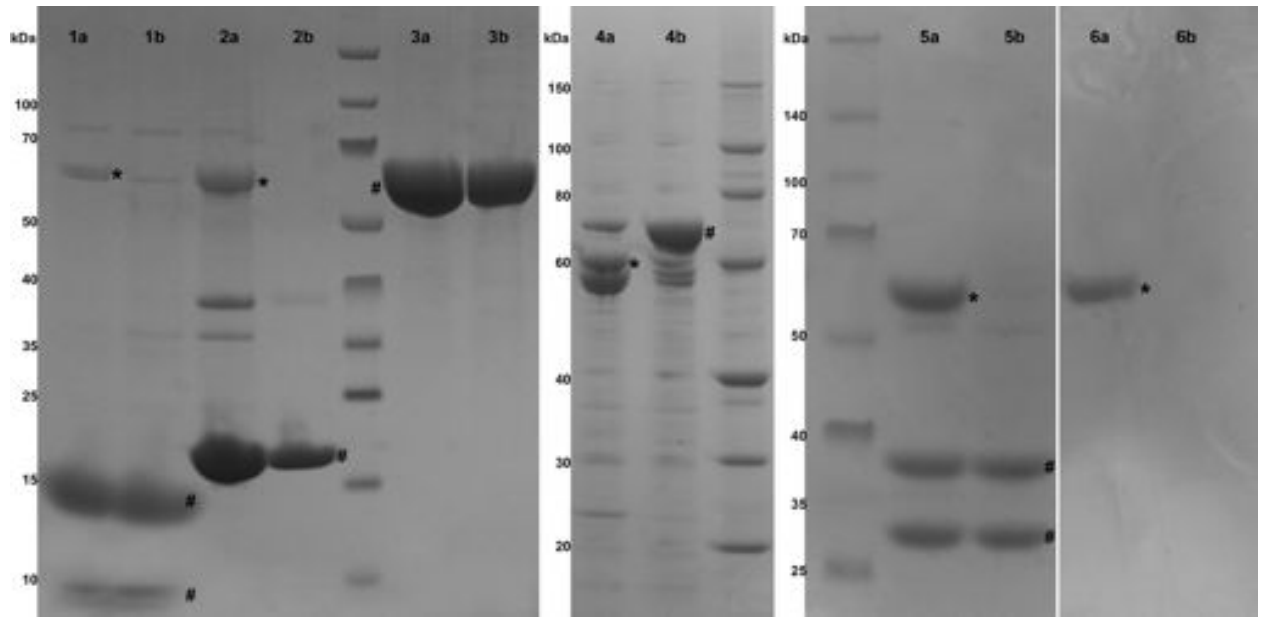

Figure 3 GroEL1ΔC is absent from protein samples purified from *M. smegmatis groEL1ΔC*. SDS-PAGE analysis of protein samples isolated from cells expressing either CFP10-ESAT6 heterodimer (1), a holo-acyl-carrier protein synthase (ACPS) (2), the β- and α-subunit D5 (3) and A3 (4) of the acyl-CoA carboxylase AccD5-AccA3 complex, the PrcA-B complex (5) or the empty expression vector (empty pMyNT) (6). **a**, proteins expressed in *M. smegmatis* mc^2155; **b**, proteins expressed in *M. smegmatis groEL1ΔC*. Purified proteins of interest are labelled with #. Visible presence of GroEL1 is depicted with *. Squares identify the bands excised for peptide mass fingerprinting. GroEL1 = 56 kDa.

requires direct protein-protein interactions between subunits of the complex as only one subunit is his-tagged.

Taking our analysis one step further, we directly tested the structural and functional properties of proteins isolated from the *groEL1ΔC* strain. We used the five expression constructs described above and transformed them into both *M. smegmatis* mc^2155 and *groEL1ΔC*. Proteins were expressed and purified using a nickel affinity column as described above. AccD5 and AccA3 protein samples were mixed in a 1:1 stoichiometry to form the high-molecular-weight AccD5-AccA3 complex. Size exclusion chromatography was performed on all samples as a final purification step.

Circular dichroism (CD) spectroscopy is a powerful tool used to visualise the secondary structure properties of protein samples. We observed that the four protein samples isolated from g*roEL1ΔC* gave virtually identical CD spectra to those purified from the wild type strain (Figure 4), implying that they are correctly folded. Furthermore, the CD spectra of the CFP10-ESAT6 complexes, showing a protein with high helical content, are comparable to those collected previously [12] and are in line with the X-ray structure, which consists of a four-helical bundle complex (PDB ID: 3FAV) [12].

Additionally, we have demonstrated carboxylase activity of the acyl-CoA carboxylase AccD5-AccA3 complex, isolated from *groEL1ΔC*, using an enzyme-coupled reaction (Figure 5). Using propionyl-CoA as a substrate, AccD5-AccA3 showed carboxylase activity with a Km = 0.1301 ± 0.0198 mM and a Vmax = 1.333 ± 0.049 mM min^{-1} mg^{-1}, data which are similar to the parameters determined using the AccD5-AccA3 complex isolated from *E. coli* [19], indicating that the AccD5-AccA3 complex isolated from *groEL1ΔC* is a functional carboxylase. Carboxylase activity requires the α-subunit of the carboxylase to be post-translationally biotinylated [19], implying that the subunits of this large megasynthase are folded correctly and, in the case of the α-subunit, correctly post-translationally modified, when isolated from *groEL1ΔC*.

Conclusions

We have developed an *M. smegmatis* expression strain that allows efficient expression and purification of mycobacterial proteins, multi-subunit protein complexes and post-translationally modified proteins while concomitantly removing the troublesome contaminant GroEL1 and consequently increasing the speed and efficiency of protein purification. The *M. smegmatis groEL1ΔC* strain is particularly suitable for laboratories performing *in vitro* activity assays and structural studies on mycobacterial proteins and protein complexes.

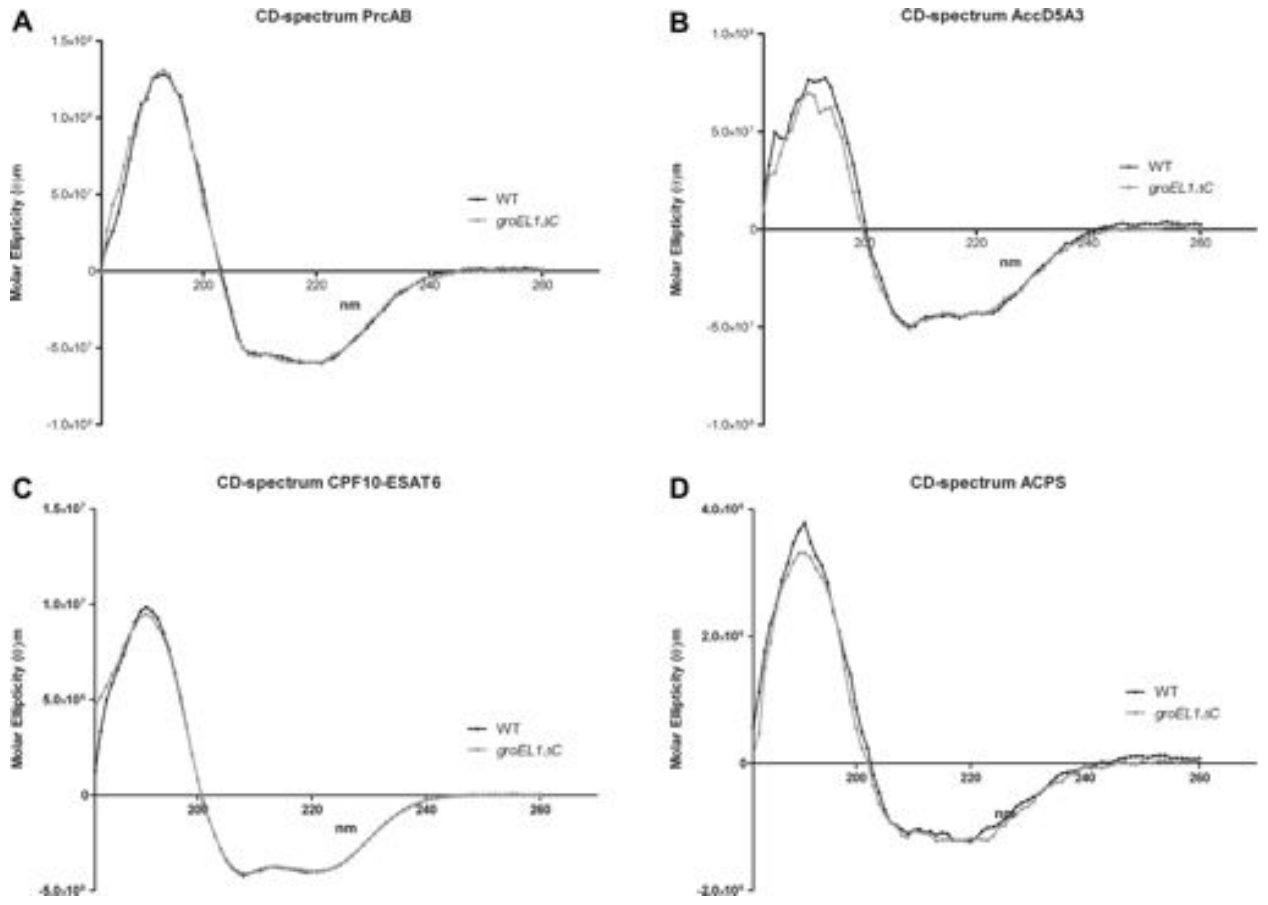

Figure 4 Proteins isolated from both strains give virtually identical CD spectra. CD spectra of the multimeric proteasome complex PrcA-B (A), the dodecameric acyl-CoA carboxylase AccD5-AccA3 complex (B), CFP10-ESAT6 heterodimer (C), and monomeric protein ACPS (D) expressed in *M. smegmatis* mc^{2}155 (WT = black) and *M. smegmatis groEL1ΔC* (grey) are virtually identical. For A and B, a concentration between 170 and 200 nM was used while for C and D, concentrations were between 5 and 10 μM.

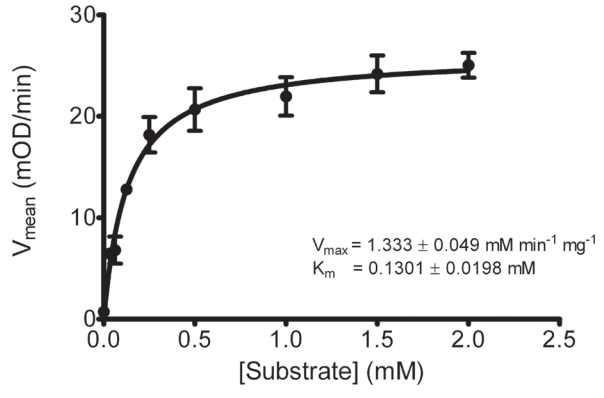

Figure 5 Kinetics of AccD5-AccA3 isolated from *M. smegmatis groEL1ΔC*. Carboxylation activity of the acyl-CoA carboxylase AccD5-AccA3 complex isolated from *groEL1ΔC* was measured using an enzyme-coupled reaction with propionyl-CoA as substrate, providing Km = 0.1469 mM and Vmax = 28.5 mOD/min.

Additional material

Additional file 1: GroEL1 is absent from an AccD5 protein sample derived from *M. smegmatis groEL1ΔC*. Results of peptide mass fingerprinting analysis of samples excised from SDS-PAGE gel (Figure 3, boxes). Shown in red are the peptides that could be identified. (a) Sample derived from *M. smegmatis* mc^{2}155. (b) Sample derived from *M. smegmatis groEL1ΔC*.

Abbreviations

PCR: Polymerase chain reaction; kDa: kilo Dalton; Hsp60: Heat shock protein 60; ADC: Albumine-dextrose-catalase; DMSO: dymethylsulfoxide; NiAc: Nickel affinity sepharose column; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electroporesis; MALDI-TOF: matrix-assisted laser desorption/ioization reflection time-of-flight.

Acknowledgements and Funding

We thank Arie Geerlof for the pMyNT expression vector, Young-Hwa Song for her contribution in the early stages of the work, the Proteomics Core Facility of EMBL Heidelberg for performing peptide mass fingerprinting, the

Mandelkow Lab (Max Planck Institute for Structural Molecular Biology, Hamburg, Germany) for access to their circular dichroism spectroscope. CW is funded by a Rubicon post-doctoral fellowship (825.08.023) from the Netherlands organization for scientific research (NWO). ME is funded by an EMBO long-term fellowship (ALTF-7272008). The project has been supported by grants awarded to MW from BMBF (Pathogenomik Plus PTJ-BIO 0313801L), from the European Commission Framework VII (NATT, 222965 and SystemTB, 241587) and from the DFG (SPP1170, WI 1058/6-3).

Author details
[1]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Building 25a, Notkestraße 85, 22603 Hamburg, Germany. [2]The Scripps Research Institute, La Jolla, 92037, USA.

Authors' contributions
EN and CP designed the study. EN made the *groEL1ΔC* strain, tested its functionality and growth, expressed and purified all proteins described except the AccD5-AccA3 complex and wrote the manuscript. CW carried out the CD measurements, provided technical assistance and participated in writing the manuscript. MA carried out all experiments concerning AccD5-AccA3. CP provided expression constructs. ME participated in testing the strain's growth and feasibility. MW organized the funding, supervised the work and helped revising the manuscript. All authors read and approved the final manuscript.

Received: 17 December 2010 Accepted: 25 March 2011
Published: 25 March 2011

References

1. Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Lott JS: **A new Gateway® vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*.** *Protein Expr Purif* 2008, **57**:81-87.
2. Poulsen C, Ahkter Y, Jeon AH, Schmitt-Ulms G, Meyer HE, Stühler K, Wilmanns M, Song YH: **Proteome-wide identification of mycobacterial pupylation targets.** *Mol Syst Biol* 2010, **6**:386-394.
3. Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR Jr, Hatfull GF: **GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in Mycobacteria.** *Cell* 2005, **123**:861-873.
4. Rinke de Wit TF, Bekelie S, Osland A, Miko TL, Hermans PW, van Soolingen D, Drijfhout JW, Schöningh R, Janson AA, Thole JE: **Mycobacteria contain two *groEL* genes: the second *Mycobacterium leprae groEL* gene is arranged in an operon with *groES*.** *Mol Microbiol* 1992, **6(14)**:1995-2007.
5. Rao T, Lund PA: **Differential expression of the multiple chaperonins of *Mycobacterium smegmatis*.** *FEMS Micobiol Lett* 2010, **310**:24-31.
6. Basu D, Khare G, Singh S, Tyagi A, Khosla S, Mande SC: **A novel nucleoid-associated protein of *Mycobacterium tuberculosis* is a sequence homolog of GroEL.** *Nucleic Acids Res* 2009, **37(15)**:4944-4954.
7. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual.** Cold Spring harbor: Cold Spring harbor laboratory press; 1989.
8. Recht J, Kolter R: **Glycopeptidolipid acetylation affects sliding motility and biofilm formation in *Mycobacterium smegmatis*.** *J Bacteriol* 2001, **183**:5718-5724.
9. van Kessel JC, Hatfull GF: **Recombineering in *Mycobacterium tuberculosis*.** *Nat Methods* 2007, **4(2)**:147-152.
10. Bardarov S, Bardarov S Jr, Pavelka Ms Jr, Sambandamurthy V, Larsen M, Tufariello J, Chan J, Hatfull G, Jacobs WR Jr: **Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis*, *M. bovis* BCG and *M. smegmatis*.** *Microbiology* 2002, **148**:3007-3017.
11. Piuri M, Hatfull GF: **A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells.** *Mol Microbiology* 2006, **62**:1569-1585.
12. Poulsen C, Holton S, Geerlof A, Wilmanns M, Song YH: **Stoichiometric protein complex formation and over-expression using the prokaryotic native operon structure.** *FEBS Lett* 2010, **584**:669-674.
13. Daugelat S, Kowall J, Matthow J, Bumann D, Winter R, Hurwitz R, Kaufmann SH: **The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization.** *Microbes Infect* 2003, **5**:1082-1095.
14. Diacovich L, Peiru S, Kurth D, Rodriguez E, Podesta F, Khosla C, Gramajo H: **Kinetic and Structural Analysis of a New Group of Acyl-CoA Carboxylases Found in *Streptomyces coelicolor* A3(2).** *J Biol Chem* 2002, **277**:31228-31236.
15. Xu Z, Horwich AL, Sigler PB: **The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex.** *Nature* 1997, **388(6644)**:741-50.
16. Fukami TA, Yohda M, Taguchi H, Yoshida M, Miki K: **Crystal structure of chaperonin-60 from Paracoccus denitrificans.** *J Mol Biol* 2001, **312(3)**:501-9.
17. Kumar CM, Khare G, Srikanth CV, Tyagi AK, Sardesai AA, Mande SC: **Facilitated oligomerization of mycobacterial GroEL: Evidence for phosphorylation-mediated oligomerization.** *J Bacteriol* 2009, **191**:6525-6538.
18. Andersson SGE, Sharp PM: **Codon usage in the *Mycobacterium tuberculosis* complex.** *Microbiology* 1996, **142**:915-925.
19. Gago G, Kurth D, Diacovich L, Tsai SC, Gramajo H: **Biochemical and structural characterization of an essential acyl Coenzyme A carboxylase from *Mycobacterium tuberculosis*.** *J Bacteriol* 2006, **188**:477-486.

doi:10.1186/1472-6750-11-27
Cite this article as: Noens *et al.*: Improved mycobacterial protein production using a *Mycobacterium smegmatis groEL1ΔC* expression strain. *BMC Biotechnology* 2011 **11**:27.

6 References

Adamis G, Papaioannou MG, Giamarellos-Bourboulis EJ, Gargalianos P, Kosmidis J & Giamarellou H (2004) Pharmacokinetic interactions of ceftazidime, imipenem and aztreonam with amikacin in healthy volunteers. *Int J Antimicrob Agents* **23**: 144-149.

Alber BE (2011) Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl Microbiol Biotechnol* **89**: 17-25.

Alonso S, Pethe K, Russell DG & Purdy GE (2007) Lysosomal killing of Mycobacterium mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proc Natl Acad Sci U S A* **104**: 6031-6036.

Andries K, Verhasselt P, Guillemont J*, et al.* (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**: 223-227.

Arabolaza A, Shillito ME, Lin T-W*, et al.* (2010) Crystal Structures and Mutational Analyses of Acyl-CoA Carboxylase β Subunit of*Streptomyces coelicolor*. *Biochemistry* **49**: 7367-7376.

Armstrong JA & Hart PD (1975) Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. *J Exp Med* **142**: 1-16.

Arnold K, Bordoli L, Kopp J & Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195-201.

Asselineau C, Asselineau J, Laneelle G & Laneelle MA (2002) The biosynthesis of mycolic acids by Mycobacteria: current and alternative hypotheses. *Prog Lipid Res* **41**: 501-523.

Asselineau J & Laneelle G (1998) Mycobacterial lipids: a historical perspective. *Front Biosci* **3**: e164-174.

Ataka M & Asai M (1990) Analysis of the nucleation and crystal growth kinetics of lysozyme by a theory of self-assembly. *Biophys J* **58**: 807-811.

Attwood PV & Wallace JC (2002) Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc Chem Res* **35**: 113-120.

Barry CE, 3rd, Lee RE, Mdluli K, Sampson AE, Schroeder BG, Slayden RA & Yuan Y (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* **37**: 143-179.

Battye TG, Kontogiannis L, Johnson O, Powell HR & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**: 271-281.

Beer A (1852) Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Annalen der Physik und Chemie. **86:** 78–88

Berg IA, Kockelkorn D, Ramos-Vera WH*, et al.* (2010) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**: 447-460.

Berrow NS, Bussow K, Coutard B*, et al.* (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr D Biol Crystallogr* **62**: 1218-1226.

Bhatt A, Kremer L, Dai AZ, Sacchettini JC & Jacobs WR, Jr. (2005) Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *J Bacteriol* **187**: 7596-7606.

Bhatt A, Molle V, Besra GS, Jacobs WR, Jr. & Kremer L (2007) The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol Microbiol* **64**: 1442-1454.

Bhatt A, Fujiwara N, Bhatt K*, et al.* (2007) Deletion of kasB in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci U S A* **104**: 5157-5162.

Bieniossek C, Nie Y, Frey D*, et al.* (2009) Automated unrestricted multigene recombineering for multiprotein complex production. *Nat Methods* **6**: 447-450.

Blanchard JS (1996) Molecular mechanisms of drug resistance in *Mycobacterium tuberculosis*. *Annu Rev Biochem* **65**: 215-239.

Blumberg HM, Burman WJ, Chaisson RE*, et al.* (2003) American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. *Am J Respir Crit Care Med* **167**: 603-662.

Blow D (2002) Outline of crystallography for biologists. Oxford University Press.

Bourenkov GP & Popov AN (2010) Optimization of data collection taking radiation damage into account. *Acta Crystallogr D Biol Crystallogr* **66**: 409-419.

Bragg, W.L. (1913). "The Diffraction of Short Electromagnetic Waves by a Crystal". Proceedings of the Cambridge Philosophical Society **17:** 43–57.

Bramwell H, Hunter IS, Coggins JR & Nimmo HG (1996) Propionyl-CoA carboxylase from *Streptomyces coelicolor* A3(2): cloning of the gene encoding the biotin-containing subunit. *Microbiology* **142 (Pt 3)**: 649-655.

Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **83**: 91-97.

Brennan PJ & Nikaido H (1995) The envelope of mycobacteria. *Annu Rev Biochem* **64**: 29-63.

Brennan PJ & Crick DC (2007) The cell-wall core of *Mycobacterium tuberculosis* in the context of drug discovery. *Curr Top Med Chem* **7**: 475-488.

Briken V, Porcelli SA, Besra GS & Kremer L (2004) Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Mol Microbiol* **53**: 391-403.

Brown AK, Bhatt A, Singh A, Saparia E, Evans AF & Besra GS (2007) Identification of the dehydratase component of the mycobacterial mycolic acid-synthesizing fatty acid synthase-II complex. *Microbiology* **153**: 4166-4173.

Burman WJ, Gallicano K & Peloquin C (2001) Comparative pharmacokinetics and pharmacodynamics of the rifamycin antibacterials. *Clin Pharmacokinet* **40**: 327-341.

Busso D, Peleg Y, Heidebrecht T*, et al.* (2011) Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: a benchmarking study. *J Struct Biol* **175**: 159-170.

Camacho LR, Constant P, Raynaud C*, et al.* (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**: 19845-19854.

Camus JC, Pryor MJ, Medigue C & Cole ST (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**: 2967-2973.

Casenghi M, Cole ST & Nathan CF (2007) New approaches to filling the gap in tuberculosis drug discovery. *PLoS Med* **4**: e293.

Chan J, Xing Y, Magliozzo RS & Bloom BR (1992) Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. *J Exp Med* **175**: 1111-1122.

Chan J, Tanaka K, Carroll D, Flynn J & Bloom BR (1995) Effects of nitric oxide synthase inhibitors on murine infection with *Mycobacterium tuberculosis*. *Infect Immun* **63**: 736-740.

Chayen NE (2005) Methods for separating nucleation and growth in protein crystallisation. *Prog Biophys Mol Biol* **88**: 329-337.

Chayen NE & Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**: 147-153.

Chayen NE, Stewart PDS, Maeder DL & Blow DM (1990) An Automated-System for Microbatch Protein Crystallization and Screening. *Journal of Applied Crystallography* **23**: 297-302.

Chayen NE, Boggon TJ, Cassetta A*, et al.* (1996) Trends and challenges in experimental macromolecular crystallography. *Q Rev Biophys* **29**: 227-278.

Chen GJ, Qiu N, Karrer C, Caspers P & Page MG (2000) Restriction site-free insertion of PCR products directionally into vectors. *Biotechniques* **28**: 498-500, 504-495.

Cole ST, Brosch R, Parkhill J*, et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763.

Cox JS, Chen B, McNeil M & Jacobs WR, Jr. (1999) Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**: 79-83.

Cronan JE, Jr. & Waldrop GL (2002) Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res* **41**: 407-435.

Crowle AJ, Ross EJ & May MH (1987) Inhibition by 1,25(OH)2-vitamin D3 of the multiplication of virulent tubercle bacilli in cultured human macrophages. *Infect Immun* **55**: 2945-2950.

Crowle AJ, Dahl R, Ross E & May MH (1991) Evidence that vesicles containing living, virulent *Mycobacterium tuberculosis* or *Mycobacterium avium* in cultured human macrophages are not acidic. *Infect Immun* **59**: 1823-1831.

Cynamon M, Sklaney MR & Shoen C (2007) Gatifloxacin in combination with rifampicin in a murine tuberculosis model. *J Antimicrob Chemother* **60**: 429-432.

Daffe M & Draper P (1998) The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* **39**: 131-203.

Dakshinamurti K & Chauhan J (1988) Regulation of biotin enzymes. *Annu Rev Nutr* **8**: 211-233.

Dalton T, Cegielski P, Akksilp S*, et al.* (2012) Prevalence of and risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in eight countries: a prospective cohort study. *Lancet* **380**: 1406-1417.

Daniel J, Oh TJ, Lee CM & Kolattukudy PE (2007) AccD6, a member of the Fas II locus, is a functional carboxyltransferase subunit of the acyl-coenzyme A carboxylase in *Mycobacterium tuberculosis*. *J Bacteriol* **189**: 911-917.

Dannenberg AM, Jr. (1994) Roles of cytotoxic delayed-type hypersensitivity and macrophage-activating cell-mediated immunity in the pathogenesis of tuberculosis. *Immunobiology* **191**: 461-473.

Debye P (1947) Molecular-weight determination by light scattering. *J Phys Colloid Chem* **51**: 18-32.

Delbridge LM & O'Riordan MX (2007) Innate recognition of intracellular bacteria. *Curr Opin Immunol* **19**: 10-16.

Delgado MA, Elmaoued RA, Davis AS, Kyei G & Deretic V (2008) Toll-like receptors control autophagy. *EMBO J* **27**: 1110-1121.

Diacovich L, Peiru S, Kurth D, Rodriguez E, Podesta F, Khosla C & Gramajo H (2002) Kinetic and Structural Analysis of a New Group of Acyl-CoA Carboxylases Found in *Streptomyces coelicolor* A3(2). *Journal of Biological Chemistry* **277**: 31228-31236.

Diacovich L, Mitchell DL, Pham H*, et al.* (2004) Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* **43**: 14027-14036.

Doluisio JT & Swintosky JV (1965) Biopharmaceutics. I. Factors influencing drug distribution in the body. Penetration of drugs through biological membranes--storage and elimination--factors influencing routes of drug administration. *Am J Pharm Sci Support Public Health* **137**: 144-168.

Draper P (1998) The outer parts of the mycobacterial envelope as permeability barriers. *Front Biosci* **3**: D1253-1261.

Ducati RG, Basso LA & Santos DS (2007) Mycobacterial shikimate pathway enzymes as targets for drug design. *Curr Drug Targets* **8**: 423-435.

Ehebauer MT & Wilmanns M (2011) The progress made in determining the *Mycobacterium tuberculosis* structural proteome. *Proteomics* **11**: 3128-3133.

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

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

Erb TJ, Berg IA, Brecht V, Muller M, Fuchs G & Alber BE (2007) Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc Natl Acad Sci U S A* **104**: 10631-10636.

Erfle JD (1973) Acetyl-CoA and propionyl-CoA carboxylation by Mycobacterium phlei. Partial purification and some properties of the enzyme. *Biochim Biophys Acta* **316**: 143-155.

Ericsson UB, Hallberg BM, Detitta GT, Dekker N & Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**: 289-298.

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**: 72-82.

Fan C, Chou CY, Tong L & Xiang S (2012) Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J Biol Chem* **287**: 9389-9398.

Fenn JB, Mann M, Meng CK, Wong SF & Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**: 64-71.

Fenton MJ & Vermeulen MW (1996) Immunopathology of tuberculosis: roles of macrophages and monocytes. *Infect Immun* **64**: 683-690.

Ferreras JA, Ryu JS, Di Lello F, Tan DS & Quadri LE (2005) Small-molecule inhibition of siderophore biosynthesis in *Mycobacterium tuberculosis* and Yersinia pestis. *Nat Chem Biol* **1**: 29-32.

Flynn JL & Chan J (2005) What's good for the host is good for the bug. *Trends Microbiol* **13**: 98-102.

Frehel C, de Chastellier C, Lang T & Rastogi N (1986) Evidence for inhibition of fusion of lysosomal and prelysosomal compartments with phagosomes in macrophages infected with pathogenic *Mycobacterium avium*. *Infect Immun* **52**: 252-262.

Fujii K, Saito H, Tomioka H, Mae T & Hosoe K (1995) Mechanism of action of antimycobacterial activity of the new benzoxazinorifamycin KRM-1648. *Antimicrob Agents Chemother* **39**: 1489-1492.

Gago G, Kurth D, Diacovich L, Tsai SC & Gramajo H (2006) Biochemical and structural characterization of an essential acyl coenzyme A carboxylase from *Mycobacterium tuberculosis*. *J Bacteriol* **188**: 477-486.

Gago G, Diacovich L, Arabolaza A, Tsai S-C & Gramajo H (2011) Fatty acid biosynthesis in actinomycetes. *FEMS Microbiology Reviews* **35**: 475-497.

Gande R, Dover LG, Krumbach K, Besra GS, Sahm H, Oikawa T & Eggeling L (2007) The Two Carboxylases of *Corynebacterium glutamicum* Essential for Fatty Acid and Mycolic Acid Synthesis. *Journal of Bacteriology* **189**: 5257-5264.

Gande R, Gibson KJC, Brown AK*, et al.* (2004) Acyl-CoA Carboxylases (accD2 and accD3), Together with a Unique Polyketide Synthase (Cg-pks), Are Key to Mycolic Acid Biosynthesis in Corynebacterianeae Such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **279**: 44847-44857.

Gatfield J & Pieters J (2000) Essential role for cholesterol in entry of mycobacteria into macrophages. *Science* **288**: 1647-1650.

Gaynor CD, McCormack FX, Voelker DR, McGowan SE & Schlesinger LS (1995) Pulmonary surfactant protein A mediates enhanced phagocytosis of *Mycobacterium tuberculosis* by a direct interaction with human macrophages. *J Immunol* **155**: 5343-5351.

Girardin SE, Travassos LH, Herve M*, et al.* (2003) Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. *J Biol Chem* **278**: 41702-41708.

Glickman MS, Cox JS & Jacobs WR, Jr. (2000) A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell* **5**: 717-727.

Gokhale RS, Saxena P, Chopra T & Mohanty D (2007) Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. *Nat Prod Rep* **24**: 267-277.

Goldstone RM, Moreland NJ, Bashiri G, Baker EN & Shaun Lott J (2008) A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. *Protein Expr Purif* **57**: 81-87.

Goren MB (1972) Mycobacterial lipids: selected topics. *Bacteriol Rev* **36**: 33-64.

Gouet P, Robert X & Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**: 3320-3323.

Greenfield NJ (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**: 2876-2890.

Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ & Sassetti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**: e1002251.

Guerardel Y, Maes E, Elass E*, et al.* (2002) Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains. *J Biol Chem* **277**: 30635-30648.

Guinier A (1939). "La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques." Ann. Phys. (Paris) **12:** 161-237.

Guruprasad K, Reddy BV & Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* **4**: 155-161.

Gutierrez MG, Master SS, Singh SB, Taylor GA, Colombo MI & Deretic V (2004) Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell* **119**: 753-766.

Haas DJ & Rossmann MG (1970) Crystallographic studies on lactate dehydrogenase at-75 degrees C. *Acta Crystallogr B* **26**: 998-1004.

Haase FC, Henrikson KP, Treble DH & Allen SH (1982) The subunit structure and function of the propionyl coenzyme A carboxylase of *Mycobacterium smegmatis*. *J Biol Chem* **257**: 11994-11999.

Hall PR, Wang YF, Rivera-Hainaj RE, Zheng X, Pustai-Carey M, Carey PR & Yee VC (2003) Transcarboxylase 12S crystal structure: hexamer assembly and substrate binding to a multienzyme core. *EMBO J* **22**: 2334-2347.

Henrikson KP & Allen SH (1979) Purification and subunit structure of propionyl coenzyme A carboxylase of *Mycobacterium smegmatis*. *J Biol Chem* **254**: 5888-5891.

Holden HM, Benning MM, Haller T & Gerlt JA (2001) The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc Chem Res* **34**: 145-157.

Holton SJ, King-Scott S, Eddine AN, Kaufmann SHE & Wilmanns M (2006) Structural diversity in the six-fold redundant set of acyl-CoA carboxyltransferases in *Mycobacterium tuberculosis*. *FEBS Letters* **580**: 6898-6902.

Huang CS, Ge P, Zhou ZH & Tong L (2011) An unanticipated architecture of the 750-kDa α6β6 holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature* **481**: 219-223.

Huang CS, Sadre-Bazzaz K, Shen Y, Deng B, Zhou ZH & Tong L (2010) Crystal structure of the α6β6 holoenzyme of propionyl-coenzyme A carboxylase. *Nature* **466**: 1001-1005.

Hummel W (1999) Large-scale applications of NAD(P)-dependent oxidoreductases: recent developments. *Trends Biotechnol* **17**: 487-492.

Hunaiti AA & Kolattukudy PE (1984) Source of methylmalonyl-coenzyme A for erythromycin synthesis: methylmalonyl-coenzyme A mutase from Streptomyces erythreus. *Antimicrob Agents Chemother* **25**: 173-178.

Hunaiti AR & Kolattukudy PE (1982) Isolation and characterization of an acyl-coenzyme A carboxylase from an erythromycin-producing Streptomyces erythreus. *Arch Biochem Biophys* **216**: 362-371.

Hunter RL, Olsen MR, Jagannath C & Actor JK (2006) Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. *Ann Clin Lab Sci* **36**: 371-386.

Inoue H, Nojima H & Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.

Jackson M, Stadthagen G & Gicquel B (2007) Long-chain multiple methyl-branched fatty acid-containing lipids of *Mycobacterium tuberculosis*: biosynthesis, transport, regulation and biological activities. *Tuberculosis (Edinb)* **87**: 78-86.

Jia L, Tomaszewski JE, Hanrahan C*, et al.* (2005) Pharmacodynamics and pharmacokinetics of SQ109, a new diamine-based antitubercular drug. *Br J Pharmacol* **144**: 80-87.

Jitrapakdee S, St Maurice M, Rayment I, Cleland WW, Wallace JC & Attwood PV (2008) Structure, mechanism and regulation of pyruvate carboxylase. *Biochem J* **413**: 369-387.

Johnston K, Clements A, Venkataramani RN, Trievel RC & Marmorstein R (2000) Coexpression of proteins in bacteria using T7-based expression plasmids: expression of heteromeric cell-cycle and transcriptional regulatory complexes. *Protein Expr Purif* **20**: 435-443.

Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**: 133-144.

Kaplan G, Post FA, Moreira AL*, et al.* (2003) *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. *Infect Immun* **71**: 7099-7108.

Kapopoulou A, Lew JM & Cole ST (2011) The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)* **91**: 8-13.

Kaufmann SH & Young DB (1992) Vaccination against tuberculosis and leprosy. *Immunobiology* **184**: 208-229.

Kaufmann SH, Hussey G & Lambert PH (2010) New vaccines for tuberculosis. *Lancet* **375**: 2110-2119.

Khasnobis S, Escuyer VE & Chatterjee D (2002) Emerging therapeutic targets in tuberculosis: post-genomic era. *Expert Opin Ther Targets* **6**: 21-40.

Khomyakova M, Bukmez O, Thomas LK, Erb TJ & Berg IA (2011) A methylaspartate cycle in haloarchaea. *Science* **331**: 334-337.

Kikuchi S, Rainwater DL & Kolattukudy PE (1992) Purification and characterization of an unusually large fatty acid synthase from *Mycobacterium tuberculosis* var. bovis BCG. *Arch Biochem Biophys* **295**: 318-326.

Kim Y, Quartey P, Li H*, et al.* (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods* **5**: 853-854.

Kimura Y, Kojyo T, Kimura I & Sato M (1998) Propionyl-CoA carboxylase of Myxococcus xanthus: catalytic properties and function in developing cells. *Arch Microbiol* **170**: 179-184.

Knowles JR (1989) The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* **58**: 195-221.

Kobayashi M, Kubota M & Matsuura Y (1999) Crystallization and improvement of crystal quality for x-ray diffraction of maltooligosyl trehalose synthase by reductive methylation of lysine residues. *Acta Crystallogr D Biol Crystallogr* **55**: 931-933.

Kochhar S, Hunziker PE, Leong-Morgenthaler P & Hottinger H (1992) Evolutionary relationship of NAD(+)-dependent D-lactate dehydrogenase: comparison of primary structure of 2-hydroxy acid dehydrogenases. *Biochem Biophys Res Commun* **184**: 60-66.

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**: 1277-1282.

Kremer L, Nampoothiri KM, Lesjean S*, et al.* (2001) Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of *Mycobacterium tuberculosis* fatty acid synthase II. *J Biol Chem* **276**: 27967-27974.

Kremer L, Dover LG, Carrere S*, et al.* (2002) Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from *Mycobacterium tuberculosis*. *Biochem J* **364**: 423-430.

Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774-797.

Kurth DG, Gago GM, de la Iglesia A*, et al.* (2009) ACCase 6 is the essential acetyl-CoA carboxylase involved in fatty acid and mycolic acid biosynthesis in mycobacteria. *Microbiology* **155**: 2664-2675.

Laskowski RA, Moss DS & Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **231**: 1049-1067.

Lederer E, Adam A, Ciorbaru R, Petit JF & Wietzerbin J (1975) Cell walls of Mycobacteria and related organisms; chemistry and immunostimulant properties. *Mol Cell Biochem* **7**: 87-104.

Lehmann J (1946) Para-aminosalicylic acid in the treatment of tuberculosis. *Lancet* **1**: 15.

Lemassu A & Daffe M (1994) Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis*. *Biochem J* **297 (Pt 2)**: 351-357.

Li C, Schwabe JW, Banayo E & Evans RM (1997) Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc Natl Acad Sci U S A* **94**: 2278-2283.

Lin TW, Melgar MM, Kurth D*, et al.* (2006) Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences* **103**: 3072-3077.

Linding R, Russell RB, Neduva V & Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**: 3701-3708.

Lindqvist Y & Schneider G (1996) Protein-biotin interactions. *Curr Opin Struct Biol* **6**: 798-803.

Liu LY, Bates ME, Jarjour NN, Busse WW, Bertics PJ & Kelly EA (2007) Generation of Th1 and Th2 chemokines by human eosinophils: evidence for a critical role of TNF-alpha. *J Immunol* **179**: 4840-4848.

Liu PT, Stenger S, Li H*, et al.* (2006) Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* **311**: 1770-1773.

Lombard J & Moreira D (2011) Early evolution of the biotin-dependent carboxylase family. *BMC Evol Biol* **11**: 232.

Marmiesse M, Brodin P, Buchrieser C*, et al.* (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**: 483-496.

Marrakchi H, Ducasse S, Labesse G*, et al.* (2002) MabA (FabG1), a *Mycobacterium tuberculosis* protein involved in the long-chain fatty acid elongation system FAS-II. *Microbiology* **148**: 951-960.

Martineau AR, Wilkinson RJ, Wilkinson KA*, et al.* (2007) A single dose of vitamin D enhances immunity to mycobacteria. *Am J Respir Crit Care Med* **176**: 208-213.

Mathur M & Kolattukudy PE (1992) Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. bovis Bacillus Calmette-Guerin. *J Biol Chem* **267**: 19388-19395.

Matsumoto M, Hashizume H, Tomishige T*, et al.* (2006) OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. *PLoS Med* **3**: e466.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**: 658-674.

McIlleron H, Wash P, Burger A, Norman J, Folb PI & Smith P (2006) Determinants of rifampin, isoniazid, pyrazinamide, and ethambutol pharmacokinetics in a cohort of tuberculosis patients. *Antimicrob Agents Chemother* **50**: 1170-1177.

McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ*, et al.* (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**: 735-738.

McNeil M, Daffe M & Brennan PJ (1991) Location of the mycolyl ester substituents in the cell walls of mycobacteria. *J Biol Chem* **266**: 13217-13223.

McPherson A (1999) Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press.

Mdluli K & Spigelman M (2006) Novel targets for tuberculosis drug discovery. *Curr Opin Pharmacol* **6**: 459-467.

Michaelis L, Menten ML, Johnson KA & Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **50**: 8264-8269.

Minnikin DE, Minnikin SM, Goodfellow M & Stanford JL (1982) The mycolic acids of Mycobacterium chelonei. *J Gen Microbiol* **128**: 817-822.

Misaki A, Seto N & Azuma I (1974) Structure and immunological properties of D-arabino-D-galactans isolated from cell walls of Mycobacterium species. *J Biochem* **76**: 15-27.

Monfeli RR & Beeson C (2007) Targeting iron acquisition by *Mycobacterium tuberculosis*. *Infect Disord Drug Targets* **7**: 213-220.

Mosmann TR & Coffman RL (1989) TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu Rev Immunol* **7**: 145-173.

Mosmann TR, Cherwinski H, Bond MW, Giedlin MA & Coffman RL (1986) Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol* **136**: 2348-2357.

Mudd S, Polevitzky K, Anderson TF & Chambers LA (1941) Bacterial Morphology as Shown by the Electron Microscope: II. The Bacterial Cell-wall in the Genus Bacillus. *J Bacteriol* **42**: 251-264.

Mueller-Dieckmann J (2006) The open-access high-throughput crystallization facility at EMBL Hamburg. *Acta Crystallogr D Biol Crystallogr* **62**: 1446-1452.

Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.

Musayev F, Sachdeva S, Scarsdale JN, Reynolds KA & Wright HT (2005) Crystal structure of a substrate complex of *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein synthase III (FabH) with lauroyl-coenzyme A. *J Mol Biol* **346**: 1313-1321.

Navarathna DH, Harris SD, Roberts DD & Nickerson KW (2010) Evolutionary aspects of urea utilization by fungi. *FEMS Yeast Res* **10**: 209-213.

Nicholson S, Bonecini-Almeida Mda G, Lapa e Silva JR*, et al.* (1996) Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. *J Exp Med* **183**: 2293-2302.

Niu C, Yin J, Cherney MM & James MN (2011) Expression, purification and preliminary crystallographic analysis of Rv2247, the beta subunit of acyl-CoA carboxylase (ACCD6) from *Mycobacterium tuberculosis*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**: 1637-1640.

Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT & Wilmanns M (2011) Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1ΔC expression strain. *BMC Biotechnology* **11**: 27.

Noll H & Bloch H (1955) Studies on the chemistry of the cord factor of *Mycobacterium tuberculosis*. *J Biol Chem* **214**: 251-265.

Noll H, Bloch H, Asselineau J & Lederer E (1956) The chemical structure of the cord factor of *Mycobacterium tuberculosis*. *Biochim Biophys Acta* **20**: 299-309.

Noss EH, Pai RK, Sellati TJ*, et al.* (2001) Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19-kDa lipoprotein of *Mycobacterium tuberculosis*. *J Immunol* **167**: 910-918.

O'Brien L, Carmichael J, Lowrie DB & Andrew PW (1994) Strains of *Mycobacterium tuberculosis* differ in susceptibility to reactive nitrogen intermediates in vitro. *Infect Immun* **62**: 5187-5190.

Oh TJ, Daniel J, Kim H, Sirakova T & Kolattukudy PE (2006) Identification and Characterization of Rv3281 as a Novel Subunit of a Biotin-dependent Acyl-CoA Carboxylase in *Mycobacterium tuberculosis* H37Rv. *Journal of Biological Chemistry* **281**: 3899-3908.

Pantoliano MW, Petrella EC, Kwasnoski JD*, et al.* (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**: 429-440.

Pappin DJ, Hojrup P & Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**: 327-332.

Patterson, A.L. (1935) A direct method for the determination of the components of interatomic distances in crystals. *Z. Krist.* **A90:** 517-542.

Paul TR & Beveridge TJ (1992) Reevaluation of envelope profiles and cytoplasmic ultrastructure of mycobacteria processed by conventional embedding and freeze-substitution protocols. *J Bacteriol* **174**: 6508-6517.

Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**: 961-1004.

Perrakis A & Romier C (2008) Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview. *Methods Mol Biol* **426**: 247-256.

Peterson DO & Bloch K (1977) *Mycobacterium smegmatis* fatty acid synthetase. Long chain transacylase chain length specificity. *J Biol Chem* **252**: 5735-5739.

Petoukhov MV & Svergun DI (2013) Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int J Biochem Cell Biol* **45**: 429-437.

Peyron P, Bordier C, N'Diaye EN & Maridonneau-Parini I (2000) Nonopsonic phagocytosis of Mycobacterium kansasii by human neutrophils depends on cholesterol and is mediated by CR3 associated with glycosylphosphatidylinositol-anchored proteins. *J Immunol* **165**: 5186-5191.

Pitarque S, Larrouy-Maumus G, Payre B, Jackson M, Puzo G & Nigou J (2008) The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface. *Tuberculosis (Edinb)* **88**: 560-565.

Porath J & Flodin P (1959) Gel filtration: a method for desalting and group separation. *Nature* **183**: 1657-1659.

Porath J, Carlsson J, Olsson I & Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**: 598-599.

Porod G (1982) General Theory. In *Small Angle X-ray Scattering* (ed. O. Glatter and O. Kratky), pp. 17-51. London: Academic Press.

Portevin D, de Sousa-D'Auria C, Montrozier H*, et al.* (2005) The acyl-AMP ligase FadD32 and AccD4-containing acyl-CoA carboxylase are required for the synthesis of mycolic acids and essential for mycobacterial growth: identification of the carboxylation product and determination of the acyl-CoA carboxylase components. *J Biol Chem* **280**: 8862-8874.

Pratscher J, Dumont MG & Conrad R (2011) Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. *Proc Natl Acad Sci U S A* **108**: 4170-4175.

Primm TP, Andersen SJ, Mizrahi V, Avarbock D, Rubin H & Barry CE, 3rd (2000) The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. *J Bacteriol* **182**: 4889-4898.

Quemard A, Sacchettini JC, Dessen A, Vilcheze C, Bittman R, Jacobs WR, Jr. & Blanchard JS (1995) Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*. *Biochemistry* **34**: 8235-8241.

Rodriguez E & Gramajo H (1999) Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of *Streptomyces coelicolor* A3(2). *Microbiology* **145 (Pt 11)**: 3109-3119.

Rodriguez E, Banchio C, Diacovich L, Bibb MJ & Gramajo H (2001) Role of an Essential Acyl Coenzyme A Carboxylase in the Primary and Secondary Metabolism of *Streptomyces coelicolor* A3(2). *Applied and Environmental Microbiology* **67**: 4166-4176.

Roessle MW, Klaering R, Ristau U*, et al.* (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography* **40**: S190-S194.

Romier C, Ben Jelloul M, Albeck S*, et al.* (2006) Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr* **62**: 1232-1242.

Rook GA, Steele J, Fraher L, Barker S, Karmali R, O'Riordan J & Stanford J (1986) Vitamin D3, gamma interferon, and control of proliferation of *Mycobacterium tuberculosis* by human monocytes. *Immunology* **57**: 159-163.

Russell AD & Chopra I (1996) *Understanding antibacterial action and resistance*. Ellis Horwood, London ; New York.

Russell DG (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* **5**: 39-47.

Russell DG, Cardona PJ, Kim MJ, Allain S & Altare F (2009) Foamy macrophages and the progression of the human tuberculosis granuloma. *Nat Immunol* **10**: 943-948.

Ryndak M, Wang S & Smith I (2008) PhoP, a key player in *Mycobacterium tuberculosis* virulence. *Trends Microbiol* **16**: 528-534.

Rypniewski WR, Holden HM & Rayment I (1993) Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **32**: 9851-9858.

Sassetti CM, Boyd DH & Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Schaeffer ML, Agnihotri G, Volker C, Kallender H, Brennan PJ & Lonsdale JT (2001) Purification and biochemical characterization of the *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein synthases KasA and KasB. *J Biol Chem* **276**: 47029-47037.

Schlesinger LS (1993) Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. *J Immunol* **150**: 2920-2930.

Scorpio A & Zhang Y (1996) Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* **2**: 662-667.

Sibley LD, Hunter SW, Brennan PJ & Krahenbuhl JL (1988) Mycobacterial lipoarabinomannan inhibits gamma interferon-mediated activation of macrophages. *Infect Immun* **56**: 1232-1236.

Sirakova TD, Thirumala AK, Dubey VS, Sprecher H & Kolattukudy PE (2001) The *Mycobacterium tuberculosis* pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J Biol Chem* **276**: 16833-16839.

Sirakova TD, Dubey VS, Kim HJ, Cynamon MH & Kolattukudy PE (2003) The largest open reading frame (pks12) in the *Mycobacterium tuberculosis* genome is involved in pathogenesis and dimycocerosyl phthiocerol synthesis. *Infect Immun* **71**: 3794-3801.

Slayden RA & Barry CE, 3rd (2002) The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **82**: 149-160.

Smejkalova H, Erb TJ & Fuchs G (2010) Methanol assimilation in Methylobacterium extorquens AM1: demonstration of all enzymes and their regulation. *PLoS One* **5**.

Smith NH, Hewinson RG, Kremer K, Brosch R & Gordon SV (2009) Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* **7**: 537-544.

Smith S, Witkowski A & Joshi AK (2003) Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res* **42**: 289-317.

St Maurice M, Reinhardt L, Surinya KH, Attwood PV, Wallace JC, Cleland WW & Rayment I (2007) Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* **317**: 1076-1079.

Stover CK, Warrener P, VanDevanter DR*, et al.* (2000) A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. *Nature* **405**: 962-966.

Strope PK, Nickerson KW, Harris SD & Moriyama EN (2011) Molecular evolution of urea amidolyase and urea carboxylase in fungi. *BMC Evol Biol* **11**: 80.

Svergun DI (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* **25**: 495-503.

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

Telenti A, Philipp WJ, Sreevatsan S*, et al.* (1997) The emb operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nat Med* **3**: 567-570.

Thoma-Uszynski S, Stenger S, Takeuchi O*, et al.* (2001) Induction of direct antimicrobial activity through mammalian toll-like receptors. *Science* **291**: 1544-1547.

Tolia NH & Joshua-Tor L (2006) Strategies for protein coexpression in *Escherichia coli*. *Nat Methods* **3**: 55-64.

Tong L (2005) Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell Mol Life Sci* **62**: 1784-1803.

Tong L (2013) Structure and function of biotin-dependent carboxylases. *Cell Mol Life Sci* **70**: 863-891.

Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D & Gokhale RS (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* **428**: 441-445.

Trivedi OA, Arora P, Vats A*, et al.* (2005) Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. *Mol Cell* **17**: 631-643.

van Crevel R, Ottenhoff TH & van der Meer JW (2002) Innate immunity to *Mycobacterium tuberculosis*. *Clin Microbiol Rev* **15**: 294-309.

van de Guchte M, Penaud S, Grimaldi C*, et al.* (2006) The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci U S A* **103**: 9274-9279.

Veyron-Churlet R, Zanella-Cleon I, Cohen-Gonsaud M, Molle V & Kremer L (2010) Phosphorylation of the *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein reductase MabA regulates mycolic acid biosynthesis. *J Biol Chem* **285**: 12714-12725.

Via LE, Lin PL, Ray SM*, et al.* (2008) Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* **76**: 2333-2340.

Vilcheze C, Morbidoni HR, Weisbrod TR, Iwamoto H, Kuo M, Sacchettini JC & Jacobs WR, Jr. (2000) Inactivation of the inhA-encoded fatty acid synthase II (FASII) enoyl-acyl carrier protein reductase induces accumulation of the FASI end products and cell lysis of *Mycobacterium smegmatis*. *J Bacteriol* **182**: 4059-4067.

Wakil SJ, Stoops JK & Joshi VC (1983) Fatty acid synthesis and its regulation. *Annu Rev Biochem* **52**: 537-579.

Walter TS, Meier C, Assenberg R*, et al.* (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**: 1617-1622.

Wayne LG & Sramek HA (1994) Metronidazole is bactericidal to dormant cells of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **38**: 2054-2058.

Wendt KS, Schall I, Huber R, Buckel W & Jacob U (2003) Crystal structure of the carboxyltransferase subunit of the bacterial sodium ion pump glutaconyl-coenzyme A decarboxylase. *EMBO J* **22**: 3493-3502.

Wishart DS, Knox C, Guo AC*, et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**: D668-672.

Wlodawer A, Minor W, Dauter Z & Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **275**: 1-21.

Wright DH, Brown GH, Peterson ML & Rotschafer JC (2000) Application of fluoroquinolone pharmacodynamics. *J Antimicrob Chemother* **46**: 669-683.

Xiang S & Tong L (2008) Crystal structures of human and Staphylococcus aureus pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nat Struct Mol Biol* **15**: 295-302.

Zeng J, Zhang L, Li Y, Wang Y, Wang M, Duan X & He ZG (2010) Over-producing soluble protein complex and validating protein-protein interaction through a new bacterial co-expression system. *Protein Expr Purif* **69**: 47-53.

Zhang H, Boghigian BA & Pfeifer BA (2010) Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in *Escherichia coli*. *Biotechnol Bioeng* **105**: 567-573.

Zhang H, Yang Z, Shen Y & Tong L (2003) Crystal Structure of the Carboxyltransferase Domain of Acetyl-Coenzyme A Carboxylase. *Science* **299**: 2064-2067.

Zhang Y, Heym B, Allen B, Young D & Cole S (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* **358**: 591-593.

Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G & Daffe M (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**: 5672-5680.

7 Acknowledgements

"When you rise in the morning, give thanks for the light, for your life, for your strength. Give thanks for your food and for the joy of living. If you see no reason to give thanks, the fault lies in yourself." - Tecumseh

I express my heartfelt gratitude to my supervisor **Dr. Matthias Wilmanns** for having allowed me to work on a challenging project and for his esteemed and valuable guidance over the course of my PhD. My sincere thanks to the members of my Thesis Advisory Committee - **Dr. Rob Meijers**, **Prof. Dr. Robert Russell,** and **Dr. Carsten Sachse** - for all their constructive feedback and overwhelming, warm support.

I profusely thank my associate advisor **Matthias Ehebauer** for his supervision in the lab, for the numerous interesting discussions, for the overall care he extended throughout my EMBL time, and for being the best colleague one could ever have.

Elke Noens, Yusuf Akhter, Barbara Tizzano, and other members of the TB subgroup have been wonderful friends and sources of constant encouragement. The world's greatest lab buddy, Koen Temmerman, invariably had the right words to make everything easier. Georg Graf zu Eulenburg and Daniel Passon kept inventing amazing ways of lighting up otherwise low and dull moments.

I am thankful to Anne Tuukkanen and Arjen Jacobi for having extended kind help with SAXS and EM analyses. I am highly appreciative of the advice offered by Spyros Chatziefthimiou, Marie-Cécile Pelissier, and Guillaume Pompidor in solving crystallography-related problems. I owe a huge deal of thanks to the wet lab management (Annabel Parret, Vivian Pogenberg, Morlin Milewski, and Alexandra Koetter) and the SPC facility (Dr. Meijers, Stephane Boivin, and team) for their relentless hard work in nurturing an exceptional atmosphere conducive to research. Very special thanks to the outstanding EMBL Hamburg admin and service teams, the EMBL Graduate Office and the PhD Office, Faculty of Biosciences, University of Heidelberg for their excellent administrative and moral support. Financial assistance from the European Union's Seventh Framework Programme (through NATT and SysteMTb grants to Dr. Wilmanns) is gratefully acknowledged.

The ups and downs have all been merry rides, thanks to Senthilprasad's unswerving optimism and confidence. My mentor Brinda Arumugam cannot be thanked enough for her boundless positivity and inspiration. Deepitha and Vijayan have been making my life in Germany, right from day one, so much more memorable and fun. Senthilvelrajan has been a solution giver and patient listener, at the receiving end of many rants. The journey would have been impossible without incessant boosts from Shankar, who 'has always had to be the one who knew, and who had the burden of knowing'.

I wholeheartedly thank my brothers Sivakumar and Niranjankumar, my sister-in-law Shunmugapriya, and the entire extended family for their everlasting patience, faith, and love. This work is dedicated to my parents **Nagajothi** and **Anandhakrishnan**, who gave me life, and to my precious nephew **Skandan**, who has given life a whole new meaning.