Characterisation of orphan cytochrome P450s from *Mycobacterium tuberculosis* H37Rv

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2017

NUR DAYANA NISBAR

SCHOOL OF CHEMISTRY

Blank page

TABLE OF CONTENT

WORD COUNT: 46458

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ºC	Degree Celsius
Å	Angstrom (10^{-10} m)
AD	Adrenodoxin
ADR	Adrenodoxin reductase
BCG	Bacille Calmette-Guérin
bp	Base pair
CD	Circular dichroism
CDP	Cyclic dipeptide
CO	Carbon monoxide
CMPG	Carr-Purcell-Meiboom-Gill
CPR	Cytochrome P450 reductase
CYP or P450	Cytochrome P450
cYY	Cyclo-*L*-Tyr-*L*-Tyr
DSC	Differential scanning calorimetry
DNA	Deoxyribonucleic acid
DTT	1,4-Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
EPR	Electron paramagnetic resonance
FAD	Flavin adenine dinucleotide
FBDD	Fragment-based drug discovery
FBS	Fragment-based screening
FDR	Flavodoxin reductase
Fdx	Ferredoxin
FMN	Flavin mononucleotide
Fpr	Flavoprotein reductase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus

HS	High-spin
HTS	High-throughput screening
IPTG	Isopropyl-*beta*-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
L	Litre
LAM	Lipoarabinomannan
LB	Luria-Bertani
LS	Low-spin
M	Molar
MCD	Magnetic circular dichroism
MDR-TB	Multi-drug resistant tuberculosis
MIC	Minimum inhibitory concentration
Mtb	*Mycobacterium tuberculosis*
MW	Molecular weight
MWCO	Molecular weight cut off
NanoESI-MS	Nano-electrospray ionization mass spectrometry
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
Ni-IDA	Nickel iminodiacetic acid
NMR	Nuclear magnetic resonance
OD	Optical density
P450 BM3	CYP102A1 from *Bacillus megaterium*
P450cam	CYP101A1 from *Pseudomonas putida*
PAS	Para-aminosalicylic acid
PCR	Polymerase chain reaction
Pd	Putidaredoxin
PdR	Putidaredoxin reductase
PDB	Protein Data Bank

PEG	Polyethylene glycol
PIM	Phosphatidylinositol mannoside
RD	Regions of difference
RIF	Rifampicin
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
rpm	Revolutions per minute
RR	Resonance Raman
Rz	Reinheitszahl
SEC-MALS	Size exclusion chromatography with multi-angle light scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMK	Sulfomenaquinone
TB	Tuberculosis
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEV	Tobacco Etch Virus
tRNA	Transfer ribonucleic acid
TSS	Transcription start site
UV	Ultraviolet
WHO	World Health Organization
Water LOGSY	Water-Ligand Observed via Gradient Spectroscopy
XDR-TB	Extensively or extremely-drug resistant tuberculosis
ΔALA	5-aminolevulinic acid hydrochloride
ε	Extinction coefficient

ABSTRACT

Tuberculosis is a disease that kills more people every year than any other infectious disease and is caused by the human pathogen, *Mycobacterium tuberculosis* (Mtb). This disease can be treated by a standard six month course of four antimicrobial drugs that have been in use since the 1960s. However, the rise of multi-drug resistant and extensively drug-resistant strains of TB has complicated the efforts to eradicate the disease. Therefore, there is a critical need for the development of new anti-TB drugs with a novel mechanism of action that can speed up treatment duration and help avoid resistance. The discovery of twenty genes encoding cytochrome P450 enzymes in the Mtb H37Rv genome sequence has pointed to the significance of these enzymes in the physiology and pathogenicity of this bacterium. Consequently, the characterisation of these Mtb P450 enzymes may define their physiological roles of which can be a novel anti-tubercular drug target.

To date, the characterisations of selected Mtb P450 enzymes have highlighted their diverse and unexpected roles in the metabolism of cholesterol and lipids and the production of secondary metabolites. Biochemical and biophysical studies of these enzymes provided knowledge of their active site properties that may be exploited for drug discovery. Therefore, with the prospect of defining novel functions and identifying novel drug targets, characterisations of the remaining orphan Mtb P450s is of interest. *M. tuberculosis* CYP141A1 and CYP143A1 are orphan enzymes with unknown physiological function in Mtb which is characterised in this study through use of various spectroscopic and biophysical techniques.

Interestingly, CYP141A1 can be expressed in form of which 54 amino acids (Del54CYP141A1) are deleted from the N-terminus. Although Del54CYP141A1 still retain spectroscopic characteristics, this form of P450 cannot be crystallized. Optimisation of full-length CYP141A1 buffer composition resulted to the formation of reproducible crystals and determination of CYP141A1 structure. Spectroscopic and structural characterisations presented in this thesis revealed many characteristics of CYP141A1 and CYP143A1 are comparable to previous Mtb P450s reported to date. CYP141A1 and CYP143A1 active site consist of *b*-type heme iron ligated by cysteine residue and a water molecule at its proximal and distal face, respectively. Both enzymes bind tightly to azole antifungal drugs highlighting their potential as a drug target. In addition, fragment-based screening applied to CYP141A1 and CYP143A1 provided the starting point for the development of potent, isoform-specific inhibitors for both orphan Mtb P450 enzymes.

The first crystal structure of CYP141A1 and identification of new fragment binders of CYP141A1 and CYP143A1 are presented in this thesis. Overall, this research remains significant in providing new knowledge on the spectroscopic and structural properties of the *M. tuberculosis* P450s CYP141A1 and CYP143A1.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. However, this thesis is written in the University of Manchester's journal format, which allows inclusion of work already published.

COPYRIGHT STATEMENT

Dedication

I dedicate this thesis to my parents, Nisbar Mohamed Nasib and Fauziah Jaafar, who offered unconditional love and support and have always been there for me.

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to God the Almighty for His showers of blessings throughout my PhD journey. As He had said in His Holy book of Quran, “For indeed, with hardship will be ease. Indeed, with hardship will be ease.” and this excerpt had kept me going from the beginning of my PhD till the final submission of this thesis. I owe it all to the God the Almighty for granting me the wisdom, health and strength to undertake my PhD and enabling me to its completion.

I would like to express my sincere gratitude to my supervisor, Professor Andrew Munro for giving me the opportunity to do my PhD in his laboratory and providing invaluable support throughout my time as his PhD student. I am grateful to have been part of such a great research group. Special thanks go to Dr. Kirsty McLean, Dr. Hazel Girvan, Dr. Kang Lan Tee and Dr. Harshwardhan Poddar for all your help and knowledge sharing. Not forgetting, to Marina Golovanova for teaching me all the valuable laboratory skills and tricks. To all of my labmates, past and present members of the Munro-Scrutton-Leys group, thank you for the memories we shared inside and outside the laboratory and I hope that we will remain friends for many more years to come.

A great deal of thanks must go to all of my Malaysian friends in Manchester especially fellow PhD candidates, my housemates and my netball team members. Thank you for the newfound friendship that we made and all the help and advice on many different occasions. Special mentions go to Nazahiyah, Nazlia, Ku Edhura, Illida, and Rabizah for being there in good and bad times and keeping me sane during the tough times. I would also like to thank all of my friends that I made in the past and present for wishing the best for my PhD.

Last but not least, I would like to thank my family: my parents and to my brothers and sister for their continuous support, encouragement and prayers that sustained me thus far. Special thanks go to my cousins, Aisyah and Nadhirah Johari; and my aunty, Suzaira Bakar. It is amazing to have a family close by when you are so far away from home. Finally, to Jasmine Aisya, Farel Idlan and Johan Arif, thank you for being my source of motivation for me to seek as much knowledge as possible and never to give up. I pray that you will receive the best education that you could get and inspired to be the best that you can be in the future. May God bless all of you!

PREFACE TO THE JOURNAL FORMAT THESIS

This thesis has been presented in University of Manchester's Journal Format thesis which allows the candidate to write sections of their thesis in a format suitable for publication. This gives an overall structure of the thesis to consist of an abstract, introduction, four results chapters and conclusion chapter. Each results chapters is presented as a publication with an abstract, introduction, materials and methods, results, discussion and references. For consistency, text formatting, figure/table numbering and page numbering have been altered to incorporate the whole thesis. Materials included in the journal format thesis may include those which are solely and/or partly authored by the student and may be already published, accepted for publication, submitted for publication in externally refereed contexts such as journals and conference proceedings or prepared in a format which is suitable for publication. For clarification, the individual contributions of the author and co-authors and other collaborators has been detailed below.

Contributions from Authors

As a supervisor, Andrew W. Munro contributed to data analysis and manuscript preparation for all result chapters.

Chapter Two

Nur Dayana Nisbar performed bioinformatics analysis, produced the new truncated construct of CYP141A1; cloned, expressed, and purified the truncated CYP141A1 protein; performed all spectroscopic and crystallisation experiments; performed data analysis and prepared the manuscript.

Hazel M. Girvan aided in the design and construction of truncated CYP141A1.

Teresa Cortes and Douglas Young performed transcriptomic analysis of the CYP141A1 gene.

Chapter Three

Nur Dayana Nisbar expressed and purified the protein; prepared the materials needed for SEC-MALS experiments; performed crystallization experiments; performed data analysis and prepared the manuscript.

Colin W. Levy collected X-ray diffraction data and aided in solving the crystal structure of CYP141A1.

Kirsty J. Mclean aided in expression, purification and crystallization of CYP141A1.

Martin Read performed protein identification experiments.

Hilda Diana Ruiz performed SEC-MALS experiments.

Chapter Four

Nur Dayana Nisbar expressed and purified the protein; performed UV-visible spectroscopy experiments; prepared the protein for fragment-based screening and SEC-MALS experiments; performed X-ray crystallography experiments; performed data analysis and prepared the manuscript.

Le Van Duyet prepared the protein for MCD, CD and DSC experiments and collected CD and DSC spectra.

Madeline E. Kavanagh performed all mass spectrometry and fragment-based screening experiments and participated in fragment-based screening data analysis.

Kirsty J. McLean and Hazel M. Girvan performed EPR experiments, aided in spectroscopic and crystallization experiments and in data analysis.

Colin W. Levy collected X-ray diffraction data and aided in solving the crystal structure of CYP141A1.

Myles Cheesman and Justin Bradley performed MCD experiments.

Hilda Diana Ruiz performed SEC-MALS experiments.

Chapter Five

Nur Dayana Nisbar expressed and purified the protein; performed UV-visible spectroscopy experiments; prepared the protein for fragment-based screening and SEC-MALS experiments; performed X-ray crystallography experiments; performed data analysis and prepared the manuscript.

Shalini Swami aided in expression, purification and crystallization of CYP143A1; and prepared protein for resonance Raman spectroscopy.

Madeline E. Kavanagh performed all mass spectrometry and fragment-based screening experiments and in fragment-based screening data analysis.

Kirsty J. McLean and Hazel M. Girvan performed EPR experiments, aided in spectroscopic and crystallization experiments and in data analysis.

Colin W. Levy collected X-ray diffraction data and aided in solving the crystal structure of CYP143A1.

Jim Kincaid and Yilin Liu performed resonance Raman spectroscopy.

Anthony Coyne performed NMR fragment screening experiments to identify novel small molecule scaffolds for CYP143A1.

CHAPTER 1

INTRODUCTION

1.1 Tuberculosis

Tuberculosis (TB) has scourged humankind worldwide throughout known history and human prehistory. Estimation of the time of origin of mycobacteria suggest that *Mycobacterium tuberculosis*, the causative agent for tuberculosis disease, was present in East Africa as early as 3 million years ago and it may have infected early hominids at that time. (1) Archaeological evidence of tuberculosis was found in fragments of vertebrae of Egyptian mummies indicating spinal TB, and were clearly depicted in early Egyptian art more than 5000 years ago. (2) Traces of tuberculosis were also found in America verified by Peruvian mummies bone findings with characteristic Pott's deformities. (3) Furthermore, the description of tuberculosis was found in Greek literature around the time of Hippocrates, the most famous Greek physician, where it was called phthisis or consumption, (4) and was continuously described in literature throughout the Renaissance era. It was René Théophile Hyacinthe Laennec's book titled *D'Auscultation Mediate* that provides modern understanding of tuberculosis. He identified most of the physical signs of pulmonary disease and introduced terms to describe those findings that are still in use today. (5) However, the most significant event in tuberculosis history was the famous presentation of Hermann Heinrich Robert Koch titled *Die Aetiologie der Tuberculose*, to the Berlin Physiological Society on March 24, 1882. Robert Koch produced indisputable evidence that specific microbe *Mycobacterium tuberculosis*, which he called tubercle bacilli, was the fundamental cause of tuberculosis. (6) Koch's multiple contributions to bacteriology, and especially his discoveries in tuberculosis, led to him being awarded the Nobel Prize in Medicine or Physiology in 1905. Another important contribution to the diagnosis of TB was the discovery of X-rays by Wilhelm Conrad von Röntgen, through which the development, course and severity of TB could be accurately monitored and studied. (7) Both of these outstanding scientific achievements were significant in the effort to slowly eradicate tuberculosis in the world.

Tuberculosis (TB) remains one of the top 10 causes of death worldwide. In 2015, the largest number of new TB cases occurred in Asia, with 61% of new cases, followed by Africa with 26% of new cases. Six countries accounted for 60% of the global new cases with India leading the count, followed by Indonesia, China, Nigeria, Pakistan and South Africa. There were an estimated 10.4 million new TB cases worldwide in 2015, of which 5.9 million (56%) were among men, 3.5 million (34%) among women and 1.0 million (10%) among children. Moreover, an estimated 1.4 million TB deaths were recorded. Globally, the rate of decline in TB incidence remained at 1.5% from 2014 to 2015 and the number of TB deaths fell by 22 % between 2000 and 2015. As people living with HIV are 20 to 30 times more likely to develop active TB disease than are people without HIV, at least one-third of people living with HIV worldwide in 2015 were infected with TB bacteria. People living with HIV accounted for 1.2 million (11%) of all new TB cases and 0.4 million deaths resulted from TB disease among people living with HIV. In addition, globally in 2015 there were an estimated 480,000 new cases of multidrug resistant (MDR) TB and an additional 100,000 people with rifampicin-resistant TB.

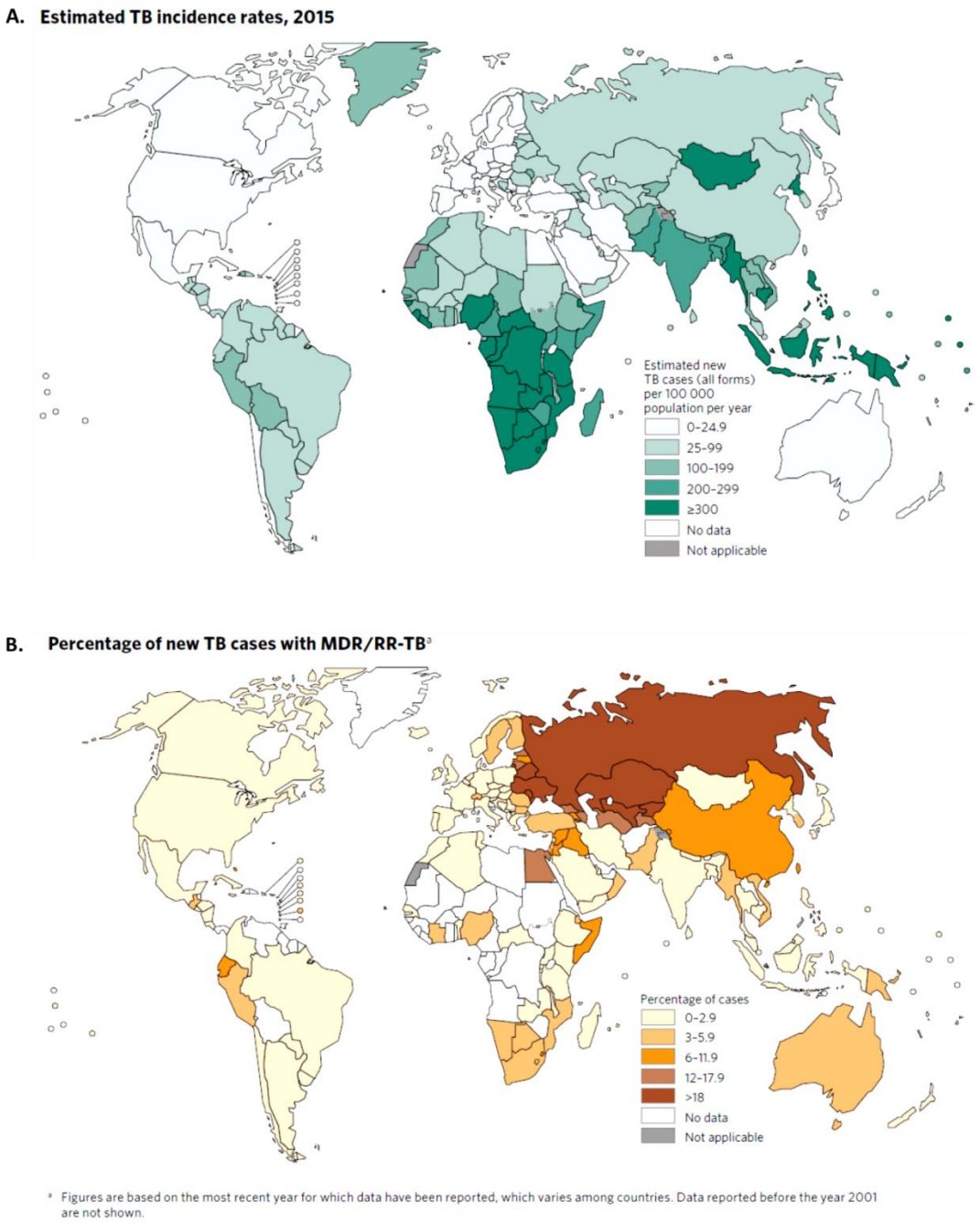


Figure 1.1 Global TB infections.
(A) Estimated TB incidences rates in 2015. The annual number of incident TB cases relative to population size varied widely among countries in 2015, from under 10 per 100 000 population in most high-income countries to 150–300 in most of the 30 high TB burden countries. (B) Percentage of new TB cases with MDR/RR-TB at the country level. Globally in 2015, an estimated 3.9% (95% confidence interval [CI]: 2.7–5.1%) of new cases had MDR/RR-TB. Figure adapted from WHO Global tuberculosis report 2016. (8)

In the recent World Health Assembly and at the United Nations General Assembly, participating governments have agreed on targets to end the TB epidemic by 2030 within the context of the End TB Strategy and Sustainable Development Goals (SDGs), respectively. In

the End TB Strategy, the aim includes a 90% reduction in TB deaths and an 80% reduction in TB incidence by 2030, compared with 2015. However, the first milestones of the End TB Strategy, set for 2020, are a 35% reduction in the absolute number of TB deaths and a 20% reduction in the TB incidence rate compared with levels in 2015. To reach these milestones, the TB incidence rate needs to fall by 4-5% per year and the proportion of people with TB who die from TB needs to be reduced to 10% globally by 2020. The rate of progress in reducing the burden of TB disease in the top six countries will also have a major influence on whether or not the 2020 global milestones are achieved. However, despite the steady slow decline in TB incidence, prevalence and mortality observed over the last decade, eradication of the disease at global level is still a challenge and massive resource investment is still required.

Tuberculosis is an infectious disease, caused by the pathogen *M. tuberculosis*, which typically affects the lungs (pulmonary TB) and also affects other part of the body including the glands, bones and nervous system (extra pulmonary TB). The disease is spread when people who are sick with pulmonary TB expel the bacteria into the air by coughing, sneezing or spitting. Only a few droplets of these bacteria need to be inhaled by a person to get infected with the disease. In most healthy people, the immune system will kill the bacteria and no symptoms are observed. However, sometimes the immune system cannot kill the bacteria and the symptoms are still not observed, but the bacteria will remain in the body. This is known as latent TB and which about one third of the world's population has latent TB. If the immune system cannot kill the bacteria, it can spread within the lungs or other parts of the body becoming an active form of TB and symptoms will develop within a few weeks or months. When a person develops active TB disease, the symptoms will start with cough, lack of appetite, unexplained weight loss, night sweats, fever and fatigue that may be mild for many months. The development of active TB is usually slow and symptoms might not begin until months or even years after the first infection. This can lead to delays in seeking care and results in transmission of the bacteria to others.

Several tests can be used to diagnose TB. The most used and established TB diagnostic test is the sputum smear microscopy. Sputum samples are examined under a

microscope to see the presence of the acid-fast TB bacteria. Acid-fast microscopy is easy and quick but it does not confirm a diagnosis of TB because some acid-fast bacilli are not *M. tuberculosis.* Therefore, an additional culture method needs to be performed to confirm the diagnosis. In addition, microscopy detects only half the number of TB cases and cannot detect drug-resistance. A more accurate and rapid diagnostic test for TB is the rapid molecular test Xpert MTB/RIF®, which simultaneously detects TB and resistance to rifampicin within 2 hours. This rapid test was introduced in 2010 and is now recommended by WHO as the initial diagnostic test in all persons with sign and symptoms of TB. Globally, the use of rapid molecular tests is increasing and many countries are phasing out the use of smear microscopy for diagnostic purposes.

1.2 Tuberculosis treatment

The understanding of how to treat tuberculosis disease has evolved considerably throughout time. Before the middle of the 19th century, religion and superstition played significant roles in efforts to treat this disease, which included taking strange potions, consuming questionable remedies and depending on mystical powers. A more sensible approach was introduced in 1854 by Hermann Brehmer with the concept of fresh air, rest, good nutrition and isolation in sanatoria. Many such sanatoria were established throughout the late 19th century. The purpose of sanatorium treatment was to support the patient's immune response and to isolate infected persons from the public. As a result, there was a slow decrease in TB cases in the 1900s. Although patients recovered well during sanatorium treatments, the long-term results were not good as over 60% of patients discharged from sanatoria died of their disease within 6 years. (9) The breakthrough in tuberculosis treatment came with the discovery of the causative bacterium, *Mycobacterium tuberculosis*, in 1882 by Robert Koch. This discovery directed the effort to treat tuberculosis by killing the bacterium itself. (10) Antimicrobial chemotherapy was the term used to describe drug treatment aimed at reducing the population of a particular microbe, in this case *M. tuberculosis* (Mtb), in the person's body. In 1944, Schatz, Bugie and Waksman reported their discovery of streptomycin, an antibiotic isolated from *Streptomyces griseus* that was proven to have

activity against Mtb. (11) Another new medication, para-aminosalicylic acid (PAS) was developed by Jörgen Lehmann in 1948 and used alongside streptomycin. In 1952, isoniazid was discovered to have anti-tubercular activity by three independent laboratories. (12) The use of isoniazid, PAS and streptomycin reduced drug resistance and improved effectiveness of TB treatment. These discoveries were the beginning of a TB-specific treatment era, as several new anti-TB drugs were developed subsequently, including pyrazinamide (1954), cycloserine (1955), ethambutol (1961) and rifampicin (1963). Consequently, a combination therapy regime was introduced that, at that time, had a treatment duration of 18 months or more. In addition, aminoglycosides such as capreomycin, vicomycin, kanamycin, amikacin and quinolones were used in drug resistance situations. Table 1.1 shows a summary of commonly used TB drugs and their mechanisms of action.

Table 1.1 Main tuberculosis drugs used in current treatment regimen and their targets and mode of action.
Table adapted from Zumla et al. (13).

Drug (year of discovery)	Targets	Mechanisms of action	References
First-line drugs			
Isoniazid (1952)	Multiple targets including acyl carrier protein reductase (InhA)	Inhibition of cell wall mycolic acid synthesis and other multiple effects on DNA, lipids, carbohydrates and NAD metabolism	(14)
Pyrazinamide (1954)	S1 component of 30S ribosomal subunit	Inhibits translation and trans-translation, acidifies cytoplasm	(15)
Ethambutol (1961)	Arabinosyl transferase	Inhibition of cell wall arabinogalactan synthesis	(16)
Rifampicin (1963)	RNA polymerase β subunit	Inhibition of RNA synthesis	(17)
Second-line drugs			
Streptomycin (1944)	S12 protein and 16S rRNA components of 30S ribosomal subunit	Inhibition of protein synthesis	(18)
PAS (1948)	Dihydropteroate synthase	Inhibition of folic acid and iron metabolism	(19)
Cycloserine (1955)	D-alanine racemase and ligase	Inhibition of peptidoglycan synthesis	(20)
Kanamycin (1957)	30S ribosomal subunit	Inhibition of protein synthesis	(21)
Ethionamide (1961)	Acyl carrier protein reductase (InhA)	Inhibition of mycolic acid synthesis	(14)
Capreomycin (1963)	Interbridge B2a between 30S and 50S ribosomal subunit	Inhibition of protein synthesis	(22)
Amikacin (1972)	30S ribosomal subunit	Inhibition of protein synthesis	(22)

The standard recommended treatment for new patients with presumed drug susceptible pulmonary TB requires treatment duration of a minimum of 6 months in two phases: two months intensive phase of the four first-line drugs (isoniazid, rifampicin, pyrazinamide and ethambutol) followed by the continuation treatment phase of isoniazid plus rifampicin. Currently, the cure rates of this regimen are more than 95% when administered under directly observed therapy. However, there are significant challenges associated with this standard treatment which can result in TB treatment failure. Patient adherence to the treatment is a major challenge given the lengthy treatment duration, the amount of pills needed to be consumed each day and the high cost of treatment which is excessive for many patients. In addition, some patients cannot tolerate the side effects of the medications, such as allergic reactions, gastrointestinal disorder and hepatotoxic effects, and they may decide to stop the treatment altogether. Drug intolerance, toxicities and loss of efficacy are significant in TB patients co-infected with HIV as there are known pharmacokinetic drug-drug interactions between TB drugs and antiretroviral therapy (ART) drugs, and increased risk of adverse effects. Besides that, the effectiveness of the treatment approach is also an issue as the treatment guidelines for drug-susceptible and drug-resistant tuberculosis have some variations between countries. Diversities of the treatment guidelines include the duration of standard treatment, treatment of only isoniazid-resistant TB, treatment of tuberculous meningitis and treatment monitoring. (23) Moreover, inappropriate guidelines, non-compliance with guidelines and even absence of guidelines for the medical practitioners can also hamper the effective treatment of TB. The rise of multidrug-resistant (MDR) tuberculosis also complicates the matter as treatment of drug-resistant TB should always be guided by an experienced physician and the initial treatment regimen should be personalized according to the results of drug-susceptibility testing of the *M. tuberculosis* isolate from the patient. However, this test is not available for over 95% of patients with MDR-TB and consequently the standard treatment is used, resulting in the majority of MDR-TB cases being missed or mistreated.

Since the current 6-month four-drug combination was introduced in the 1970s, no new TB drug classes have been developed or approved for drug-susceptible TB. Not until recently has a promising new era in tuberculosis drug development begun with the

emergence of a number of anti-TB drug candidates in the pipeline. The need for the development of new TB drugs and shorter treatment regimens is obvious. Over the past decade, positive advances in TB therapeutics have been reported as a result of significant investment by scientists, funding bodies and high-profile advocacy by the WHO's STOP TB department and other organizations. This accelerated rate of TB drug development is ensured to be continuous as "Intensified research and innovation" is the third pillar of the recently announced WHO End TB strategy. (8) As of August 2015, there are nine anti-TB drugs in advanced phases of clinical development for the treatment of drug-susceptible, multidrug-resistant TB or latent TB infection (LTBI), of which six are new and three are already approved or repurposed. The six new compounds are bedaquiline, delamanid, PBTZ169, pretomanid, Q203 and sutezolid. The three approved or repurposed drugs undergoing further testing are rifampicin, rifapentine and linezolid.

1.3 Mycobacterium tuberculosis

Mycobacterium is a genus within the order Actinomycetales that comprises a large number of well characterised species, several of which are associated with human and animal disease. *M. tuberculosis* (Mtb) is only known to populate and infect humans, and consequently it is the etiologic agent of tuberculosis disease in humans. Mtb is a non-motile, rod-shaped bacterium that is usually 2-4 µm in length and 0.2-0.5 µm in width. It is also an obligate aerobe that requires high levels of oxygen to survive. For this reason, Mtb is always found in the well-aerated upper lobes of the lungs. In addition, the bacterium is a facultative intracellular parasite by inhibiting phagosome-lysosome fusion and has a slow generation time of 15-20 hours. *M. tuberculosis* cannot be identified as either Gram-positive or Gram-negative bacteria because the Gram stain does not permeate its cell surface. However, it can be classified as acid-fast bacteria as it can be stained by acid-fast dyes such as the Ziehl-Neelsen stain, after which the bacilli appear pink in a contrasting background. The impermeability of Mtb to certain dyes and stains is mainly due to its unusual cell wall composition.

The cell wall core is composed of three distinct macromolecules: peptidoglycan, arabinogalactan and mycolic acid. Mycolic acids are formed by long meromycolate and short α-chains fatty acids, which form the characteristic thick, waxy lipid coat of mycobacteria and which are the major contributors to the impermeability of the cell wall and to bacterial pathogenicity. (24) Above the mycolic acids are the free lipids and interspersed within the cell wall core are the cell wall proteins, the phosphatidylinositol mannosides, the phthiocerol-containing lipids, lipomannan and lipoarabinomannan. Fractionation of the cell wall components using various solvents resulted in solubilised fraction of free lipids, proteins, lipoarabinomannan and phosphatidylinositol mannosides, and an insoluble fraction of the mycolic acid-arabinogalactan-peptidoglycan complex. It is hypothesised that soluble lipids, proteins and lipoglycans are the signalling, effector molecules in the disease process, whereas the insoluble cell wall core is essential for the viability of the cell and should be addressed in the context of drug development. (25) The biosynthetic pathways of all three components of the cell wall core reveal significant drug targets that are the basis of present drugs and which have potential for new drug development. (26)

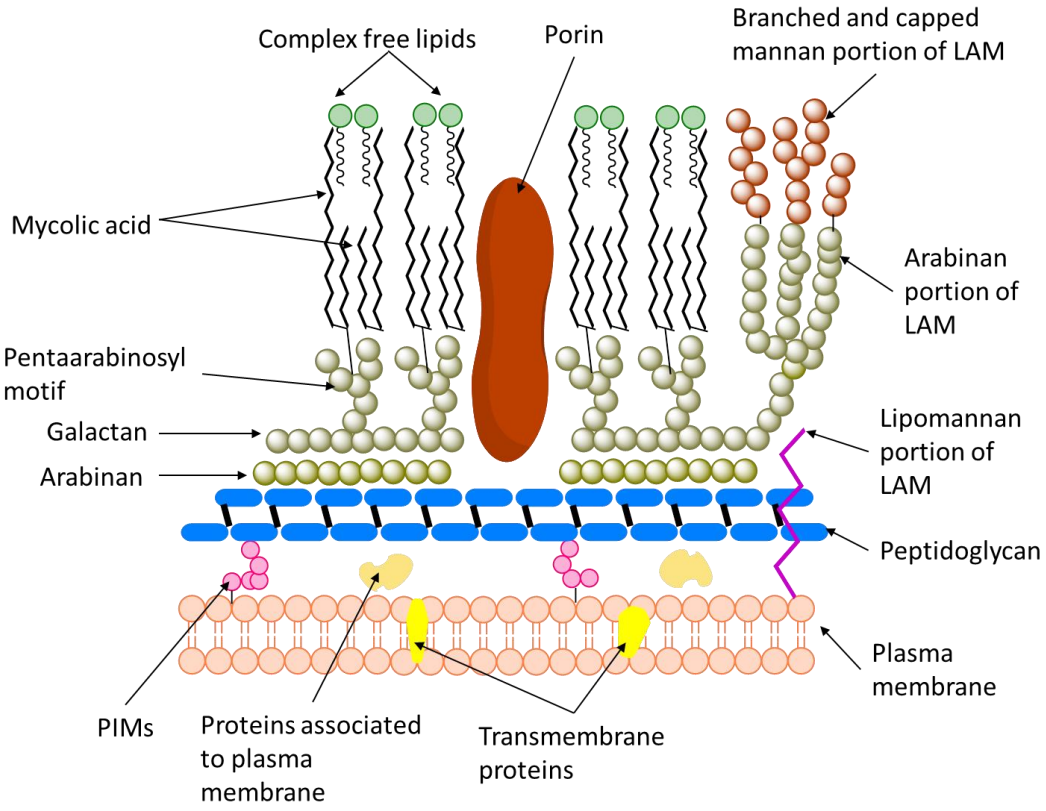


Figure 1.2 The mycobacterial cell wall.
Mycobacteria are characterised by the presence of a thick mycolate-rich outer layer providing efficient permeability barrier. LAM: lipoarabinomannan; PIMs: phosphatidylinositol mannosides. Figure adapted from Medjahed et al. (27)

1.4 The Cytochromes P450

The discovery of the cytochromes P450 begins around 1940s and 1950s through pioneering studies of Mueller and Miller and of Julius Axelrod, who reported the presence of oxidative enzymes involved in the metabolism of xenobiotics and located in the membranous fractions of liver cells. (28) In addition, the paper by Ken Ryan and Lewis Engel revealed the role of P450 in steroid metabolism and this led to subsequent studies that discovered the functions for P450s. The name P450 was derived due to the discovery of a pigment (P) in liver microsomal fractions with an absorption maximum at 450 nm in the presence of a reducing agent and carbon monoxide (CO) by Martin Klingenberg in 1958. (29) Following this finding, in 1962 Tsuneo Omura and Ryo Sato confirmed P450s as hemoproteins and the name cytochrome was associated with it. (30)

Moreover, the development of spectrophotometry enabled the characterization of the cytochrome P450 UV-visible absorption spectrum, and most importantly the identification of the specific absorbance maximum of the ferrous/CO-bound form. In their resting state, the cytochromes P450 binds heme *b* in its ferric state and they occupy either a low-spin (LS) or a high-spin (HS) state, depending on the organization of electrons in the 3d orbitals of the iron. The LS form is typical for substrate-free cytochrome P450 enzymes and is usually associated with the heme iron being hexa-coordinate. Four equatorial ligands to the heme iron originate from nitrogen atoms in the heme pyrrole rings. The fifth (proximal) ligand comes from the side chain of a phylogenetically conserved cysteine thiolate and the sixth (distal) ligand originates from the oxygen of a water molecule. The characteristic absorption maximum at ~450 nm in ferrous cytochromes P450 bound to CO is a consequence of the nature of the proximal axial ligand to the heme iron (the cysteine thiolate). The electron donating property of the cysteine thiolate bond is important for the P450s' catalytic activity in reducing and activating heme iron-bound dioxygen.

1.4.1 Cytochrome P450 catalytic cycle

As shown in Figure 1.3, the cytochrome P450 catalytic cycle begins at the resting state with a ferric heme iron in the LS state ligated to a cysteine thiolate in the proximal position, and to a water molecule in the distal position. Upon substrate binding, the water molecule is displaced and a HS ferric heme iron is formed. This results in an increase in the ferric heme iron reduction potential that favours electron transfer to the heme iron from the redox partner. The first electron reduces the ferric heme iron to the ferrous state, which can then bind dioxygen resulting to the formation of a ferric superoxy species which has a more favourable (higher) redox potential. A second electron is then transferred from the redox partner, leading to the production of a ferric peroxy complex, which is rapidly protonated to the ferric hydroperoxy state (compound 0). Further protonation at the distal oxygen atom caused heterolysis of the O-O bond and the production of a ferryl-oxo form (compound 1) species and a molecule of water. Compound 1 is a reactive and highly transient species considered to effect transfer of its oxygen atom to the substrate, forming a hydroxylated

product (though other oxidised products may also be formed). Dissociation of the oxidised product results in the re-binding of a water molecule as the sixth ligand to the restored ferric heme iron, completing the cycle. However, there are three well known non-productive pathways as follows: (i) the autoxidation shunt: the ferric superoxy species releases superoxide to return to the ferric state, (ii) the peroxide shunt: compound 0 is protonated and collapses to the ferric state with hydrogen peroxide formation, and (iii) the oxidase shunt: the compound 1 oxygen is reduced and protonated, and the ferric state is restored along with production of a molecule of water. These non-productive pathways may occur in absence of substrates or due to inappropriately positioned substrates in the active site, or due to failure in timely delivery of electrons or protons. As cytochromes P450 are monooxygenases, the classical reaction catalysed is the hydroxylation of organic substrates on carbon atoms. However, P450s are versatile in their reactivity and perform several other forms of chemical transformations, as shown in Figure 1.4. Studies of microbial P450s have revealed diverse roles such as the hydroxylation of the immediate precursor to the antibiotic erythromycin (6-deoxyerythronolide B) and the epoxidation of the anticancer agents epothilones C and D (forming epothilones A and B, respectively). (31,32)

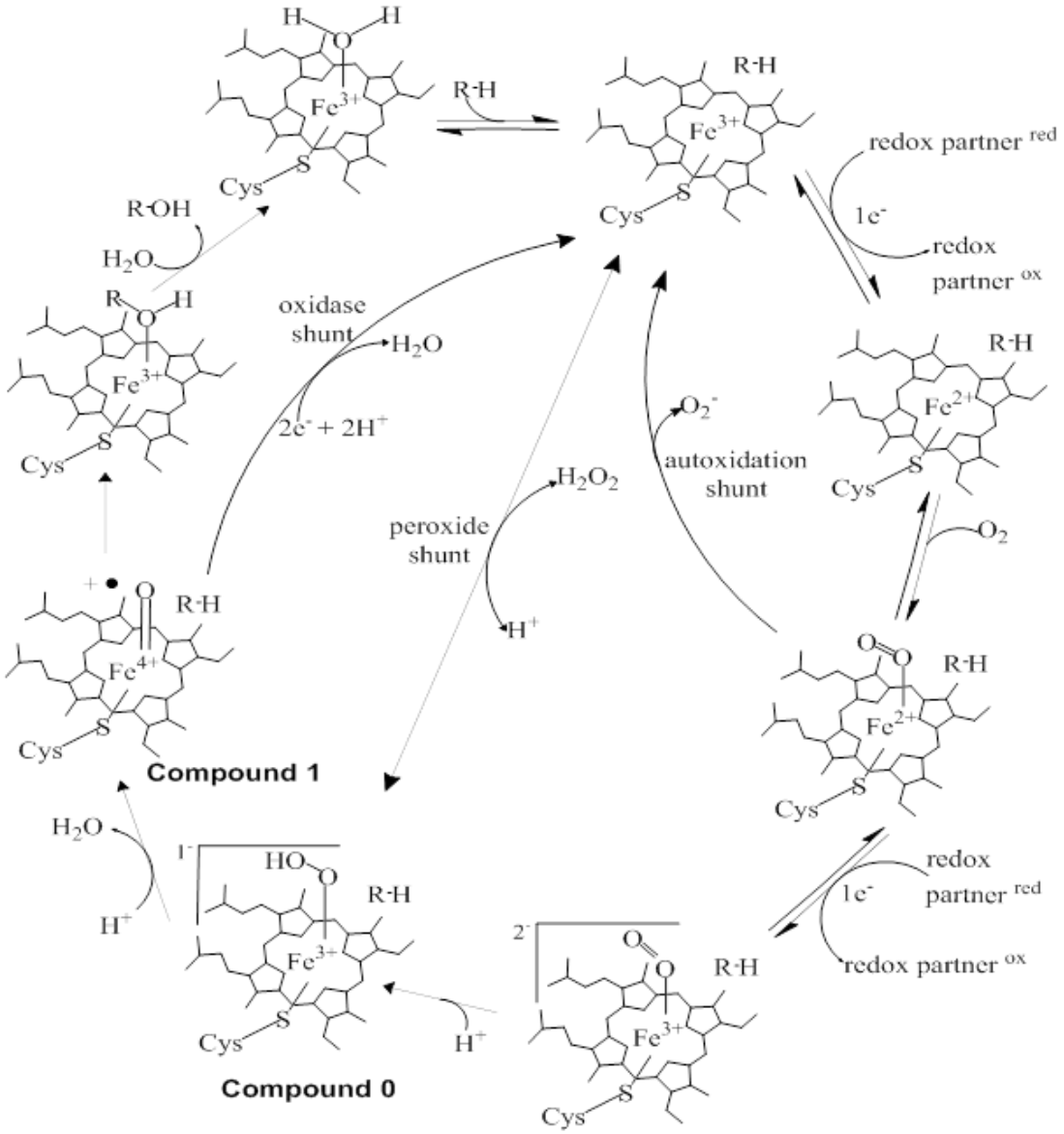


Figure 1.3 The catalytic cycle of cytochrome P450.
Upon substrate binding, the distal water ligand is displaced and causes the ferric heme iron spin-state to shift from low-spin (S = ½) to high-spin (S = 5/2). This leads to a more positive potential of the heme iron and favours electron transfer from a redox partner to reduce the heme iron to the ferrous state. Oxygen binds to the ferrous heme iron to form the ferrous-oxy intermediate. A second electron is delivered by the redox partner and reduces the heme iron to the ferric peroxy state. Two successive protonations produce the ferric hydroperoxo intermediate (Compound 0) and leads to scission of the bound dioxygen and a production of a water molecule. After a further protonation, the heme is transformed to a ferryl-oxo species (Compound 1) with a porphyrin π radical cation which is likely to be the catalytically relevant substrate oxidant in most P450 reactions. Compound 1 attacks the nearby substrate and effects its hydroxylation. Hydroxylated product (R-OH) departs, allowing water to rebind the ferric heme iron and complete the cycle. Non-productive pathways leading to collapse of oxy intermediates in the cycle are also indicated.

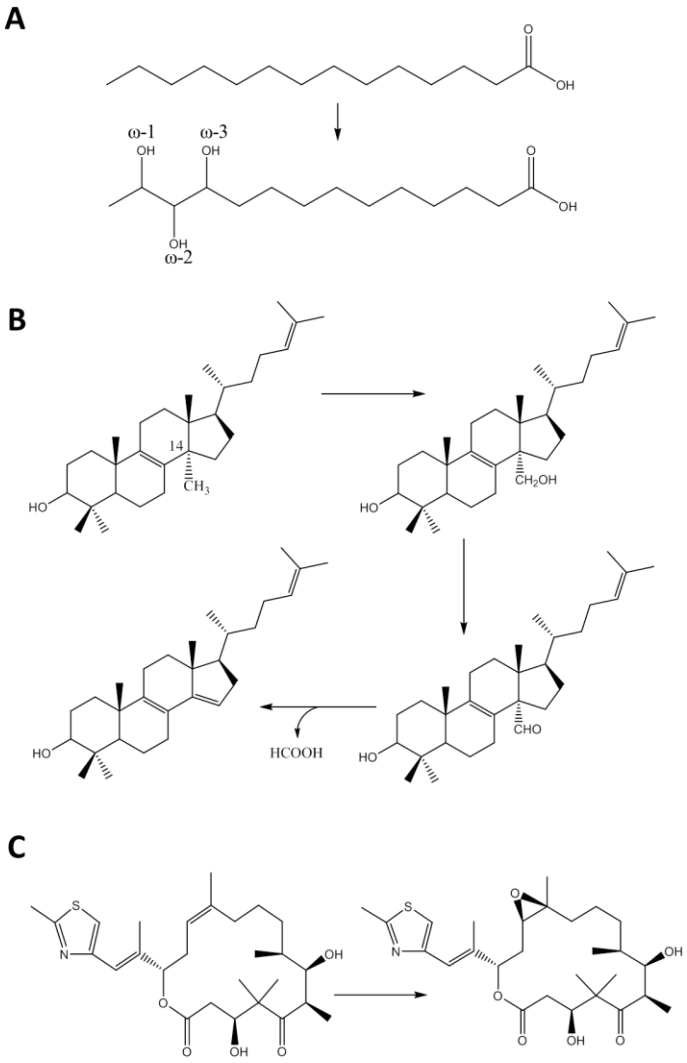


Figure 1.4 Different cytochrome P450 reactions catalysed by different microbial P450s.
(A) C-H bond hydroxylation: the hydroxylation of tetradecanoic acid at the ω-1, ω-2 or ω-3 position (indicated) by *Bacillus megaterium* CYP102A1 (P450 BM3). (33) (B) Demethylation: the 3-step oxidation of the sterol lanosterol catalysed by the fungal CYP51 sterol demethylase, forming the product, the 14α-demethylated ergosterol. (34) (C) Epoxidation: the epoxidation of epothilone D to form epothilone B by EpoK, a cytochrome P450 from the bacterium *Sorangium cellulosum.* (32)

1.4.2 Electron transfer partners of cytochrome P450

As detailed above, the typical monooxygenation function of cytochrome P450 requires delivery of two electrons and two protons to the heme iron, according to the equation:

$$RH + O_2 + 2e^- + 2H^+ \rightarrow ROH + H_2O$$

The two electrons are derived from NADPH or NADH cofactors (according to the preference of the specific redox partners used) and transported to the P450 heme via one or more redox partner protein systems. Early studies of cytochrome P450 identified two distinct classes of redox partner systems, which were initially considered the only major classes of P450 redox partners. The first major class, referred to as class I, is a bacterial-type system which is typified by the *Pseudomonas putida* CYP101A1 (P450cam, a camphor hydroxylase) redox partners. Detailed analysis of the P450cam redox partners demonstrated that these were a FAD-containing reductase (putidaredoxin reductase, PdR) that transferred electrons from NADH to a 2Fe-2S ferredoxin (putidaredoxin, Pd), which then mediated single electron transfers to P450cam. (35–37) This three-component cytochrome P450 redox system is typical of that found for many other prokaryotic cytochromes P450, although the nature of the iron-sulfur cluster is variable (typically either a 2Fe-2S or a 3Fe-4S cluster). The *Mycobacterium smegmatis* FdxA protein was shown to be a 7Fe ferredoxin and was suggested to support activity of one or more of the numerous P450s in this bacterium and in the pathogen *M. tuberculosis.* (38) The second major class, class II, involves a FAD- and FMN-containing enzyme termed cytochrome P450 reductase (CPR) found in the mammalian liver and in other tissues in eukaryotes. CPR participates in several P450-dependent reactions in other mammalian organs and in various eukaryotic cells. The CPR FAD-binding domain is the entry point for electrons derived from NADPH, with the FMN-domain receiving single electrons from the FAD and passing these to the P450. (39) This eukaryotic CPR/P450 two-component system was shown to be membrane anchored for both these enzymes as a consequence of hydrophobic N-terminal transmembrane peptide sections that embed these proteins into the microsomal membranes. However, deviation from these classical redox partner systems was demonstrated by eukaryotic mitochondrial P450 systems that used a class I redox system (40,41) and by the cineole degrading prokaryotic P450cin from *Citrobacter braakii* (CYP176A1) that uses a FMN-binding flavodoxin as its natural redox partner, which is in turn reduced by a FAD-dependent reductase. (42) Moreover, studies in recent years have revealed a diversity of redox partner systems that corresponds to the diversity of P450 reactions. (43–45)

1.4.3 The structure of cytochrome P450

Commonality between all cytochromes P450 can be seen through their structural arrangement. The first cytochrome P450 structure (P450cam, CYP101A1) was solved in the early 1980s, (46–48) followed by the second P450 structure for the heme (P450) domain of the soluble P450-CPR fusion enzyme P450 BM3 (CYP102A1). (49) Structural features general to the P450s are exemplified by both the P450cam and P450 BM3 atomic structures. The crystal structure of P450cam is shown in Figure 1.5. The overall fold of the cytochromes P450 resemble a trigonal prism with two major structural domains. The heme cofactor buried at the centre of the molecule is conserved across the P450 superfamily. The two major domains are the larger (α helix-rich) domain and the smaller (β sheet-rich) domain. In general, the position of structural elements close to the heme, especially helices I and L, are well-conserved while those regions controlling substrate specificity deviate the most, including the B' helix. The most conserved element of the P450 structure is the β-bulge segment prior to the L helix where the Cys proximal ligand resides. The rigid architecture of this segment protects the Cys ligand and enables it to accept hydrogen bonds from peptide NH groups. These hydrogen bonds aid in regulating the heme iron redox potential and therefore the cysteine loop region is important in facilitating oxygenase catalysis in the P450s. A similar structural arrangement is also found in two closely related enzymes; nitric oxide synthase (NOS) and chloroperoxidase (CPO) that, like P450s, catalyse monooxygenation reactions. (50,51) There are several conserved residues important for P450 catalysis, such as threonine (Thr252 in P450cam) that is involved in distortion of the I helix, and a phenylalanine near the conserved cysteine (Phe393 in P450 BM3, 7 residues before the cysteine) that is an important regulator of heme iron redox potential. In P450cam, the Thr252 side-chain OH donates a H-bond to a peptide carbonyl oxygen that would normally be involved in an α-helical H-bond, resulting in local helical distortion. However, this Thr is not totally conserved in P450s, as the erythromycin hydroxylase P450eryF contains an Ala and P450cin has Asn instead of a Thr, which also displays a similar distortion in the I helix as seen in P450cam. (31,52) This arrangement is considered to be crucial for proton delivery to heme iron-oxy species (ferrix-superoxy and

ferric-hydroperoxy) to generate the active Fe^{IV}=O (compound I) substrate oxidising species. Studies of BM3 F393A/H variants revealed the role of this conserved Phe residue in tuning the reduction potential of the heme iron and in its reactivity with dioxygen. (53) CYP74 (allene oxide synthase) and CYP8 (prostacyclin synthase) enzymes do not have this Phe residue and these P450s do not perform oxygenation chemistry, instead catalysing molecular rearrangements. (54) It is thought that glutamate and arginine (Glu320 and Arg323 in P450 BM3) in an EXXR motif in the K helix region are important in heme binding and maintenance of P450 tertiary structure. However, studies of CYP157C1 revealed an absence of this arginine residue in the sequence, but still bind heme and form a normal P450 ferrous-CO complex. (55) To date, advancements in the field of structural biology have enabled the determination of many P450 structures (by X-ray crystallography) including various substrate/ligand-bound complex structures, P450s in various conformational states and a few new P450-redox protein complexes. This abundance of new structural information is useful in providing a better understanding of P450 versatility and how the P450 active site adapts to substrates of diverse sizes, shapes and chemical character.

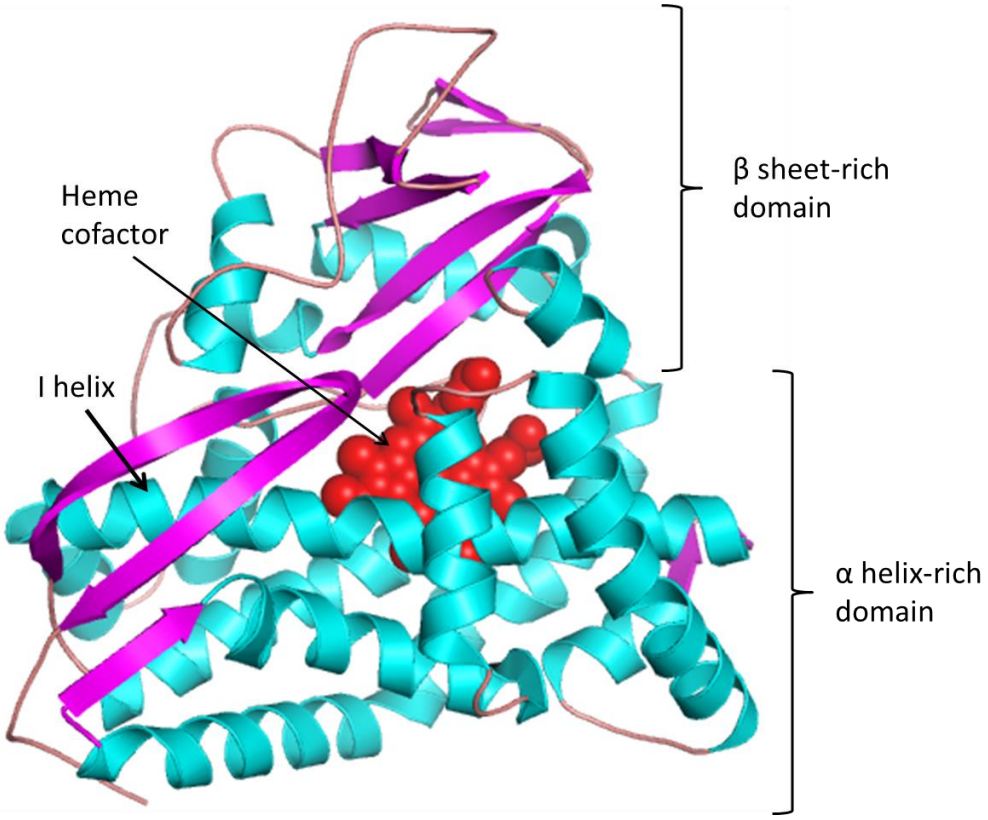


Figure 1.5 General structure of cytochrome P450.
The typical fold for a P450 enzyme is shown using *P. putida* CYP101A1 (P450cam) from its crystal structure (PDB ID: 2CPP). (48) The P450s adopt an overall structural fold resembling a trigonal prism, with the heme cofactor (red sphere) buried at the centre of the molecule. There are two major domains of a P450: the larger α helix-rich domain and the smaller β sheet-rich domain. Major structural elements include the long I helix, which runs across the distal face of the heme and contains several residues important for catalysis. The positions of structural elements close to the heme are generally well conserved in the P450s, while there is considerable deviation in the structural organisation of elements that control substrate specificity.

1.5 Mycobacterium tuberculosis P450s

In 1998, determination of the genome sequence of *M. tuberculosis* (Mtb) H37Rv by Stewart Cole and co-workers provided valuable information regarding the biology of this human pathogen and that is critical for the development of TB research. The genome sequence revealed many novel gene products of unknown function and some other unusual features. These include a large proportion of genes involved in lipid metabolism (with some previously only found in animals and plants), novel biosynthesis pathways for complex coat lipids and carbohydrates, novel gene families and several other genes encoding proteins or enzymes that were previously thought to be exclusive to eukaryotes. (56) Among the most

unexpected findings from Mtb H37Rv genome sequence was the presence of 20 different genes encoding different cytochrome P450 isoforms. Although genomic analysis of related actinobacteria has subsequently shown that many of these organisms are relatively P450-rich, (57,58) the relatively small size of the Mtb H37Rv genome (4.4 Mb) means that this is an unexpectedly large number of P450 genes in a bacterium. The cohort represents a 200-fold greater P450 gene "density" than is seen in humans (57 P450s in a 3 Gb genome) and there are no P450s in *Escherichia coli*, which has a similarly sized genome to Mtb. (59) The large number of P450s in Mtb indicates their potential importance in Mtb physiology or pathogenicity, and therefore it is of interest to decipher their physiological roles and catalytic functions.

A typical method to assign P450 function is through amino acid sequence comparisons. P450s with greater than 40% identity are classified in the same family and often corresponds to commonality in substrate selectivity. However, all the Mtb P450s have rather low levels of amino sequence identity between them and were the first representatives of new P450 families and thus, of unpredictable substrate preference. The exception was CYP51B1 which shows 30-40% identity to many eukaryotic 14α-sterol demethylases that participate in sterol synthesis. (60) CYP132 also has sufficient protein sequence similarity to a characterized CYP4 fatty acid oxidases family that enable an educated guess at its function. (61) Moreover, dispersity of the 20 Mtb P450 genes across the genome and chromosomal localization provides little clue to their biochemical functions, since most do not lie in operons or gene clusters of defined function. The only function prediction that comes from gene context is the prediction of CYP128A1 as a menaquinone hydroxylase. This is based on its gene location in an operon with the sulfotransferase *stf3*, which produces a derivative of MK-9(H_2) sulphated on its polyisoprene tail (S881).(62) Despite an apparent lack of clues to Mtb P450 physiological roles from similarities to P450s of known function and from genetic context, important insights into their essentiality to the growth, viability and virulence of the pathogen can be obtained from genetic and transcriptomic analysis. A number of in vivo and in vitro transcriptomic and microarray experiments have revealed genes that display changes in their expression levels upon exposure to conditions associated with bacterial growth, dormancy or infection. (63–65)

Similar studies have been done to probe effects of various antibiotics and other drugs known to influence Mtb growth or viability. (66–68) Three Mtb P450s (CYP121A1, CYP125A1 and CYP128A1) were experimentally demonstrated to play essential roles in Mtb by different methods. Nonetheless, these genes described as being essential to growth *in vitro* with a defined growth medium may not represent all the genes essential to the bacterium during all of its stages of growth and infection. For example, CYP139A1 appears to be nonessential for Mtb viability under standard laboratory conditions but its genomic localization near polyketide synthase (PKS) genes suggest that CYP139A1 is potentially important and 'essential' in the production of an Mtb macrolide compound. (69,70) Besides that, there is a high proportion of Mtb P450s that is orphan enzymes with no known functions that requires functional characterization which will define their physiological and infectivity-related functions and in determining which are crucial to the viability of the pathogen. To date, a growing number of Mtb P450s have been studied that provided fundamental structural/function information including the nature of heme binding site and P450 interactions with azole drugs. These studies have also revealed novel and unexpected functions for bacterial P450s and have pointed to their potential for exploitation as therapeutic targets. Major recent findings of Mtb P450 are detailed in the following sections, highlighting significant and important discoveries obtained from their biochemical and structural characterisations.

1.5.1 CYP51B1: the first prokaryotic sterol demethylase

The presence of CYP51 enzyme encoded by *Rv0764c* gene in the *M. tuberculosis* H3Rv genome was unexpected as it is the first prokaryotic sterol demethylase identified in a bacterial genome. The discovery of a prokaryotic CYP51 indicates that *CYP51* genes are not exclusive only to eukaryotes. More recently, related actinomycetes such as *Streptomyces* spp. *and Rhodococcus* spp. were also shown to encode a CYP51 enzyme. (71) Like all known bacterial P450s, the soluble nature of CYP51B1 facilitates their expression, purification and crystallization making it the first Mtb P450 to be characterised. The *E.coli* expression and isolation of Mtb CYP51B1 led to the confirmation of sterol

demethylase activity for the sterol substrates lanosterol and dihydrolanosterol and plant sterol obtusifoliol producing the respective 8,14-diene products. (72) Since a number of other components of a likely sterol biosynthetic pathway appear absent in the bacterium, the relevance of this function remains unclear. CYP51B1 displays UV-visible spectral properties typical of P450 enzymes with a Soret peak at 419 nm in the oxidised substrate-free form. (73) Upon binding of the substrate analogue estriol, the Soret peak shift towards 392 nm due to the heme-iron spin-state shift from predominantly low-spin to partially/predominantly high-spin. (72,74) Mtb CYP51B1 also forms the characteristic P450 ferrous-CO complex with a Soret maximum near 450 nm. However, this collapses over a period of several minutes to form the P420 species. Subsequent studies on the formation of P420 species revealed that the P420 form can be reversible by deprotonating the cysteine ligand to the heme iron and by substrate binding to stabilize the active form of CYP51B1 (75). Various clinically used azoles also bind particularly tight to CYP51B1 (73,76) but as it is apparently not essential for viability or infectivity of the bacterium, (64,69) it appears unlikely that CYP51B1 is a major target for azole drugs.

The crystal structure of Mtb CYP51B1 has been determined for the ligand-free, 4-phenylimidazole-bound, fluconazole-bound and estriol-bound forms (74,77) which provides the first insight into the three-dimensional structure of a sterol demethylase P450 shown in Figure 1.6. The overall structures of CYP51B1 revealed a substantial distortion of the I-helix that is different to other known P450 structures. (78) Depending on the ligand either 4-phenylimidazole or fluconazole, the precise position of the ‘kink’ differs which enables appropriate conformational changes of the enzyme. Structural analysis of CYP51B1 also defined two distinct channels that suggest discrete sites of substrate/product entry and exit through movement of the BC and FG regions. (74,77) Mapping positions of known drug-resistance conferring mutations in fungal CYP51 enzymes onto the Mtb CYP51 azole-bound structures revealed that mutations were largely located not in the active site area, but instead in the conformationally mobile or interdomain regions and at the proximal face of the heme in the cysteine binding pocket. (77,79)

Work has also progressed in the area of compound screening, and both type I and type II ligands were identified. The leading type I ligand, 4, 4'-dihydroxybenzophenone (DHBP) was crystallized in its complex with CYP51B1 and structural ordering was observed particularly in the BC loop region. (80) The type II ligands, α-ethyl-N-4-pyridinyl-benzeneacetamide (EPBA) and the related 2-(benzo[d]-2,1,3-thiadiazole-4-sulfonyl)-2-amino-2-phenyl-N-(pyridinyl-4)-acetamide (BSPPA) also produced a degree of structural reordering and crystal structures with these compounds demonstrated direct coordination of the ligand's pyridine nitrogen to the heme iron. (81) Further elaboration of these second-generation compounds containing an N-(4-pyridyl)-formamide moiety may lead to CYP51B1-specific inhibitors. A number of the second-generation compounds were shown to possess a greater binding affinity for the CYP51E sterol demethylase from the parasitic protozoan pathogen *Trypanosoma cruzi* and the molecule α-[[(4-methylcyclohexyl)carbonyl]amino]-N-4-pyridinyl-1-H-indole-3-propanamide displayed selective and potent activity against *T. cruzi* in an infected mouse macrophage model. (82)

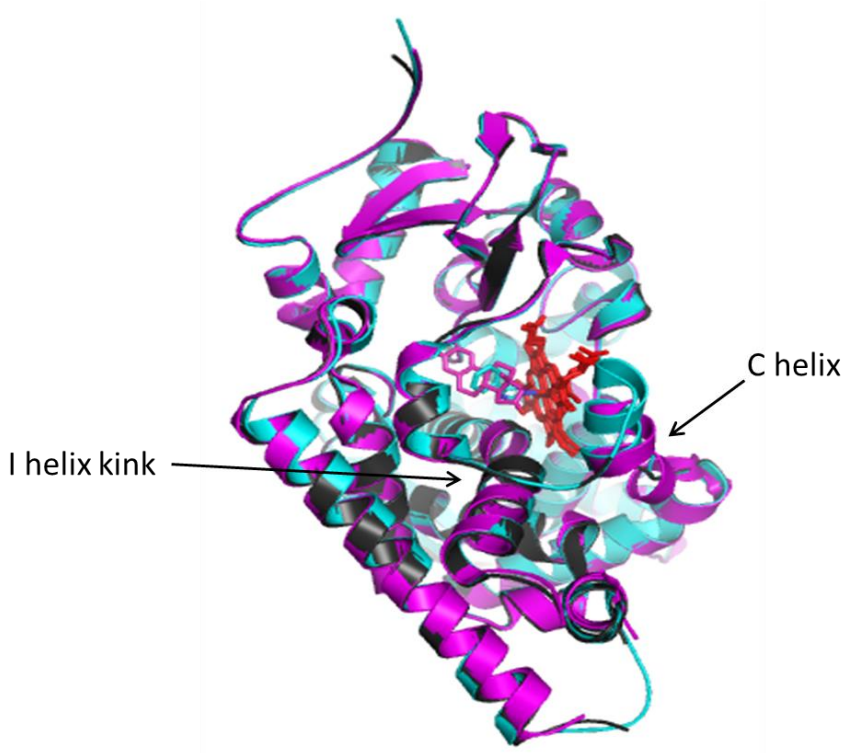


Figure 1.6 Superposition of ligand-free, estriol-bound and 4-phenylimidazole-bound *M. tuberculosis* CYP51B1 crystal structures.
The structure of CYP51BI was solved in the ligand-free form (PDB 1H5Z, in grey), and in forms bound to the substrate analog estriol and to the inhibitor 4-phenylimidazole (PDB 1X8V in magenta and PDB 1E9X in cyan, respectively). (74,77) Heme cofactors for all

structures are shown in red sticks while the estriol and 4-phenylimidazole ligands are shown in magenta and cyan sticks, respectively. The structures are viewed from the substrate entry/product exit site. The overall structures of CYP51B1 reveal a large kink of the I helix. This I helix kink is structurally perturbed upon ligand binding, enabling conformational changes that allow CYP51B1 to adapt to binding different size ligands. Moreover, in the ligand-free form, the C helix and the C-terminal part of the BC loop are disordered. Binding of 4-phenylimidazole triggers ordering of the C helix and estriol binding restores one helical turn on the C helix.

1.5.2 CYP121A1: Bacterial secondary metabolism essential for Mtb viability

The next Mtb P450 targeted for characterisation is CYP121A1 encoded by *Rv2276* gene, which exhibits limited similarity at the amino acid level with selected polyketide oxidising P450s. The *CYP121A1* gene was also shown to be an essential gene in Mtb H37Rv growth. (69,83) Its essentiality for Mtb viability was confirmed by the recent demonstration that *CYP121A1* could be deleted from the Mtb H37Rv genome only when a second copy of the gene was integrated elsewhere on the chromosome. (84) Nonetheless, studies of 100 clinical isolates have found that *CYP121A1* was among the 5.5% of Mtb genes partially or completely deleted and the physiological effects on these deletion strains are unknown. Hypotheses have been made that such deletions may confer a short-term evolutionary advantage, such as evading the host immune system, providing antibiotic resistance or limiting latency which may be advantageous to the pathogen at certain stages in infection. (85) The importance of *CYP121A1* was further reinforced in other studies that identified this gene as consistently expressed across a panel of clinical strains. (86)

CYP121A1 was successfully expressed in *E. coli* and the protein purified to homogeneity, which allowed its further characterisation. Both UV-visible spectroscopy and resonance Raman analysis indicated that, in the resting form, the ferric heme iron of CYP121A1 was predominantly in a low-spin (LS) state and on reduction to the ferrous form followed by complexation with carbon monoxide produced a stable Fe^{II}-CO complex, unlike CYP51B1 which rapidly converted to the P420 state. Other spectroscopic studies such as electron paramagnetic resonance (EPR) and magnetic circular dichroism provided evidence for the coordination of the heme iron by a proximal cysteinate. Extremely tight binding of azole anti-fungal drugs is also a feature for this P450 as the dissociation constant (K_d) values for clotrimazole, econazole and miconazole are the lowest for any of the Mtb P450s.

(87) Moreover, the order of affinities of these azoles correlated with their minimal inhibitory concentration (MIC) potency as anti-mycobacterial agents, suggesting that CYP121A1 may be the true Mtb target for the azoles. (88) Clotrimazole and econazole, when applied in combination with either of the commonly used front line anti-TB drugs rifampicin and isoniazid, displayed synergistic anti-mycobacterial activity. (89) Additionally, econazole was able to dramatically reduce the Mtb burden in the lungs and spleen of infected mice. (90) Therefore, it is likely that these azole drugs could be used to target CYP121A1 and other Mtb P450s. In previous studies, optical binding studies using a variety of lipids, steroids and polyketides failed to induce heme spectral perturbations consistent with substrate-like binding, and thus the natural substrate for CYP121A1 remained unknown. However, studies on the gene product of *Rv2275*, a gene located adjacent to *CYP121A1* have revealed the substrate for CYP121A1. *Rv2275* encodes a new class of enzyme known as a cyclic dipeptide (CDP) synthase; whose preferred substrate in this isoform is two molecules of L-tyrosine linked to their aminoacyl tRNAs and bound to the *Rv2275*-encoded CDP synthase in its dimeric form. The CDP synthase catalyzes the production of the CDP cyclo-*L*-tyrosine-*L*-tyrosine (cYY). (91) cYY was shown to bind to CYP121A1 and to induce a type I spectral shift, and is efficiently converted into a single major product. Further analyses have shown that CYP121A1 catalyzes C-C bond formation by oxidative coupling of the cYY aryl side chains to make a metabolite named mycocyclosin. (92) The physiological role of mycocyclosin is yet to be determined, but members of this diketopiperazine class of compounds have been found to play important roles in immunosuppression and blockage of cation channels. (93,94) In *B. subtilis*, a CDP synthase (*YvmC*) and a P450 (*CYP134A1*) were identified with a similar genomic orientation as seen for *Rv2275-CYP121A1*. These enzymes were shown to be involved in successive steps producing the CDP cyclo-*L*-leucine-*L*-leucine (cLL), followed by a three-step CYP143A1-mediated oxidation of the CDP to produce pulcherriminic acid, a precursor of the extracellular iron-chelating pigment pulcherrimin that is thought to play a role in ultraviolet light protection. (95) However, CYP134A1 is not highly related to CYP121A1 and performs a different non-coupling reaction, reflecting the distinct physiological functions of these pathways.

The ligand-free CYP121A1 crystal structure has been solved at very high resolution (1.06 Å) and has revealed interesting P450 structural features, including the nature of heme binding and the architecture of its active site. The heme is bound in two conformations related by 180º 'flip' of the cofactor and one of the pyrrole rings of the heme was distorted by interaction of the side chain of Pro346 (the residue immediately following the cysteine proximal ligand to the heme iron). (96) The active site cavity is quite large (1350 Å) but the region immediately above the heme was constricted due to the position of Arg386 and Ser237 side chains. A network of hydrogen bonded amino acids and water molecules were also observed. The extension of this hydrogen-bonding network from the protein surface to the active site, provide a clear pathway for efficient proton transfer to iron-oxo intermediates in the CYP121A1 catalytic cycle, with replenishment of the protons from bulk solvent. (96). The fluconazole-bound CYP121A1 structure was determined to 1.6 Å resolution, which revealed a novel mode of binding of the azole drug. Instead the fluconazole triazole nitrogen binding directly to the heme iron, the triazole nitrogen coordinated the heme iron via a bridging sixth water ligand that remains on the heme iron. This finding which was consistent with spectroscopic analysis, showed a less extensive Soret band red shift (on fluconazole binding) than is typical for such P450-azole interactions. (97) The CYP121A1 substrate-bound structure has shown that cYY was positioned in the active site with one of the tyrosyl moieties pointing towards the heme plane and hydrogen bonding with key active site residues. (92) Active site structures of CYP121A1 in complex with cYY and fluconazole are shown in Figure 1.7. Structures of CYP121A1 complexes with cYY analogs have also been solved to explore its substrate specificity and preference for other cyclic dipeptides. (98) Due to the availability of these high resolution crystal structures, fragment-based screening and drug design studies were possible and these approaches have been successful in producing preliminary novel series of potent and CYP121A1 isoform-selective inhibitors. (99) Elaboration of these fragments towards the distal active site of CYP121A1 has produced compounds with excellent selectivity against human drug metabolizing P450s and other Mtb P450 isoforms. (100) Further structural optimization to improve the cellular activity of these compounds will be required with the aim of developing novel classes of anti-TB drugs.

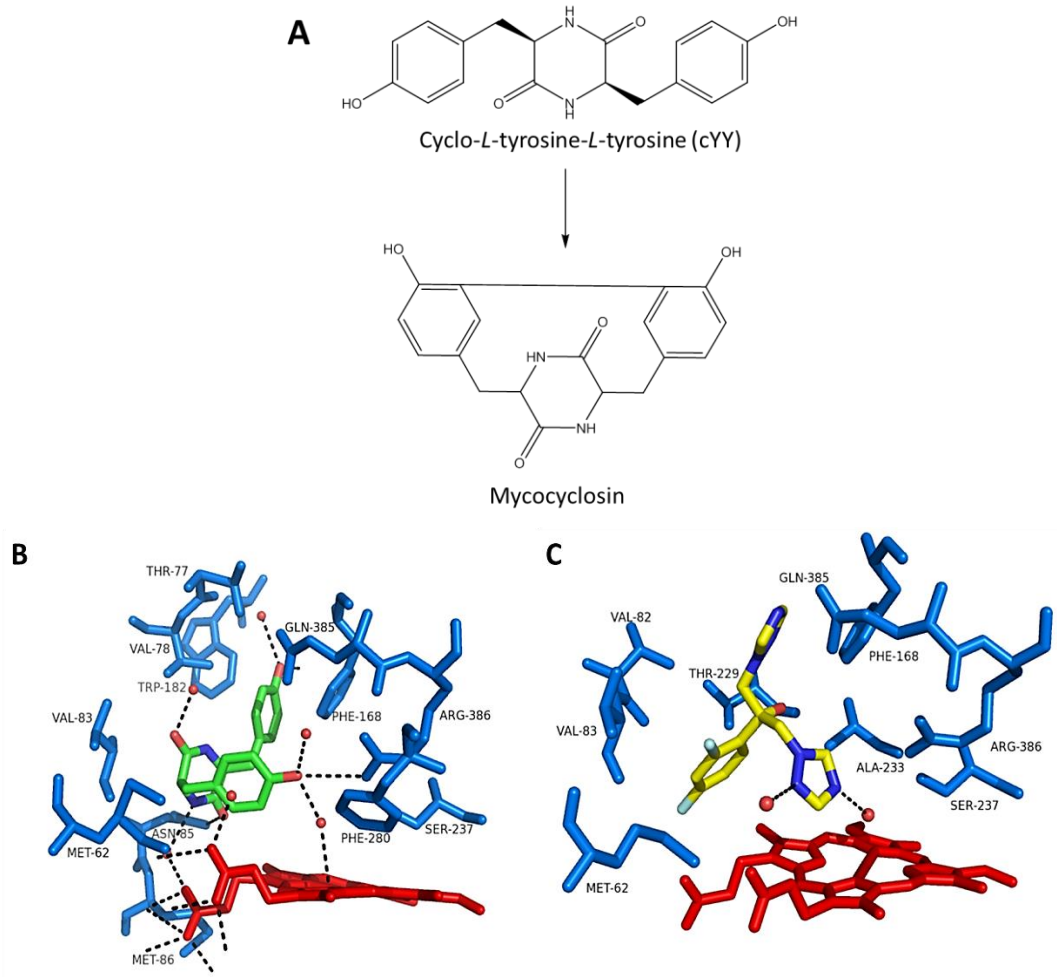


Figure 1.7 Reaction catalysed by CYP121A1 and crystal structures of CYP121A1 in complex with substrate and azole inhibitor.
(A) The C-C bond coupling of the substrate cyclo-*L*-tyrosine-*L*-tyrosine (cYY) to form a product named mycocyclosin, catalysed by *M. tuberculosis* CYP121A1. (92) (B) Stick representation of the active site of CYP121A1 in complex with cYY (PDB 3G5H). (92) cYY is shown in green, the heme is in red, CYP121A1 residues are in blue and water molecules are represented as red spheres. Probable H-bonds between CYP121A1, cYY and water molecules are indicated as dashed lines. (C) Stick representation of the CYP121A1 active site in complex with fluconazole. (97) Fluconazole is shown in yellow, CYP121A1 residues are in blue and water molecules are represented as red spheres. Probable H-bonds between fluconazole triazole nitrogen and the oxygen atoms of CYP121A1 water molecules are indicated as dashed lines, with one of these waters being the sixth ligand to the heme iron.

1.5.3 CYP125A1, CYP142A1 and CYP124A1: The cholesterol oxidase P450s

M. tuberculosis (Mtb) is capable of using cholesterol as a carbon source (101) and cholesterol has an important role in the infectivity and virulence of Mtb. (102–104) Host cholesterol is required to facilitate the entry of mycobacteria into macrophages and the catabolism of cholesterol was shown to be important for survival of pathogenic mycobacteria

in the host. (105,106) Interestingly, a large number of the genes involved in cholesterol catabolism in the Mtb-related soil bacterium *Rhodococcus* sp. strain RHA1 were also found conserved in Mtb *H37Rv*, and a degree of conservation in these actinobacterial pathways was identified. (107) Most of these genes in Mtb have been found to be under the control of *kstR* regulons, which encode a TetR-like transcriptional repressor. (108,109) Many of the genes found within these regulons overlap with the genes that have been shown to be essential *in vivo* in mouse or macrophage models of infection, including *mce4*, *igr* and the *fdaA5* operon. (58,64,110,111)

In Mtb, the *CYP125A1* (*Rv3545c*) gene that is located in the intracellular growth (*igr*) operon was shown to be the only Mtb *CYP* gene induced in macrophage infection and has been identified as essential for virulence during macrophage infection in mice. (64,111,112) *CYP125A1* is also found to be upregulated in Mtb-infected human dendritic cells that play a role in host cell immunity as well as Mtb pathogenicity. (113,114) Mtb CYP125A1 was recombinantly expressed and purified, and was shown to bind cholesterol and cholest-4-en-3-one with sub-micromolar affinity. Subsequently, CYP125A1 was found to catalyse the 27-hydroxylation of cholest-4-en-3-one on its side chain to preferentially form the (25S)-26-hydroxy product. (115) A number of CYP125A1 crystal structures were solved, including the ligand-free, cholest-4-en-3-one substrate-bound, econazole and LP10 inhibitor-bound, as well as androstenedione-bound forms. (116–118) The ligand-free crystal structure revealed a ‘letterbox’-like hydrophobic active site entry cavity that narrows in a funnel-like manner on approach to the heme centre. The catalytic site around the CYP125A1 heme iron and distal water consists of Leu117, Ala268, Val313, Phe316 and the methyl group of the conserved threonine Thr272. However, the distal water was not present in all ligand-free structures as the CYP125A1 was purified predominantly in a high-spin form. (116) The cholest-4-en-3-one bound structure reveals conformational changes in the H-helix and the N-terminal region of the I-helix that enclose the active site cavity and permit hydrophobic contacts between the I-helix and the substrate molecule. It is postulated that a minor structural rearrangement would allow a catalytically productive substrate orientation. (117) In the androstenedione-, econazole- and LP10-bound CYP125A1 structures, a similar position of the ligands were

also observed, but the pyridinyl ring of LP10 points into the heme pocket and hydrogen bonds with the active site water molecule adjacent to the axial water ligand to the heme iron.

(118)

Furthermore, it appears that a Mtb H37Rv *CYP125A1* deletion strain does not affect growth in the presence of cholesterol, which points to the presence of a compensatory cholesterol oxidase activity catalysed by one of the other Mtb P450s. *CYP142A1* (*Rv3518c*) is also located in the Mtb cholesterol regulon and was identified as the likely compensatory enzyme that maintained cholesterol degradation capacity in the absence of CYP125A1 function. When expressed recombinantly *in vitro*, CYP142A1 was discovered to catalyse either 27-hydroxylation of cholesterol/cholest-4-en-3-one (25R isomer) or generate 5-cholestenoic acid/cholest-4-en-3-one-27-oic acid from these substrates by successive sterol oxidations. (119,120) Despite having a similar substrate-oxidising role to CYP125A1, the protein sequence similarity between these two cholesterol oxidases P450 are surprisingly low (28 %). Nonetheless, the ligand-free CYP142A1 structure displays the same type of 'letterbox'-shaped access channel as described for CYP125A1 with the absence of an extended loop connecting the β1 and β2 strands. (119)

In addition, an apparently non-essential Mtb P450 (CYP124A1, the product of the*Rv2266* gene) was found to have 40% amino acid identity to CYP125A1 and can also catalyse C26-omega-hydroxylation of cholesterol and cholest-4-en-3-one, indicating evolutionary relationships. *CYP124A1* does not compensate for loss of *CYP125A1* function in Mtb, probably indicating that these steroids are unlikely to be true physiological substrates of CYP124A1. (120,121) In addition, given its preference for methyl-branched chain lipid substrates such as phytanic acid, CYP124A1 is more likely to play a role in fatty acid metabolism. CYP124A1 may also have the potential to oxidise a menaquinone-type derivative or precursor, as it is located in the same region as CYP128A1 on the Mtb chromosome. The structure of CYP124A1 was determined in the ligand-free and phytanic acid substrate-bound forms. (121) The structure of ligand-free CYP124A1 displays a high degree of similarity to CYP125A1 and CYP142A, again suggesting a common evolutionary origin. Similar to CYP142A1, the extended loop to connect the β1 and β2 regions is also not

seen in CYP124A1. Substrate binding to CYP124A1 induces a structural rearrangement of the FG helices and movement of the FG loop towards the phytanic acid ligand, closing over the access channel to the substrate cavity. The phytanic acid molecule is bound in a conformation optimal for ω-hydroxylation, with hydrophobic and polar interactions observed between different secondary structural elements and the substrate methyl-branched lipid chain and carboxylic acid groups, respectively. Additional solvent filled pockets observed in the active site cavity of phytanic acid-bound CYP124A1 suggest that these regions may be accommodated by parts of yet unknown physiological substrates for CYP124A1. (121) Structural views of CYP125A1, CYP124A1 and CYP142A1 are shown in Figure 1.8.

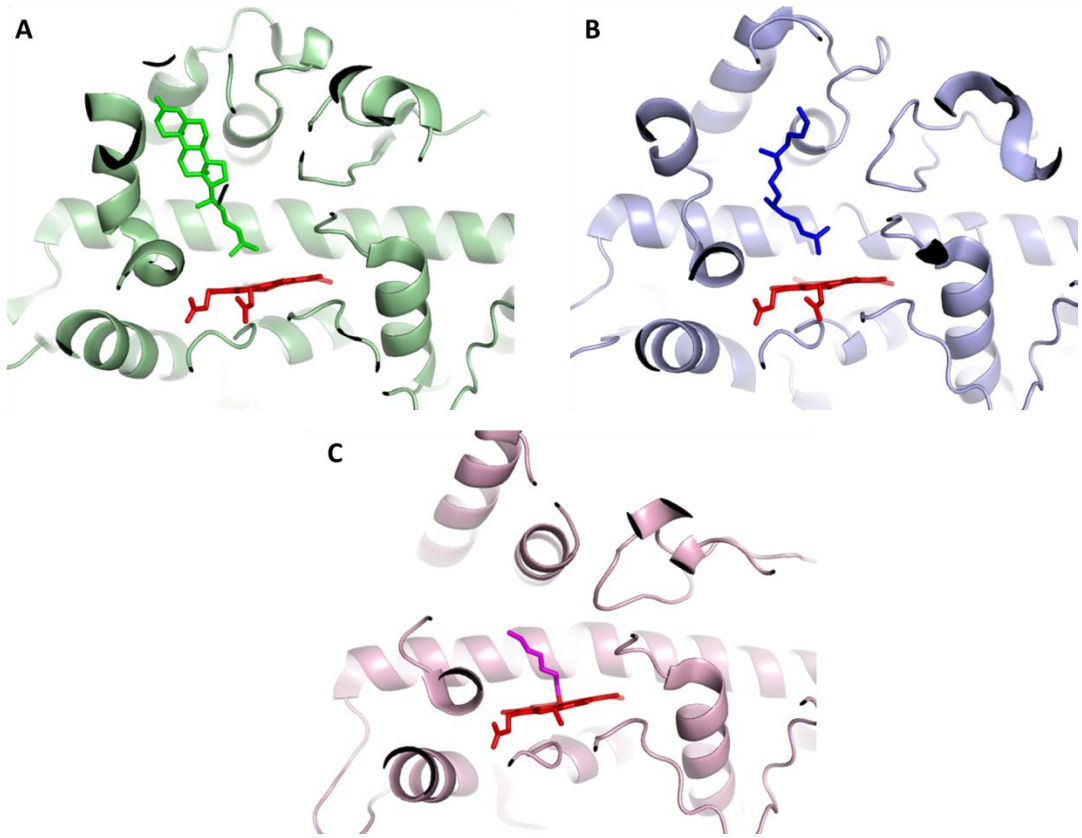


Figure 1.8 Structural views of *M. tuberculosis* P450 cholesterol oxidases.
Structural views of the active site regions of (A) CYP125A1 in complex with cholest-4-en-3-one (green sticks) (PDB 2X5W) (116), (B) CYP124A1 in complex with phytanic acid (blue sticks) (PDB 2WMH) (121) and (C) CYP142A1 in complex with tetraethylene glycol (magenta sticks) (PDB 2XKR) (119).

1.5.4 CYP128A1: Another essential gene in Mtb

In a genome-wide transposon hybridization study (TraSH), *CYP128A1* (*Rv2268c*) was the only P450 identified as essential for optimal growth of the pathogenic Mtb H37Rv strain under normal laboratory conditions. (69) The expression of the *CYP128A1* gene was found to be upregulated following Mtb starvation (122) and following exposure to the potential anti-TB drug lupulone. (123) The function of CYP128A1 was predicted based on its chromosomal location that is adjacent to a sulfotransferase, *stf3* (*Rv2267c*). Both of these genes, *CYP128A1* and *stf3*, together with *rv2269c*, were proposed to be involved in the biosynthetic pathway of sulfomenaquinone (SMK also known previously as S881) as shown in Figure 1.9 (124). Mutational analysis by thin layer chromatography (TLC) revealed that Δ*CYP128A1* and Δ*stf3* mutants did not produce a spot corresponding to SMK, but complementation with each respective gene restored SMK production. The Mtb P450 *CYP124A1* (also located adjacent to *stf3* in the Mtb genome) could not complement the lack of *CYP128A1* function. These findings confirm the role of CYP128A1 in SMK biosynthesis and indicate that it is the only Mtb P450 involved in this pathway. (124) Production of SMK involves the hydroxylation of menaquinone, a reaction type commonly mediated by P450 enzymes, followed by Stf3-mediated sulfation. CYP128A1 was predicted to catalyse an ω-hydroxylation of dihydromenaquinone, that subsequently allows the sulfation of this product at the terminal of its polyisoprenoid chain, producing SMK. (125) Menaquinone, the precursor to SMK is essential to Mtb viability because of its role in the electron transport chain and in regulating respiration. (126–128) Furthermore, since the *CYP128A1, stf3 and Rv2269c* genes involved in the production of SMK have no obvious counterparts in mycobacteria outside the Mtb complex, they may have a role specific for Mtb pathogenesis. To date, attempts to purify and characterize the CYP128A1 Mtb P450 have proven unsuccessful due to its insolubility. The hydrophobic nature of the putative dihydromenaquinone substrate suggests that it is membrane-associated, but there is no evidence that CYP128A1 has a membrane-anchor region of the type found in its eukaryotic P450 counterparts. (129)

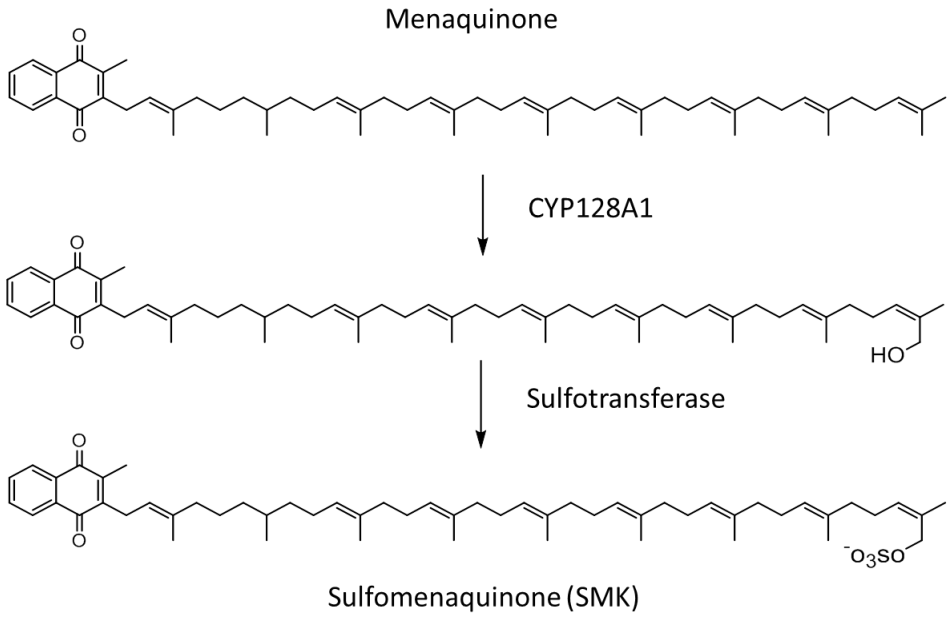


Figure 1.9 Proposed sulfomenaquinone (SMK) biosynthesis in *M. tuberculosis*. Menaquinone is hydroxylated by CYP128A1 followed by sulfation catalyzed by the sulfotransferase, Stf3.

1.5.5 CYP130A1 and CYP144A1: Characterisation of orphan P450s

In many cases, genomic localization does not give any clear indication as to what most Mtb P450s' physiological roles might be, thus making them orphan enzymes with no known function. In the search to define which particular isoforms are the key players in the physiology and/or pathogenicity of Mtb, the vaccine strain *M. bovis* BCG was analysed as a reference. *M. bovis* BCG is highly related to Mtb H37Rv, but with a series of deleted genomic regions called RD regions or 'regions of difference'. (58) These RDs are considered to contain key genes responsible for virulence and have been mapped onto the Mtb genome. Two of the Mtb P450-encoding genes are located in the *M. bovis* BCG RD regions. These genes are *CYP130A1* (*Rv1256c*) and *CYP141A1* (*Rv3121*) present in the RD 13 and RD 12 regions, respectively. (130) The potential of CYP130A1 and CYP141A1 as determinants of Mtb virulence is a major drive for their enzymatic characterisation. The crystal structures of CYP130A1 in ligand-free and econazole-bound forms were determined, which revealed interesting aspects of CYP130A1 oligomerisation. The ligand-free CYP130A1 crystallizes as a monomer in an open conformation and the econazole-bound CYP130A1 crystallizes as a dimer in a closed conformation, with an extensive dimerization

interface, shown in Figure 1.10 (131). This finding could not be replicated in the solution state binding studies, but it does reveal that econazole displays apparent cooperative binding to CYP130A1 and can promote its dimerization. (131) Although there is no compelling evidence for *CYP130A1* gene essentiality, work has progressed using compound library screening to identify ligands binding this P450. High-throughput compound screening studies identified a number of type II ligands, predominantly heterocyclic arylamines, and crystal structures reveal ligation of the molecules to the heme iron via nitrogen atoms. (132) The substrate selectivity for CYP130A1 remains undefined; hence it is presently difficult to determine its role in Mtb.

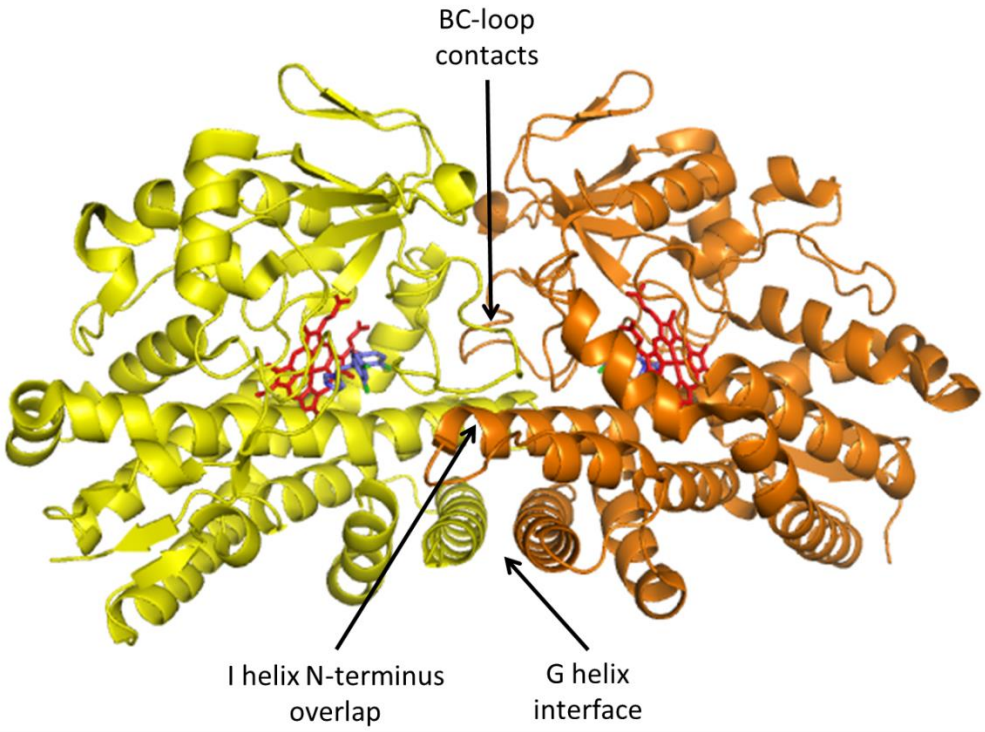


Figure 1.10 Dimerization of *M. tuberculosis* CYP130A1.
Crystal structure of CYP130A1 in complex with econazole (PDB 2UVN) (131). The monomers are coloured in yellow and orange, heme is in red and econazole is in blue. The dimerization interfaces of CYP130A1 are formed largely by interactions between the G helices in anti-parallel orientations, overlapping N termini of the I helices and multiple contacts in the BC-loop regions, as shown.

Another orphan Mtb P450 that was characterised is CYP144A1 (*Rv1777*), which is located in a region of the *Mtb* genome rich in genes encoding proteins of unknown function.

The *CYP144A1* gene was found to be non-essential for growth *in vitro* through gene knockout studies. However, the deletion strain was more sensitive to azole inhibition in culture, suggesting an important role for CYP144A1 in cell physiology and/or in mediating azole resistance. (119) Furthermore, the *CYP144A1* gene was upregulated in Mtb when treated with vancomycin at 10x its MIC, possibly indicating the involvement of CYP144A1 in the Mtb stress response mechanism. (133) Evolutionary analysis revealed that CYP144A1-like genes were conserved within the *Mycobacterium* genus, suggesting the retention of an important catalytic role. (134) Spectroscopic analysis of purified CYP144A1 demonstrates heme thiolate coordination in its LS, ferric form. However, the cysteinate is protonated to its thiol form in both ligand-free and the CO-bound ferrous forms. These data suggest that the binding of the CYP144A1 substrate may be crucial for maintaining cysteine thiolate coordination and catalytic activity. The tendency to form the P420 state in reduced/reduced-CO bound forms of CYP144A1 produces some issues in quantifying the P450 using the Omura and Sato method. However, a CYP144A1 heme extinction coefficient can be determined using the pyridine hemochromagen technique. (119,135) Various azole antifungals also binds tightly to CYP144A1 making it another possible target for novel anti-TB drug development. CYP144A1 was also reported to be produced in two different forms, one being the full-length version and the other a truncated version generated from a leaderless transcript. (136) However, the crystal structures of both forms of CYP144A1 were similar to each other and show that the active site cavity is large and predominantly hydrophobic, which possibly means that its natural substrates are bulky hydrophobic molecules. The residues located around the active site cavity were also different from other known structures of Mtb P450s hence suggesting a unique role and substrate selectivity. (134)

1.5.6 Mtb P450 redox partners systems

Redox partners systems are necessary to provide electrons to Mtb P450s to enable them to oxidize their substrates. For bacterial P450s, a typical class I redox partner system involves a NAD(P)H-binding FAD-containing ferredoxin reductase (FDR) that transfers

electrons to the P450 via an iron-sulfur cluster-containing ferredoxin or (in some cases) a FMN-binding flavodoxin. (44) The proposed redox partner system for Mtb P450s is shown in Figure 1.11. Based on the Mtb H37Rv genome, there are several genes encoding ferredoxin-type proteins and there is no evidence for stand-alone flavodoxin-encoding genes. Two of the ferredoxins that are likely P450 redox partner systems are those located chromosomally adjacent to P450s. First, the 3Fe-4S ferredoxin (Fdx) encoded by *Rv0763c* is chromosomally adjacent to CYP51B1 and was shown to support the sterol demethylase activity of CYP51B1. (72,75) Second, the 3Fe-4S ferredoxin encoded by *Rv1786* lies adjacent to CYP143A1. Other potential redox partners are FdxA (encoded by *Rv2007c*) and FdxC (encoded by *Rv1177*) which are likely bind to 4Fe-4S or 3Fe-4S cofactors. *Rv3503c* also encodes a putative 3Fe-4S ferredoxin, FdxD. A potential NAD(P)H-dependent flavin-containing reductase is the flavoprotein reductase A (FprA) encoded by *Rv3106*, which is an adrenodoxin reductase (ADR) homologue. (137) In eukaryotic mitochondrial P450 systems, a prokaryotic-like class I redox system operates which involves the FAD-containing ADR and the 2Fe-2S adrenodoxin AD. (41) FprA is a soluble enzyme that transfers electrons to Fdx and thus is a potential Mtb P450 redox partner system. (138) Another FAD-containing reductase protein, FdR encoded by *Rv0688*, was also shown to participate in a productive redox chain involving CYP51B1 and Fdx to reconstitute lanosterol demethylation. (139) Another potential FAD-containing ferredoxin reductase redox partner is encoded by the *Rv1896c* gene. There are also single component ferredoxin reductase/ferredoxin fusion proteins that could provide more efficient systems for electron transfer to Mtb P450s. These fusion proteins are FprB (an FprA-Fdx like fusion protein) and FdxB (a putative ferredoxin reductase-ferredoxin fusion in which the ferredoxin is likely to bind to a 2Fe-2S cluster). To date, much work remains to be done to characterize these potential Mtb P450 redox partner systems.

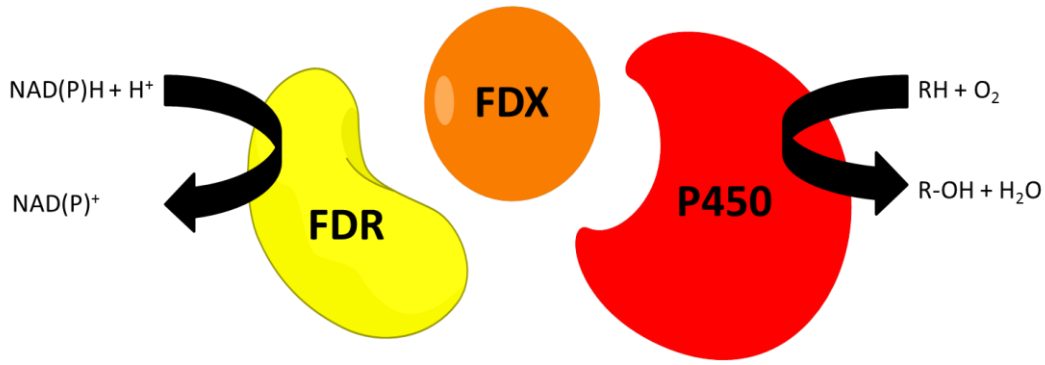


Figure 1.11 Proposed redox partner system for *M. tuberculosis* P450s.
The Class I redox partner system usually consists of a NAD(P)H-binding FAD-containing reductase enzyme (FDR), which transfers electrons to the cytochrome P450 via an iron-sulfur cluster-containing ferredoxin (FDX) protein. Substrate binding to the P450 typically increases its heme iron reduction potential, facilitating electron transfer from NAD(P)H via FDR and FDX partners.

1.5.7 Azoles as anti-tubercular drugs

Studies on Mtb P450s have indicated the potential of these enzymes as novel drug targets for TB treatment, based on the facts that some of the P450s are encoded by genes essential for Mtb growth (CYP121A1, CYP125A1 and CYP128A1) and have important roles in cholesterol metabolism and secondary metabolite production. The therapeutic potential of Mtb P450s is further evident as almost all of the Mtb P450s characterised bind very tightly to the most potent antimycobacterial azoles, which are clotrimazole, econazole and miconazole. Various studies have shown that clotrimazole and econazole are effective against persistent and multidrug-resistant strains of Mtb (140,141) as well as cooperative in their antimycobacterial activity with the current anti-TB drugs rifampicin and isoniazid. (89) In addition, econazole was also found to be effective in reducing bacterial burden by 90 % in the lungs and spleen of infected mice. (90) Crystallographic studies of Mtb P450s have shown the binding mode of various azoles and provide an explanation to the mechanism of inhibition of these azoles. They inhibit P450 catalysis by reversibly coordinating the ferric heme iron through an azole nitrogen in the sixth distal ligand position, replacing the weakly bound distal water molecule and competitively inhibiting substrate access to active site. Consequently, these findings have inspired further studies on the development of new scaffolds based around imidazole, triazole or other heme-coordinating moieties. Many of these new molecules were shown to have good MICs against Mtb. (105,142) However, most

of these azoles may not be suitable as scaffolds for front-line oral anti-TB drug candidates due to their poor bioavailability and cross-reactivity with human P450s and other proteins, as well as the potential for development of resistance towards azoles. Nonetheless, there are improvements in pharmacokinetic and toxicity profiles of several novel azole drugs, which could make the application of appropriate azoles in anti TB therapy successful.

1.6 Drug discovery approaches

1.6.1 High-throughput screening

As there are an increasing number of potential therapeutic targets emerging from the field of functional genomics and a rapid development of large compound libraries derived from parallel and combinatorial chemical synthesis, there is a need for high-throughput drug discovery screening technology. High-throughput screening (HTS) is a drug discovery approach that has gained widespread popularity over the last two decades and is widely exploited by pharmaceutical industries. The main objective of using a HTS approach is to accelerate the drug discovery process by screening large compound libraries at a rate that may exceed thousands of compounds per day or per week against selected and specific targets. A typical HTS may screen up to 10000 compounds per day and ultra-HTS can conduct 100000 assays per day. (143)

Two stages of compound screening are usually performed in HTS. A primary screen is usually less quantitative than the secondary screen and will only identify positive hits. The secondary screening (involving biological and biochemical assays) will then verify the hits. Standard assays such as ELISA, proliferation/cytotoxicity assays, reporter assays and binding assays are usually applied to HTS. These assays are classified as either heterogeneous or homogeneous assays depending on the degree of complexity in performing the assay. Homogeneous assays involves a one-step mixing of all assay components, incubating and reading the assay plate, which is advantageous in terms of speed. However, heterogeneous assays, for example ELISA, are widely used in industries as they are more sensitive. (144) A variety of strategies have been used for HTS assays,

including the measurement of catalytic activity from a purified enzyme, a reconstituted complex, a cellular extract or a phenotype. Given the right compound library, assay design becomes the determining factor for identifying compound hits using HTS (145). Moreover, micro-titre plate-based assay methodologies using radiometric, fluorescence, colorimetric and other detection methods, in assay volumes of a few microlitres, can be performed using automated screening platforms, and were developed to enhance the application of HTS in both industry and academia. (143)

A fundamental feature of HTS is that it assumes no prior knowledge of the compound binding site on the target protein. As a consequence, HTS can identify molecules with chemical novelty and novelty in mechanism of action. (146) Despite the value of HTS being criticised recently, HTS data quality remains very high as the assay measurement reflects the true value and is monitored by rigorous quality assurance methods. (147,148) In addition, the cost of implementing HTS is relatively low considering the number of compounds being screened and the sustainability of its infrastructure. (149) Although HTS is considered to be the opposite of rational drug design, the high rate of identifying compound hits that can be developed into leads is still advantageous for pharmaceutical companies. The success rate of finding leads for targets in HTS is approximately 50%. (150) Between 48 and 84% of the molecules in lead optimisation stage in three pharmaceutical companies have been derived from HTS. HTS has led to the development of a number of drugs including Tykerb (GlaxoSmithKline), Januvia (Merck) and Sprycel (Bristol-Myers Squibb). (149) As the timeline for HTS to product launch can exceed 10 years, many of the drugs launched in the upcoming decade are expected to be derived from HTS.

1.6.2 Fragment-based screening

In parallel to HTS, there is a need for more rational approaches in drug discovery. Among the latest developments in drug discovery is an approach known as fragment-based drug discovery (FBDD), or fragment-based screening (FBS). Rather than screening millions of fully built, drug-sized chemical compounds as lead starting points, FBDD begins with screening less numerous collections of smaller compounds that may only exhibit weak

binding affinity. Fragments are usually defined as having less than 20 non-hydrogen or 'heavy' atoms. Follow-up medicinal chemistry strategies are then applied to increase binding affinity by elaborating the starting fragment hits. In this case, structural information plays an important role in the fragment-based drug design strategies, which involves the attempt to build a lead 'piece-by-piece'. Therefore, FBDD is considered as structure-based drug design. (151)

There are several advantages of implementing FBS in drug discovery. The generally accepted advantage of FBS compared with the more conventional approach of screening typical lead-like compounds (MW 300-500) is that the smaller the fragment, the higher the hit rate. Many drug lead compounds in HTS compound libraries yielded few hits, or hits that were false positives when screened against newer or more difficult targets. (152) Therefore, a very wide range of molecules may need to be explored to find the rare ones that actually bind. In FBS, the sampling of chemical space is much more efficient with fragments than with larger molecules. In addition, a higher proportion of diverse drug space can be covered by only few thousand fragments, which are practically easier to assemble, maintain and screen. Although fragments form fewer interactions with the protein target, which results in lower binding affinities, smaller fragments can have higher ligand efficiency, leading to smaller drugs with better chances for favourable pharmacokinetics. (153,154)

Since the success of the technique called 'SAR by NMR' which stands for 'structure-activity relationships (SAR) by nuclear magnetic resonance (NMR)' (155), several advanced biophysical techniques such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), X-ray crystallography and many more are applied concurrently during FBS hit validation. (156) The possibility of applying various well-known and practical biophysical techniques has resulted in the rise of FBDD approaches in both industry and academia. To date, more than 30 drug candidates from FBS have entered the clinic, with two approved and several more in advanced trials. (157)

1.6.3 Application of drug discovery approaches to Mtb P450s

In the past decade, both HTS and FBS approaches have been applied to develop new P450 inhibitors to target Mtb P450 enzymes. New scaffolds for inhibition of CYP51B1 were identified by HTS of libraries of organic compounds. (80,81) The HTS assay used was based on the spectral changes of the P450 associated with the either substrate binding (type I) or inhibitor binding (type II) to the heme. Three hits were identified out of the 20000 compounds screened, which are alpha-ethyl-N-4-pyridinyl-benzeneacetamide (EPBA), 2-(benzo[d]-2,1,3-thiadiazole-4-sulfonyl)-2-amino-2-phenyl-N-(pyridinyl-4)-acetamide (BSPPA) and 4,4'-dihydroxybenzophenone (DHBP). Crystal structures of CYP51B1 complexes with both EPBA and BSPPA demonstrated direct coordination of the ligand pyridine nitrogen group to the heme iron. In addition, EPBA was also shown to be inhibitory to the growth of Mtb. (81) A common N-(4-pyridyl)-formamide moiety was used to develop second-generation molecules that may lead to CYP51B1-specific inhibitors. (82) In addition, two arylamine compounds were also identified to bind CYP130A1 from a HTS assay, and crystal structures revealed ligation of the molecules to the heme iron via nitrogen atoms. (132) It was suggested that certain compounds of this type could serve as selective inhibitors of subsets of P450 enzymes. However, toxicity issues of various arylamines might compromise their exploitation as antibiotics. (158)

Most recently, a FBDD campaign targeting several of the Mtb P450s was initiated to develop more potent and specific inhibitor for these enzymes. The first successful application of FBS approaches to a Mtb P450 was reported for CYP121A1. (99) Multiple biophysical techniques such as differential scanning fluorimetry (DSF), NMR and ITC were used, revealing several fragment hits. However, four fragment hits characterised by X-ray crystallography enabled fragment elaboration by fragment-merging and retrofragmentation strategies. Low nanomolar affinity leads that made polar contacts with the heme iron were successfully developed, but they did not show antimycobacterial activity. This may be due to cell wall permeability issues and drug-efflux mechanisms. (99,100) Nonetheless, an array of FBS applications against 7 of the Mtb P450 enzymes has proved useful in building relationships between the enzymes, identifying promiscuous and more selective fragment

scaffolds for these enzymes, and illustrating the distinct binding properties of enzymes within a family of proteins that share a conserved catalytic mechanism and cofactor. (159)

1.7 Summary and research objectives

1.7.1 Mtb P450s as novel drug targets

The discovery of 20 genes encoding cytochrome P450 enzymes in the *Mycobacterium tuberculosis* H37Rv genome sequence has pointed to the significance of these enzymes in the physiology and pathogenicity of this human pathogen. Biochemical characterisation of various Mtb P450s has proven that these enzymes have unique and unexpected roles in the bacterium, such as host cholesterol oxidation and secondary metabolite production. In addition, the gene essentialities of some of these P450s have been reported, implicating these enzymes as potential drug targets for new classes of anti-TB drugs. The findings that many of the Mtb P450s bind azole drug inhibitors with high affinity have driven research towards novel drug discovery. The development of Mtb P450 inhibitors has progressed beyond azole antifungals in the area of compound screening and fragment-based approaches. Such studies provide the foundations for rational design and synthesis of more potent and selective Mtb P450 inhibitors. Ideally, novel azole derivatives or new chemical scaffolds developed could simultaneously target several key Mtb P450s, hence minimizing the possibility of drug resistance which is the main problem for the treatment of tuberculosis.

In summary, biophysical and structural characterisations of Mtb P450 enzymes may provide valuable insights regarding the P450 active site architecture, substrate/ligand specificity and binding mode. Consequently, understanding the structural characteristics of Mtb P450 enzymes may define the physiological roles of these enzymes in the bacterium. Based on these hypotheses, with the prospect of defining novel functions and identifying novel drug targets, characterisations of the remaining orphan Mtb P450s is of interest. *M. tuberculosis* CYP141A1 and CYP143A1 are orphan enzymes with unknown physiological

function in Mtb and will be characterised in this study through use of various biophysical and structural techniques.

1.7.2 Research objectives

Characterisation of CYP141A1 is crucial in view of the fact that the CYP141A1-encoding gene (*Rv3121*) is located in one of the *M. bovis* BCG RD regions and is absent from the genome of the *M. bovis* BCG vaccine strain. Thus, CYP141A1 is potentially a determinant of Mtb virulence or important for host pathogenicity. Furthermore, the physiological role of CYP141A1 cannot be readily deduced from its amino acid similarities with other P450s of known function, or from the genomic localization of the gene. Although some aspects of the biophysical properties of CYP141A1 in Mtb were studied previously (Duyet Le Van, PhD thesis, University of Manchester 2011), there is no published work regarding the structural properties of this enzyme. CYP141A1 exhibits a unique reduced/CO complex absorbance spectrum at 440 nm which is distinct from the typical P450 spectrum, hence making it 'P440' rather than P450. An explanation for this unusual shift of the typical heme Fe^{II}-CO absorption spectrum is still unknown. In addition, CYP141A1 retains high affinity for azole antifungal drugs and thus is another potential P450 drug target for these compounds in Mtb. Efforts have been made to crystallize CYP141A1 in our laboratory previously, but the crystals produced poor diffraction data, possibly reflecting heterogeneous protein samples or weakly diffracting crystal forms. Therefore, optimisation of expression, purification and crystallization methodologies of CYP141A1 is necessary in order to produce a high quality crystal structure. Structural information for the ligand-free or ligand-bound forms of CYP141A1 is still unavailable, but much needed in order to reveal key features of its active site cavity and to define critical residues in substrate binding and catalysis. Moreover, the availability of a CYP141A1 crystal structure will be beneficial to the field of protein engineering and structure-based drug design.

In addition, in order for P450s to function, redox partner systems are required. Mtb P450s follows the classical prokaryotic class I systems in which a ferredoxin (FDX) mediates the electron transport between a ferredoxin reductase (FDR) and the P450. Out of the 20

Mtb P450s present in the genome of Mtb H37Rv, only the genes for CYP51B1 and CYP143A1 are located adjacent to ferredoxins. In a previous study, the 3Fe-4S ferredoxin adjacent to CYP51B1 was expressed and purified from *E. coli*, and was shown to support electron transfer and sterol demethylase activity of CYP51B1. (75) A second 3Fe-4S ferredoxin is encoded by *Rv1786*, which is adjacent to the gene encoding CYP143A1. In unpublished work, the *Rv1786* gene product was expressed and shown to bind a 3Fe-4S cluster by EPR analysis, and should thus be the preferred partner of CYP143A1. Therefore, it is also suggested that CYP143A1 has an important function in Mtb. The gene pair of *Rv1785c* (*CYP143A1*) and *Rv1786* is located in a region that is close to several PE and PPE genes (many of which are virulence factors, containing Pro-Glu or Pro-Pro-Glu motifs at their N-terminal regions), and to esat-6 family genes. Esat-6, the product of gene *esxA*, is an infectivity-related protein that is also deleted from the genome of the avirulent *M. bovis* BCG. (160) In the case of CYP143A1, expression, purification and crystallization methodologies were already been established by a previous PhD student (Shalini Swami, PhD thesis, University of Manchester 2015). However, structural data have not been yet been published and ligand-bound forms and improved structural resolutions are aims of the current project. Importantly, the availability of a CYP143A1 ligand-free crystal structure has provided the platform to move this research towards defining its natural substrate and through developing isoform-specific inhibitors by using drug discovery approaches. In this study, fragment-based screening approach are applied to identify novel ligands for CYP143A1 and to define CYP143A1 ligand preferences. Biophysical and structural characterisation of fragment-bound forms of CYP143A1 may provide a clearer understanding of the binding mode of these ligands, which will be a starting point for the development of isoform-specific inhibitors using fragment growing/merging/elaboration strategies.

1.8 References

1. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, Supply P, Vincent V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005;1(1):0055–61.

2. Cave A. The evidence for the incidence of tuberculosis in ancient Egypt. Br J Tuberc. 1939;33(3):142–52.

3. Daniel T. The origins and precolonial epidemiology of tuberculosis in the Americas: can we figure them out? Int J Tuberc Lung Dis. 2000;4(5):395–400.

4. Meachen GN. A short history of tuberculosis. London : John Bale, Sons & Danielsson Ltd. 1936.

5. Daniel TM. Rene Theophile Hyacinthe Laennec and the founding of pulmonary medicine. Int J Tuberc Lung Dis. 2004;8(5):517–8.

6. Daniel TM. Robert Koch and the pathogenesis of tuberculosis. Int J Tuberc lung Dis Off J Int Union against Tuberc Lung Dis. 2005;9(11):1181–2.

7. Glasser O. W. C. Roentgen and the discovery of the Roentgen rays. Am J Roentgenol. 1995 Nov;165(5):1033–40.

8. World Health Organization (WHO). Global tuberculosis report 2016. Geneva: WHO Press. www.who.int/tb/publications/global_report/en/ Date last accessed: July 7, 2017. Date last updated: 2016.

9. Ayvazian LF. History of tuberculosis. Lung Biol Heal Dis. 1993;66:1–20.

10. Gradmann C. Robert Koch and the white death: From tuberculosis to tuberculin. Microbes Infect. 2006;8(1):294–301.

11. Schatz A, Bugie E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. Clin Orthop Relat Res. 2005 Jan 1;55(437):3–6.

12. Riva MA. From milk to rifampicin and back again: history of failures and successes in the treatment for tuberculosis. J Antibiot (Tokyo). 2014 Sep 6;67(9):661–5.

13. Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov. 2013;12(5):388–404.

14. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR. *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science. 1994;263(5144):227–30.

15. Shi W, Zhang X, Jiang X, Yuan H, Lee JS, Barry CE, Wang H, Zhang W, Zhang Y. Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. Science. 2011;333(6049):1630–2.

16. Mikusova K, Slayden RA, Besra GS, Brennan PJ. Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. Antimicrob Agents Chemother. 1995 Nov;39(11):2484–9.

17. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell. 2001;104(6):901–12.

18. Schatz A, Bugle E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. Exp Biol Med. 1944 Jan 1;55(1):66–9.

19. Chakraborty S, Gruber T, Barry CE, Boshoff HI, Rhee KY. Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in *Mycobacterium tuberculosis*. Science. 2013;339(6115):88–91.

20. Bruning JB, Murillo AC, Chacon O, Barletta RG, Sacchettini JC. Structure of the *Mycobacterium tuberculosis* D-alanine:D-alanine ligase, a target of the antituberculosis drug D-cycloserine. Antimicrob Agents Chemother. 2011;55(1):291–301.

21. Salian S, Matt T, Akbergenov R, Harish S, Meyer M, Duscha S, Shcherbakov D, Bernet BB, Vasella A, Westhof E, Böttger EC. Structure-activity relationships among the kanamycin aminoglycosides: Role of ring I hydroxyl and amino groups. Antimicrob Agents Chemother. 2012;56(12):6104–8.

22. Sirgel FA, Tait M, Warren RM, Streicher EM, Böttger EC, van Helden PD, Gey van Pittius NC, Coetzee G, Hoosain EY, Chabula-Nxiweni M, Hayes C, Victor TC, Trollip A. Mutations in the *rrs* A1401G gene and phenotypic resistance to amikacin and capreomycin in *Mycobacterium tuberculosis*. Microb Drug Resist. 2012 Apr;18(2):193–7.

23. Horsburgh CR, Barry CE, Lange C. Treatment of tuberculosis. N Engl J Med. 2015;373(22):2149–60.

24. Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence. 2013 Jan 27;4(1):3–66.

25. Brennan PJ. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis. 2003;83(1–3):91–7.

26. Barry CE, Crick DC, McNeil MR. Targeting the formation of the cell wall core of *M. tuberculosis*. Infect Disord Drug Targets. 2007 Jun;7(2):182–202.

27. Medjahed H, Gaillard JL, Reyrat JM. *Mycobacterium abscessus*: a new player in the mycobacterial field. Trends Microbiol. 2010;18(3):117–23.

28. Estabrook RW. A passion for P450s (rememberances of the early history of research on cytochrome P450). Drug Metab Dispos. 2003;31(12):1461–73.

29. Klingenberg M. Pigments of rat liver microsomes. Arch Biochem Biophys. 1958;75(2):376–86.

30. Omura T, Sato R. Liver microsomes spermidine in the extraction of the deoxyribosyl-synthesizing system from *Escherichia*. J Biol Chem. 1962;237(April):1375–6.

31. Cupp-Vickery JR, Han O, Hutchinson RC, Poulos TL. Substrate-assisted catalysis in cytochrome P450eryF. Nat Struct Biol. 1996;3:632–7.

32. Ogura H, Nishida CR, Hoch UR, Perera R, Dawson JH, Ortiz De Montellano PR. EpoK, a cytochrome P450 involved in biosynthesis of the anticancer agents epothilones A and B. Substrate-mediated rescue of a P450 enzyme. Biochemistry. 2004;43(46):14712–21.

33. Boddupalli SS, Pramanik BC, Slaughter CA, Estabrook RW, Peterson JA. Fatty acid monooxygenation by P450BM-3: Product identification and proposed mechanisms for the sequential hydroxylation reactions. Arch Biochem Biophys. 1992

Jan;292(1):20–8.

34. Lepesheva GI, Waterman MR. Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. Biochim Biophys Acta. 2007 Mar;1770(3):467–77.

35. Tyson CA, Lipscomb JD, Gunsalus IC. The role of putidaredoxin and P450 cam in methylene hydroxylation. J Biol Chem. 1972 Sep 25;247(18):5777–84.

36. Gunsalus IC, Wagner GC. [17] Bacterial P-450cam methylene monooxygenase components: Cytochrome m, putidaredoxin, and putidaredoxin reductase. Methods Enzymol. 1978;52(C):166–88.

37. Holden M, Mayhew M, Bunk D, Roitberg A, Vilker V. Probing the interactions of putidaredoxin with redox partners in camphor P450 5-monooxygenase by mutagenesis of surface residues. J Biol Chem. 1997 Aug 29;272(35):21720–5.

38. Ricagno S, de Rosa M, Aliverti A, Zanetti G, Bolognesi M. The crystal structure of FdxA, a 7Fe ferredoxin from *Mycobacterium smegmatis*. Biochem Biophys Res Commun. 2007;360(1):97–102.

39. Gutierrez A, Grunau A, Paine M, Munro AW, Wolf CR, Roberts GCK, Scrutton NS. Electron transfer in human cytochrome P450 reductase. Biochem Soc Trans. 2003;31(3):497–501.

40. Ewen KM, Kleser M, Bernhardt R. Adrenodoxin: The archetype of vertebrate-type [2Fe-2S] cluster ferredoxins. Biochim Biophys Acta - Proteins Proteomics. 2011;1814(1):111–25.

41. Müller JJ, Lapko A, Bourenkov G, Ruckpaul K, Heinemann U. Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. J Biol Chem. 2001;276(4):2786–9.

42. Hawkes D, Slessor K, Bernhardt P, Voss JJ De. Cloning, expression and purification of Cindoxin, an unusual Fmn-Containing cytochrome P450 redox partner. ChemBioChem. 2010;11(8):1107–14.

43. McLean KJ, Sabri M, Marshall KR, Lawson RJ, Lewis DG, Clift D, Balding PR, Dunford AJ, Warman AJ, McVey JP, Quinn A-M, Sutcliffe MJ, Scrutton NS, Munro AW. Biodiversity of cytochrome P450 redox systems. Biochem Soc Trans. 2005;33(4):796–801.

44. Hannemann F, Bichet A, Ewen KM, Bernhardt R. Cytochrome P450 systems-biological variations of electron transport chains. Biochim Biophys Acta - Gen Subj. 2007;1770(3):330–44.

45. McLean KJ, Luciakova D, Belcher J, Tee KL, Munro AW. Biological diversity of cytochrome P450 redox partner systems. Adv Exp Med Biol. 2015;851:299–317.

46. Poulos TL, Finzel BC, Gunsalus IC, Wagner GC, Kraut J. The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. J Biol Chem. 1985;260(30):16122–30.

47. Poulos TL, Finzel BC, Howard a J. Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450. Biochemistry. 1986;25(18):5314–22.

48. Poulos TL, Finzel BC, Howard AJ. High resolution crystal structure of Cytochrome P450cam. J Mol Biol. 1987;195:687–700.

49. Ravichandran KG, Boddupalli SS, Hasemann CA, Peterson JA, Deisenhofer J.

Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science. 1993;261(5122):731–6.

50. Crane BR. The structure of nitric oxide synthase oxygenase domain and inhibitor complexes. Science. 1997;278(5337):425–31.

51. Sundaramoorthy M, Terner J, Poulos TL. The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid. Structure. 1995;3(12):1367–78.

52. Hawkes DB, Adams GW, Burlingame AL, Ortiz De Montellano PR, De Voss JJ. Cytochrome P450cin (CYP176A), isolation, expression, and characterization. J Biol Chem. 2002;277(31):27725–32.

53. Ost TWB, Miles CS, Munro AW, Murdoch J, Reid GA, Chapman SK. Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3. Biochemistry. 2001;40(45):13421–9.

54. Brash AR. Mechanistic aspects of CYP74 allene oxide synthases and related cytochrome P450 enzymes. Phytochemistry. 2009;70(13–14):1522–31.

55. Rupasinghe S, Schuler MA, Kagawa N, Yuan H, Lei L, Zhao B, Kelly SL, Waterman MR, Lamb DC. The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. FEBS Lett. 2006 Nov 27;580(27):6338–42.

56. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537–44.

57. Bentley SD, Thomson NR, James KD, Harris DE, Quail MA, Harper D, Bateman A, Brown S, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Larke L, Murphy L, Oliver K, Rabbinowitsch E, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature. 2002;3(2):141–7.

58. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci USA. 2003;100(13):7877–82.

59. Blattner FR, Iii GP, Bloch CA, Perna NT, Burland V, Riley M, Collado-vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. The Complete genome sequence of *Escherichia coli* K-12. Science. 1997;277(5331):1453–74.

60. Waterman MR, Lepesheva GI. Sterol 14alpha-demethylase, an abundant and essential mixed-function oxidase. Biochem Biophys Res Commun. 2005;338(1):418–22.

61. Recchi C, Sclavi B, Rauzier J, Gicquel B, Reyrat J. *Mycobacterium tuberculosis* Rv1395 Is a class III transcriptional regulator of the AraC family involved in cytochrome P450 regulation. J Biol Chem. 2003;278(36):33763–73.

62. Holsclaw CM, Sogi KM, Gilmore SA, Schelle MW, Leavell MD, Bertozzi CR, Leary JA. Structural characterization of a novel sulfated menaquinone produced by stf3 from *Mycobacterium tuberculosis*. ACS Chem Biol. 2008;3(10):619–24.

63. Bhatt A, Fujiwara N, Bhatt K, Gurcha SS, Kremer L, Chen B, Chan J, Porcelli SA., Kobayashi K, Besra GS, Jacobs WR, Betts JC, Lukey PT, Robb LC, McAdam RA.,

Duncan K, Cox JS, Chen B, McNeil M, et al. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Nature. 2002;104(3):79–83.

64. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12989–94.

65. Cox RA, Garcia MJ. Adaptation of mycobacteria to growth conditions: A theoretical analysis of changes in gene expression revealed by microarrays. Gordon S V., editor. PLoS One. 2013 Apr 12;8(4):e59883.

66. Betts JC, McLaren A, Lennon MG, Kelly FM, Lukey PT, Blakemore SJ, Duncan K. Signature gene expression profiles discrimate between isoniazid- thiolactomycin-, and triclosan-treated *Mycobacterium tuberculosis*. AntimicrobAgents Chemother. 2003;47(9):2903–13.

67. Waddell SJ, Stabler RA, Laing K, Kremer L, Reynolds RC, Besra GS. The use of microarray analysis to determine the gene expression profiles of *Mycobacterium tuberculosis* in response to anti-bacterial compounds. Tuberculosis. 2004;84(3–4):263–74.

68. Brown PO, Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Schoolnik GK, Brown PO, Schoolnik GK. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc Natl Acad Sci U S A. 1999;96(22):12833–8.

69. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol. 2003;48(1):77–84.

70. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. Ramakrishnan L, editor. PLoS Pathog. 2011 Sep 29;7(9):e1002251.

71. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW. The P450 Superfamily : Update on new sequences, gene-mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol. 1993 Jan;12(1):1–51.

72. Bellamine A, Mangla AT, Nes WD, Waterman MR. Characterization and catalytic properties of the sterol 14alpha-demethylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 1999;96(16):8937–42.

73. Souter A, Mclean KJ, Smith WE, Munro AW. The genome sequence of *Mycobacterium tuberculosis* reveals cytochromes P450 as novel anti-TB drug targets. J Chem Technol Biotechnol. 2000;75(May):933–41.

74. Podust LM, Yermalitskaya L V., Lepesheva GI, Podust VN, Dalmasso EA, Waterman MR. Estriol bound and ligand-free structures of sterol 14α-demethylase. Structure. 2004;12(11):1937–45.

75. McLean KJ, Warman AJ, Seward HE, Marshall KR, Girvan HM, Cheesman MR, Waterman MR, Munro AW. Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. Biochemistry. 2006;45(27):8427–43.

76. Aoyama Y, Horiuchi T, Gotoh O, Noshiro M, Yoshida Y. CYP51-like gene of *Mycobacterium tuberculosis* actually encodes a P450 similar to eukaryotic CYP51. J Biochem. 1998;124(4):694–6.

77. Podust LM, Poulos TL, Waterman MR. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. Proc Natl Acad Sci U S A. 2001;98(6):3068–73.

78. Denisov IG, Shih AY, Sligar SG. Structural differences between soluble and membrane bound cytochrome P450s. J Inorg Biochem. 2012;108:150–8.

79. Lepesheva GI, Virus C, Waterman MR. Conservation in the CYP51 family. Role of the B helix/BC loop and helices F and G in enzymatic function. Biochemistry. 2003;42(30):9091–101.

80. Eddine AN, von Kries JP, Podust M V, Warrier T, Kaufmann SHE, Podust LM. X-ray structure of 4,4 '-dihydroxybenzophenone mimicking sterol substrate in the active site of sterol 14 alpha-demethylase (CYP51). J Biol Chem. 2008;283(22):15152–9.

81. Podust LM, Von Kries JP, Eddine AN, Kim Y, Yermalitskaya L V., Kuehne R, Ouellet H, Warrier T, Alteköster M, Lee JS, Rademann J, Oschkinat H, Kaufmann SHE, Waterman MR. Small-molecule scaffolds for CYP51 inhibitors identified by high-throughput screening and defined by X-ray crystallography. Antimicrob Agents Chemother. 2007;51(11):3915–23.

82. Chen CK, Doyle PS, Yermalitskaya L V., Mackey ZB, Ang KKH, McKerrow JH, Podust LM. Trypanosoma cruzi CYP51 inhibitor derived from a *Mycobacterium tuberculosis* screen hit. Matlashewski G, editor. PLoS Negl Trop Dis. 2009 Feb 3;3(2):e372.

83. Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2003 Jun 10;100(12):7213–8.

84. McLean KJ, Carroll P, Lewis DG, Dunford AJ, Seward HE, Neeli R, Cheesman MR, Marsollier L, Douglas P, Smith WE, Rosenkrands I, Cole ST, Leys D, Parish T, Munro AW. Characterization of active site structure in CYP121: A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem. 2008;283(48):33406–16.

85. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere Y-OL, Aman K, Kato-Maeda M, Small PM. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A. 2004 Apr 6;101(14):4865–70.

86. Gao Q, Kripke K, Saldanha A, Yan W. Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. Microbiology. 2005;151(1):5–14.

87. McLean KJ, Cheesman MR, Rivers SL, Richmond A, Leys D, Chapman SK, Reid GA, Price NC, Kelly SM, Clarkson J, Smith WE, Munro AW. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. J Inorg Biochem. 2002;91:527–41.

88. McLean KJ, Marshall KR, Richmond A, Hunter IS, Fowler K, Kieser T, Gurcha SS, Besra GS, Munro AW. Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes. Microbiology. 2002;148(10):2937–49.

89. Ahmad Z, Sharma S, Khuller GK. In vitro and ex vivo antimycobacterial potential of azole drugs against *Mycobacterium tuberculosis* H37Rv. FEMS Microbiol Lett. 2005;251(1):19–22.

90. Ahmad Z, Sharma S, Khuller G. Azole antifungals as novel therapeutic agents

against murie tuberculosis. FEMS Microbiol Lett. 2006;261(2):181–6.

91. Gondry M, Sauguet L, Belin P, Thai R, Amouroux R, Tellier C, Tuphile K, Jacquet M, Braud S, Courçon M, Masson C, Dubois S, Lautru S, Lecoq A, Hashimoto S, Genet R, Pernodet J-L. Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat Chem Biol. 2009 Jun 3;5(6):414–20.

92. Belin P, Le Du MH, Fielding A, Lequin O, Jacquet M, Charbonnier J-B, Lecoq A, Thai R, Courçon M, Masson C, Dugave C, Genet R, Pernodet J-L, Gondry M. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(18):7426–31.

93. Borthwick AD. 2,5-diketopiperazines: Synthesis, reactions, medicinal chemistry, and bioactive natural products. Chem Rev. 2012 Jul 11;112(7):3641–716.

94. Cornacchia C, Cacciatore I. 2, 5-Diketopiperazines as neuroprotective agents. Mini Rev Med Chem. 2012;12:2–12.

95. Cryle MJ, Bell SG, Schlichting I. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: A cyclo-l-leucyl-l-leucyl dipeptide oxidase. Biochemistry. 2010;49(34):7282–96.

96. Leys D, Mowat CG, McLean KJ, Richmond A, Chapman SK, Walkinshaw MD, Munro AW. Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450. J Biol Chem. 2003;278(7):5141–7.

97. Seward HE, Roujeinikova A, McLean KJ, Munro AW, Leys D. Crystal structure of the *Mycobacterium tuberculosis* P450 CYP121-fluconazole complex reveals new azole drug-P450 binding mode. J Biol Chem. 2006;281(51):39437–43.

98. Fonvielle M, Le Du MH, Lequin O, Lecoq A, Jacquet M, Thai R, Dubois S, Grach G, Gondry M, Belin P. Substrate and reaction specificity of *Mycobacterium tuberculosis* cytochrome P450 CYP121: Insights from biochemical studies and crystal structures. J Biol Chem. 2013;288(24):17347–59.

99. Hudson SA, McLean KJ, Surade S, Yang YQ, Leys D, Ciulli A, Munro AW, Abell C. Application of fragment screening and merging to the discovery of inhibitors of the *Mycobacterium tuberculosis* cytochrome P450 CYP121. Angew Chemie - Int Ed. 2012;51(37):9311–6.

100. Kavanagh ME, Coyne AG, McLean KJ, James GG, Levy CW, Marino LB, De Carvalho LPS, Chan DSH, Hudson SA, Surade S, Leys D, Munro AW, Abell C. Fragment-Based Approaches to the Development of *Mycobacterium tuberculosis* CYP121 Inhibitors. J Med Chem. 2016;59(7):3272–302.

101. Russell DG, VanderVen BC, Lee W, Abramovitch RB, Kim MJ, Homolka S, Niemann S, Rohde KH. *Mycobacterium tuberculosis* wears what it eats. Cell Host Microbe. 2010;8(1):68–76.

102. Larrouy-Maumus G. Cholesterol acquisition by *Mycobacterium tuberculosis*. Virulence. 2015 Jul 4;6(5):412–3.

103. Kumar GA, Jafurulla M, Chattopadhyay A. The membrane as the gatekeeper of infection: Cholesterol in host–pathogen interaction. Chem Phys Lipids. 2016;199:179–85.

104. Lovewell RR, Sassetti CM, VanderVen BC. Chewing the fat: Lipid metabolism and homeostasis during *M. tuberculosis* infection. Curr Opin Microbiol. 2016;29:30–6.

105. Pandey AK, Sassetti CM. Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci U S A. 2008;105(11):4376–80.

106. Kaul D, Anand PK, Verma I. Cholesterol-sensor initiates *M. tuberculosis* entry into human macrophages. Mol Cell Biochem. 2004;258(1–2):219–22.

107. Van der Geize R, Yam K, Heuser T, Wilbrink MH, Hara H, Anderton MC, Sim E, Dijkhuizen L, Davies JE, Mohn WW, Eltis LD. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. Proc Natl Acad Sci U S A. 2007 Feb 6;104(6):1947–52.

108. Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, Frita R, Ten Bokum A, Besra GS, Lott JS, Stoker NG. A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis.* Mol Microbiol. 2007;65(3):684–99.

109. Kendall SL, Burgess P, Balhana R, Withers M, Ten Bokum A, Lott JS, Gao C, Uhia-Castro I, Stoker NG. Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. Microbiology. 2010;156(5):1362–71.

110. Nesbitt NM, Yang X, Fontán P, Kolesnikova I, Smith I, Sampson NS, Dubnau E. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. Infect Immun. 2010;78(1):275–82.

111. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. J Exp Med. 2003;198(5):693–704.

112. Kendall SL, Rison SCG, Movahedzadeh F, Frita R, Stoker NG. What do microarrays really tell us about *M. tuberculosis*? Trends Microbiol. 2004 Dec;12(12):537–44.

113. Tailleux L, Waddell SJ, Pelizzola M, Mortellaro A, Withers M, Tanne A, Castagnoli PR, Gicquel B, Stoker NG, Butcher PD, Foti M, Neyrolles O. Probing host pathogen cross-talk by transcriptional profiling of both *Mycobacterium tuberculosis* and infected human dendritic cells and macrophages. PLoS One. 2008;3(1):e1403.

114. Wolf AJ, Linas B, Trevejo-Nuñez GJ, Kincaid E, Tamura T, Takatsu K, Ernst JD. *Mycobacterium tuberculosis* Infects Dendritic Cells with High Frequency and Impairs Their Function In Vivo. J Immunol. 2007;179(4):2509–19.

115. Capyk JK, Kalscheuer R, Stewart GR, Liu J, Kwon H, Zhao R, Okamoto S, Jacobs WR, Eltis LD, Mohn WW. Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27 steroids. J Biol Chem. 2009;284(51):35534–42.

116. McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D, Munro AW. The structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection. J Biol Chem. 2009;284(51):35524–33.

117. Ouellet H, Guan S, Johnston JB, Chow ED, Kells PM, Burlingame AL, Cox JS, Podust LM, De Montellano PRO. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. Mol Microbiol. 2010;77(3):730–42.

118. Ouellet H, Kells PM, Ortiz De Montellano PR, Podust LM. Reverse type I inhibitor of *Mycobacterium tuberculosis* CYP125A1. Bioorganic Med Chem Lett. 2011;21(1):332–7.

119. Driscoll MD, McLean KJ, Levy C, Mast N, Pikuleva IA, Lafite P, Rigby SEJ, Leys D, Munro AW. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: Evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. J Biol Chem. 2010;285(49):38270–82.

120. Johnston JB, Ouellet H, Montellano PRO De. Functional redundancy of steroid C 26 - monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. J Biol Chem. 2010;285(47):36352–60.

121. Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR, Montellano PRO De. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(49):20687–92.

122. Betts J, Lukey P, Robb L, McAdam R. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Nature. 2002;104(3):79–83.

123. Wei J, Liang J, Shi Q, Yuan P, Meng R. Genome-wide transcription analyses in *Mycobacterium tuberculosis* treated with lupulone. Brazilian J Microbiol. 2014;45(1):333–41.

124. Sogi KM, Holsclaw CM, Fragiadakis GK, Nomura DK, Leary JA, Bertozzi CR. Biosynthesis and regulation of sulfomenaquinone, a metabolite associated with virulence in *Mycobacterium tuberculosis*. ACS Infect Dis. 2016;2(11):800–6.

125. Holsclaw CM, Sogi KM, Gilmore SA, Schelle MW, Leavell MD, Bertozzi CR, Leary JA. Structural characterization of a novel sulfated menaquinone produced by stf3 from *Mycobacterium tuberculosis*. ACS Chem Biol. 2008;3(10):619–24.

126. Dhiman RK, Mahapatra S, Slayden RA, Boyne ME, Lenaerts A, Hinshaw JC, Angala SK, Chatterjee D, Biswas K, Narayanasamy P, Kurosu M, Crick DC. Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence. Mol Microbiol. 2009;72(1):85–97.

127. Honaker RW, Dhiman RK, Narayanasamy P, Crick DC, Voskuil MI. DosS responds to a reduced electron transport system to induce the *Mycobacterium tuberculosis* DosR regulon. J Bacteriol. 2010;192(24):6447–55.

128. Upadhyay A, Fontes FL, Gonzalez-Juarrero M, McNeil MR, Crans DC, Jackson M, Crick DC. Partial saturation of menaquinone in *Mycobacterium tuberculosis* : function and essentiality of a novel reductase, MenJ. ACS Cent Sci. 2015;1(6):292–302.

129. McLean KJ, Belcher J, Driscoll MD, Fernandez CC, Le Van D, Bui S, Golovanova M, Munro AW. The *Mycobacterium tuberculosis* cytochromes P450: physiology, biochemistry & molecular intervention. Future Med Chem. 2010;2(8):1339–53.

130. Brosch R, Gordon S V, Pym A, Eiglmeier K, Garnier T, Cole ST. Comparative genomics of the mycobacteria. Int J Med Microbiol. 2000;290(2):143–52.

131. Ouellet H, Podust LM, Ortiz De Montellano PR. *Mycobacterium tuberculosis* CYP130: Crystal structure, biophysical characterization, and interactions with antifungal azole drugs. J Biol Chem. 2008;283(8):5069–80.

132. Podust LM, Ouellet H, von Kries JP, Ortiz de Montellano PR. Interaction of *Mycobacterium tuberculosis* CYP130 with heterocyclic arylamines. J Biol Chem. 2009;284(37):25211–9.

133. Provvedi R, Boldrin F, Falciani F, Palu G, Manganelli R. Global transcriptional response to vancomycin in *Mycobacterium tuberculosis*. Microbiology. 2009;155(Pt

4):1093–102.

134. Chenge J, Kavanagh ME, Driscoll MD, McLean KJ, Young DB, Cortes T, Levy CW, Rigby SEJ, Leys D, Abell C, Munro AW. Structural characterization of CYP144A1 – a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*. Sci Rep. 2016;6(26628):1–12.

135. Berry EA, Trumpower BL. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. Anal Biochem. 1987 Feb 15;161(1):1–15.

136. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. Cell Rep. 2013;5(4):1121–31.

137. Bossi RT, Aliverti A, Raimondi D, Fischer F, Zanetti G, Ferrari D, Tahallah N, Maier CS, Heck AJR, Rizzi M, Mattevi A. A covalent modification of NADP+ revealed by the atomic resolution structure of FprA, a *Mycobacterium tuberculosis* oxidoreductase. Biochemistry. 2002;41(28):8807–18.

138. McLean KJ, Scrutton NS, Munro AW. Kinetic, spectroscopic and thermodynamic characterization of the *Mycobacterium tuberculosis* adrenodoxin reductase homologue FprA. Biochem J. 2003;372:317–27.

139. Zanno A, Kwiatkowski N, Vaz ADN, Guardiola-Diaz HM. MT FdR: A ferredoxin reductase from *M. tuberculosis* that couples to MT CYP51. Biochim Biophys Acta - Bioenerg. 2005;1707:157–69.

140. Ahmad Z, Sharma S, Khuller GK. The potential of azole antifungals against latent/persistent tuberculosis. FEMS Microbiol Lett. 2006 May;258(2):200–3.

141. Ahmad Z, Sharma S, Khuller GK, Singh P, Faujdar J, Katoch VM. Antimycobacterial activity of econazole against multidrug-resistant strains of *Mycobacterium tuberculosis*. Int J Antimicrob Agents. 2006;28(6):543–4.

142. Castagnolo D, Radi M, Dessì F, Manetti F, Saddi M, Meleddu R, De Logu A, Botta M. Synthesis and biological evaluation of new enantiomerically pure azole derivatives as inhibitors of *Mycobacterium tuberculosis*. Bioorg Med Chem Lett. 2009;19(8):2203–5.

143. Martis E, Radhakrishnan R. High-throughput screening: the hits and leads of drug discovery-an overview. J Appl Pharm Sci. 2011;1(1):2–10.

144. Armstrong JW. A review of high-throughput screening approaches for drug discovery. Am Biotechnol Lab. 1999;17(5):26–8.

145. Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, Auld DS. High-throughput screening assays for the identification of chemical probes. Nat Chem Biol. 2007;3(8):466–79.

146. Makley LN, Gestwicki JE. Expanding the number of “Druggable” targets: Non-Enzymes and protein-protein interactions. Chem Biol Drug Des. 2013 Jan;81(1):22–32.

147. Makarenkov V, Zentilli P, Kevorkov D, Gagarin A. An efficient method for the detection and elimination of systematic error in high-throughput screening. Bioinformatics. 2007;23(13):1648–57.

148. Coma I, Clark L, Diez E, Harper G. Process validation and screen reproducibility in high-throughput screening. J Biomol Screen. 2009;14(1):66–76.

149. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, Green DVS, Hertzberg RP, Janzen WP, Paslay JW, Schopfer U, Sittampalam GS. Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov. 2011;10(3):188–95.

150. Fox S, Farr-Jones S, Sopchak L, Boggs A, Nicely HW, Khoury R, Biros M. High-throughput screening: Update on practices and success. J Biomol Screen. 2006;11(7):864–9.

151. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov. 2007;6(3):211–9.

152. Barker A, Kettle JG, Nowak T, Pease JE. Expanding medicinal chemistry space. Drug Discov Today. 2013 Mar;18:298–304.

153. Hopkins AL, Groom CR, Alex A. Ligand efficiency: A useful metric for lead selection. Drug Discov Today. 2004;9(10):430–1.

154. Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH. The role of ligand efficiency metrics in drug discovery. Nat Rev Drug Discov. 2014;13(2):105–21.

155. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. Science. 1996;274(5292):1531–4.

156. Hubbard RE, Murray JB. Experiences in fragment-based lead discovery. Methods Enzymol. 2011;493:509–31.

157. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov. 2016;15(9):605–19.

158. Kim D, Guengerich FP. Cytochrome P450 activation of arylamines and heterocyclic amines. Annu Rev Pharmacol Toxicol. 2005 Feb;45(1):27–49.

159. Kavanagh ME, Chenge J, Zoufir A, McLean KJ, Coyne AG, Bender A, Munro AW, Abell C. A fragment profiling approach to inhibitors of the orphan *M. tuberculosis* P450 CYP144A1. Biochemistry. 2017;56(11):1559–72.

160. Mahairas G, Sabo P, Hickey M, Singh D. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol. 1996;178(5):1274–82.

CHAPTER 2

EXPRESSION, PURIFICATION AND CHARACTERISATION OF TRUNCATED CYP141A1, A CYTOCHROME P450 FROM *MYCOBACTERIUM TUBERCULOSIS* H37RV

Nur Dayana Nisbar, Hazel M. Girvan and Andrew W. Munro

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

2.1 Abstract

CYP141A1 is one of the 20 cytochrome P450 enzymes (P450) found in the *Mycobacterium tuberculosis* (Mtb) H37Rv genome. It is hypothesized that CYP141A1 could have important roles in the infectivity or virulence of Mtb, as the *CYP141A1* gene was absent in the avirulent *Mycobacterium bovis* BCG strain. CYP141A1 function in Mtb is still unknown due to the lack of information from its gene location and due to its lack of amino acid sequence similarity to other P450s of defined function. Therefore, it is crucial to have structural insights of CYP141A1 in order to decipher its natural substrate and physiological role. Attempts to obtain a high-resolution crystal structure of CYP141A1 have not been successful to date. Here we describe the production of two new constructs of truncated CYP141A1 in which the first 30 and 54 N-terminal residues of the enzyme were deleted in efforts to improve crystal packing and structural resolution. We successfully expressed, purified and characterized one of the truncated CYP141A1 enzymes, known as Del54CYP141A1. Purified Del54CYP141A1 was highly soluble and retained its typical P450 UV-visible spectroscopy characteristics. However, crystallization of Del54CYP141A1 was unsuccessful. This study provides new data on the properties of the N-terminal Del54CYP141A1 truncation mutant of CYP141A1.

2.2 Introduction

The presence of 20 different genes encoding cytochrome P450 (P450) enzymes in *Mycobacterium tuberculosis* (Mtb) H37Rv genome sequence suggest that these P450s could have novel and important roles in the bacterium. Three Mtb P450s (CYP128A1, CYP125A1 and CYP121A1) were shown to play essential roles in Mtb, as established by nutrient starvation, gene deletion, microarray and genome-wide transposon hybridization studies. (1–4) In addition, characterization of various Mtb P450 enzymes has revealed novel and/or unpredicted functions, including cholesterol metabolism (CYP125A1 (5), CYP124A1 (6) and CYP142A1 (7)) and production of secondary metabolites (CYP121A1 (8)). Revelation of the natural substrates, novel functions and their potential as drug targets for these previously characterized Mtb P450s has motivated further research in this area. Another interesting Mtb P450 for characterization is CYP141A1, which is absent from the genome of the avirulent *Mycobacterium bovis* BCG strain and this could have important roles in the infectivity or virulence of Mtb.

Although various biophysical characterization of CYP141A1 in the solution state were successfully performed, structural characterization of CYP141A1 has proved difficult as crystallization trials of CYP141A1 did not produce high quality crystals diffracting X-rays to appropriate resolution. A common reason for poorly diffracting crystals is that the N- and C-termini of proteins are often flexible and disordered, hence hindering well-ordered crystal packing. In protein crystallization, the typical solution for this problem is to generate a new protein construct with the flexible or disordered ends trimmed off, leaving the stable core of protein. (9,10) N-terminal truncation is uncommon for prokaryotic P450s, but is a routine strategy for structural studies of eukaryotic P450s that are generally membrane-bound proteins, associating with the membrane through an N-terminal anchor sequence. Almost all of the currently available eukaryotic P450 crystal structures have been determined for microsomal P450s expressed and crystallized without the N-terminal transmembrane helix. (11)

Mtb P450s, like other bacterial P450s, lack the N-terminal membrane anchor regions found in eukaryotic P450s and are thus soluble, cytoplasmic enzymes. Consequently, most Mtb P450s can be expressed and purified in substantial amounts to enable spectroscopic and structural studies. (7,12–16) Most of the Mtb P450s studied to date have been structurally characterized either in their ligand-free, substrate-bound or inhibitor-bound forms. In all cases, the wild-type full length gene sequences were used for expression, purification and characterization studies, except for CYP144A1. Expression and purification of full-length CYP144A1 revealed an alternative, shorter form of the enzyme (15) and subsequent Mtb H37Rv transcriptomics analysis identified a second, stronger transcription start site (TSS) (17) for the *CYP144A1* gene. Consequently, a new truncated version of CYP144A1 with the first 30 amino acids at its N-terminal end deleted was generated and characterized. Spectroscopic and structural characterization of both the full-length and truncated forms of CYP144A1 showed that they retained the same properties despite the truncation. (18)

In light of the success in characterizing a truncated form of CYP144A1 and in order to tackle the problem of obtaining a CYP141A1 crystal structure, we investigated the effects of N-terminal truncation on CYP141A1. We set out to produce new truncated forms of CYP141A1 with the ultimate goal of obtaining active and stable CYP141A1 construct(s) amenable to X-ray crystallography. Design of the new truncated forms of CYP141A1 was guided by the aforementioned transcriptomic studies, which highlighted a second TSS in the *CYP141A1* gene, (17) and by bioinformatics analysis identifying a disordered region of CYP141A1. Using this approach, two new CYP141A1 constructs were generated and their encoded proteins characterized. This paper presents the expression, purification and characterization of new truncated forms of CYP141A1, and provides data which shed light on the protein stability and integrity of these variants of CYP141A1.

2.3 Materials and methods

2.3.1 Bioinformatics analysis

Bioinformatics web-servers such as PONDR, (19,20) GlobPlot, (21,22) IUPred (23,24) and Metadisorder (25,26) were used to identify disordered regions of full-length CYP141A1. In addition, PPCpred (27,28) was employed to predict the propensity for truncated CYP141A1 to be expressed and purified, and for it to crystallize successfully. XtalPred-RF, (29,30), XANNpred (31,32) and Crysalis (33,34) were also used to predict the propensity for truncated CYP141A1 to produce diffraction-quality crystals. Alignments of full-length and truncated forms of CYP141A1 were generated using the EMBL-EBI Clustal Omega multiple sequence alignment program. (35,36)

2.3.2 Construction of expression plasmid vectors encoding truncated CYP141A1 genes

Two different N-terminal truncations of CYP141A1 were constructed and named Del30CYP141A1 and Del54CYP141A1, respectively. Design of the truncated CYP141A1 expression plasmid vectors is shown in Figure 2.1. The pET15b plasmid vector was modified to have a Tobacco Etch Virus (TEV) protease recognition site instead of the standard thrombin cleavage site. Consequently, the truncated *CYP141A1* genes were amplified from the full-length CYP141A1 vector using forward and reverse primers, as shown in Table 2.1, and by using CloneAmp^{TM} HiFi PCR Premix (Clontech Laboratories, Inc., USA) with thermal cycling conditions set at 98 °C for 10 s; 55 °C for 10 s; 72 °C for 7 s, repeated for 35 cycles. Following that, the PCR product were inserted into the modified pET15b plasmid vector between *Nde*I and *Bam*HI restriction sites (underlined in Table 1), by using an In-Fusion HD Cloning procedure (Clontech Laboratories, Inc., USA). The resultant expression plasmid vectors were transformed into *E. coli* strain BL21 (DE3) Gold and cells were grown overnight for plasmid extraction. The purified expression plasmid vectors for truncated CYP141A1 were sequenced in order to verify the amino acid sequences of the Del30CYP141A1 and Del54CYP141A1 constructs.

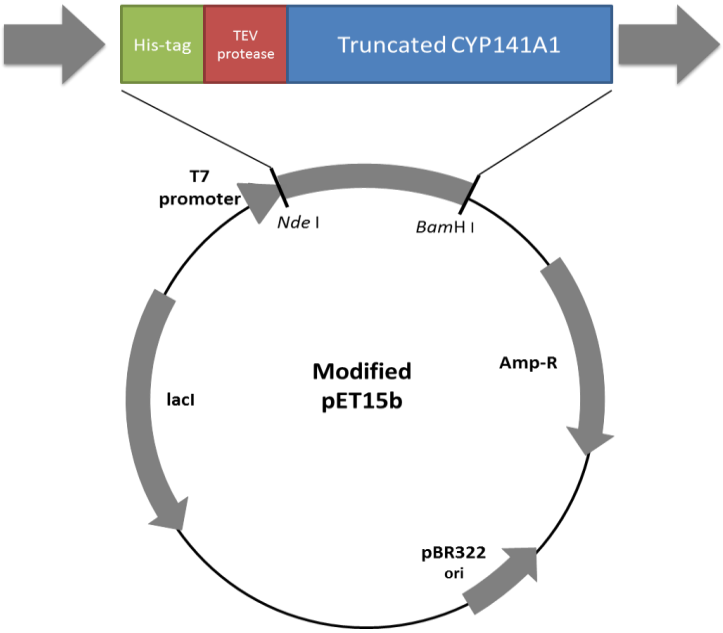


Figure 2.1 Design of the truncated CYP141A1 expression plasmid vector using a modified pET15b plasmid.

Table 2.1 Primer pairs for cloning the truncated CYP141A1 genes.

Construct	Primer pairs (from 5' to 3')	Restriction sites
Del30CYP141A1	FWD: gtatttccaaggccatatggccccgatagagttgc REV: ttagcagccggatcctcacgtcggccaggtaac	*Nde* I *Bam*H I
Del54CYP141A1	FWD: gtatttccaaggccatatgctgtccgacaagcgtttcag REV: ttagcagccggatcctcacgtcggccaggtaac	*Nde* I *Bam*H I

2.3.3 Expression of truncated CYP141A1 proteins

The standard expression protocol for producing full-length CYP141A1 was used for the expression of truncated CYP141A1 enzymes. The expression plasmids containing the truncated forms of CYP141A1 were transformed into *E. coli* BL21 Gold (DE3) cells. A single plasmid-containing colony obtained from the transformation was inoculated into 5 ml of LB medium supplemented with 50 µg/ml carbenicillin and cells were incubated overnight (37 °C, 200 rpm) for culture growth. A further culture was prepared the next day by inoculating 5 ml

of the overnight culture into 300 ml of LB medium supplemented with 50 µg/ml carbenicillin, and the cells were grown at 37 °C with shaking at 200 rpm until the OD_{600} reach 0.5. This culture was used as the inoculum for a larger cell culture growth. Truncated CYP141A1 enzymes were typically expressed in 12 litre cultures of 2xYT medium. Each 2 litre flask contained 500 ml of 2xYT medium which was sterilized and supplemented with 50 µg/ml carbenicillin before inoculating with 10 ml of the starter culture. The cells were grown at 37 °C with 200 rpm agitation until the OD_{600} reached 0.5. The incubation temperature was then reduced to 18 °C and the culture was allowed to equilibrate. When the OD_{600} reached 0.6-0.8, 0.25 mM of 5-aminolevulinic acid hydrochloride (∆ALA) was added as a heme precursor. Subsequently, 0.1 mM IPTG was added to induce gene expression and the culture was incubated for a further 20-24 hours at 18 °C with shaking at 180 rpm. Different expression conditions were also tested using various induction strengths and with growth at 25 °C and at two different ∆ALA concentrations (0.25 mM and 0.5 mM). The cells were harvested by centrifugation using Beckman Coulter Avanti J-25 High Performance centrifuge and JA-10 fixed angle rotor (6000 rpm, 4 °C, 10 min) and the cell pellets were stored at -20 °C until required for protein purification.

2.3.4 Purification of truncated CYP141A1 enzymes

The harvested cell cultures were thawed and resuspended in lysis buffer (50 mM potassium phosphate, 500 mM potassium chloride, 10 mM imidazole, 10% glycerol, pH 8.0) with addition of a protease inhibitor tablet (SIGMAFAST™ Protease Inhibitor Cocktail tablets, EDTA-free, Sigma-Aldrich, UK), 10 µg/ml lysozyme and 10 µg/ml DNase (Sigma-Aldrich, UK). The cells were kept on ice at all times and lysed by intermittent sonication 20-25 times with 20 second bursts at full power, and with 60 seconds cooling time between each burst using a Bandelin Sonopuls GM2600 sonicator. The lysate was then centrifuged using Beckman Coulter Avanti J-26XP centrifuge and JLA-16.250 fixed angle rotor (14500 rpm, 45 minutes, 4 °C) to remove cellular debris as described above. The supernatant was collected and used as a crude CYP141A1 enzyme sample for further purification. Binding of the crude enzyme with nickel affinity resin (Nickel-IDA resin, Generon, United Kingdom) pre-

equilibrated with lysis buffer (50 mM potassium phosphate, 500 mM potassium chloride, 10 mM imidazole, 10% glycerol, pH 8.0) was performed using a batch method. The loaded Ni-IDA column was washed with 20 column volumes of nickel buffer A (50 mM potassium phosphate, 500 mM potassium chloride, 10% glycerol, pH 8.0) containing 20 mM imidazole. The protein was eluted using nickel buffer A with increasing concentrations of imidazole. The eluted protein fractions (red solutions) were pooled and dialysed against two changes of buffer containing 5 litres of buffer Q (50 mM Tris-HCl, 1 mM EDTA, pH 7.5) at 4°C. Subsequently, the dialysed proteins were loaded onto a Q-Sepharose column (Q-Sepharose Fast Flow resin, GE Healthcare Life Sciences, UK) pre-equilibrated with buffer Q. The column was washed with 2 column volumes of buffer Q at 4°C and the protein was eluted with a linear gradient using buffer Q with 0-500 mM KCl. The selected eluted protein fractions were pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 20 MWCO 30000, GE Healthcare Life Sciences, UK) to < 1 ml prior to the final purification step using a gel filtration column (HiLoad 16/600 Superdex 200, GE Healthcare Life Sciences, UK). Fractions of highest purity were again pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 6 MWCO 30000, GE Healthcare Life Sciences, UK) to appropriate protein concentrations for used in protein characterization and crystallization.

2.3.5 Assessment of purity

The purity of the truncated CYP141A1 enzymes were assessed by UV visible absorption spectroscopy (determination of Reinheitszahl (Rz) value - which is the ratio of P450 Soret bound-specific absorption at ~423 nm compared with protein-specific absorption at 280 nm), and by SDS-PAGE (on 12% polyacrylamide denaturing gels) of protein samples at each stage of the purification process. Due to the fact that CYP141A1 is not converted to a ferrous-carbon monoxide complex with an A_{max} at ~450 nm (instead with this feature at 440 nm), the concentration was more accurately determined from the Soret maximum of the oxidized enzyme (in an almost completely low-spin state) using an extinction coefficient of 95 mM^{-1} cm^{-1}. (13) The Beer-Lambert law ($A = \varepsilon.c.l$) was used for the calculation of

cytochrome P450 concentration, where A is the absorbance; ε is the Soret band extinction coefficient; c is the sample concentration and l is the path length of the sample cuvette.

2.3.6 UV-visible spectroscopic characterisation of truncated CYP141A1 enzymes

UV-visible absorption spectra of the oxidized, reduced and reduced/CO-bound forms of the truncated CYP141A1 enzymes were recorded on a Cary 50 UV-visible scanning spectrophotometer (Varian, UK) using a 1 cm pathlength quartz cuvette. Spectra were recorded between 250-800 nm. Typically, purified truncated CYP141A1 enzymes at ~2-10 μM in 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5 at room temperature were used. Binding of carbon monoxide (CO) to the truncated CYP141A1 protein was performed after the reduction of the oxidized protein using a small amount of solid sodium dithionite. 30 to 60 bubbles of CO were added slowly into the ferrous P450 solution inside a fume cupboard prior to recording the CO-bound spectrum.

2.3.7 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were performed using an ÄKTA Purifier HPLC system (GE Healthcare Life Sciences, UK), coupled to a Wyatt DAWN HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX differential refractometer (Wyatt Technology, UK). Chromatographic separation was performed using Superdex-200 10/300 GL size-exclusion column (GE Healthcare Life Sciences, UK) with 100 μl sample loop at a flow rate of 0.75 ml/min, in the selected running buffer. The output was analysed by ASTRA V software (Wyatt Technology, UK). For truncated CYP141A1 enzymes, the value used for dn/dc was 0.183 ml/g. The light scattering system can provide accurate estimates of molar mass for homogeneous samples and can indicate where peaks are composed of mixtures of either different oligomers or conformers, as a well as giving an estimation of polydispersity. SEC-MALS experiments were performed by Ms Hilda Diana Ruiz in the Biomolecular Analysis Facility, University of Manchester.

2.3.8 Crystallization of truncated CYP141A1 proteins

Initial crystallization screening was performed using sparse-matrix kits: PACT *premier*™ HT-96 (Molecular Dimensions, UK), JCSG-*plus*™ HT-96 (Molecular Dimensions, UK) and SG1™ Screen HT-96 (Molecular Dimensions, UK), resulting in 288 different conditions. Sitting drops were dispensed using a robotic platform (mosquito® Crystal, TTP Labtech, UK) by mixing 400 nl protein solution (40 mg/ml truncated CYP141A1 in buffer consisting of 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5) with 400 nl reservoir solution.

2.4 Results

2.4.1 Expression and purification of truncated CYP141A1 enzymes

The N-terminal amino acid sequences removed to generate Del30CYP141A1 and Del54CYP141A1, as well as the final protein sequence lengths and expected molecular weight of the truncated CYP141A1 proteins, are shown in Table 2.2. Typically, overexpression of *M. tuberculosis* P450 enzymes in *E. coli* strains can be initially identified by the appearance of a dark brown/red colour in the harvested cell culture pellets. For Del30CYP141A1 expression, the cell pellets were mainly light brown, suggesting poor protein expression or sub-stoichiometric heme incorporation under the expression conditions tested. Consequently, extraction of the expressed Del30CYP141A1 produced a light brown solution, confirming poor protein expression/heme content. The crude lysate of Del30CYP141A1 expressed under this particular condition (0.5 mM ΔALA, 0.1 mM IPTG, 18 °C) was subjected to nickel affinity chromatography in order to isolate pure protein for enzyme characterization. Del30CYP141A1 protein was eluted from Ni-IDA resin using 100, 200 and 300 mM imidazole concentrations in binding buffer. Analysis by SDS-PAGE (Figure 2.2A) showed that the first elution fraction contained a small amount of Del30CYP141A1 (with a band at approximately 43 kDa), but had a major protein contaminant (seen between the 60-80 kDa marker bands). The following elution fractions using 200 and 300 mM imidazole did not contain this major contaminant. However, the amount of Del30CYP141 was lower (faint bands at approximately 43 kDa). The concentration of Del30CYP141A1 in

the first elution fraction was approximated to be 0.2 μM with an R_z value of approximately 0.25, calculated from the UV-visible spectrum (Figure 2.2B). As the eluted Del30CYP141A1 was low in concentration and purity/heme content, no further work was done on this enzyme.

Table 2.2 N-terminal amino acid sequences removed from intact CYP141A1 to generate Del30CYP141A1 and Del54CYP141A1, including final sequence length and expected molecular weight.
The expected molecular weight is that including the N-terminal 6xHis-tag and TEV protease recognition site.

Protein	Removed N-terminal amino acid sequence	Final sequence length (amino acids)	Expected molecular weight (kDa)
CYP141A1		400	45.9
Del30CYP141A1	MTSTSIPTFPFDRPVPTEPSPMLSELRNSCP	370	42.8
Del54CYP141A1	MTSTSIPTFPFDRPVPTEPSPMLSELRNSCPVAPIELPSGHTAWLVTRFDDVKG	346	40.2

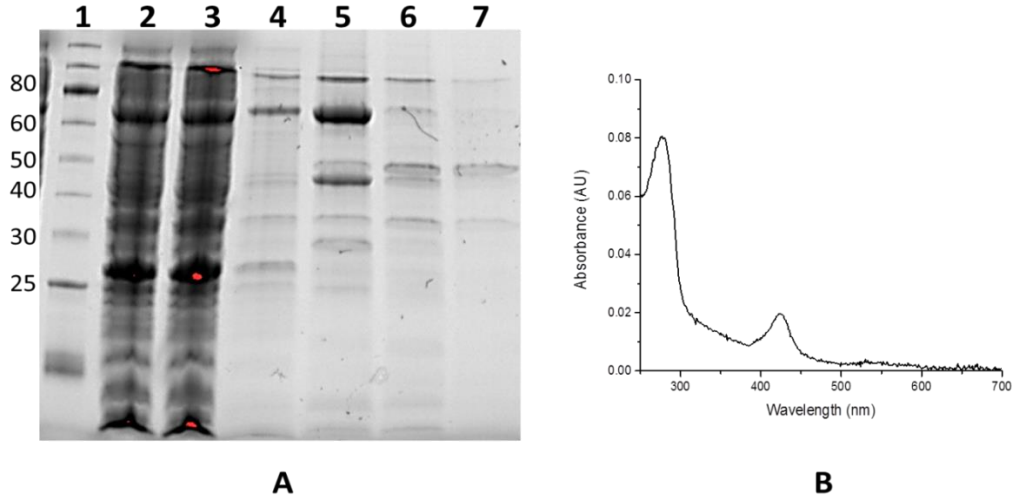


Figure 2.2 Purification of Del30CYP141A1.
(A) SDS-PAGE gel analysis of Del30CYP141A1 nickel affinity chromatography. Lane 1: NEB Unstained Protein Ladder, Broad range marker, bands labelled in kDa; Lane 2: Crude lysate; Lane 3: Flow-through fraction; Lane 4: Wash fraction; Lanes 5-7: Elution fractions using 100, 200 and 300 mM imidazole, respectively. (B) UV-visible absorption spectrum of oxidized form of Del30CYP141A1 with an absorption maximum at 423 nm possibly due to imidazole-bound Del30CYP141A1. The calculated Rz value i.e. (A_{423nm}/A_{280nm}) is approximately 0.25, indicating poor purity/sub-stoichiometric heme content.

In contrast, Del54CYP141A1 was successfully overexpressed in *E. coli* under various growth conditions, as evidenced initially by the appearance of a dark brown/red colour in the harvested cell culture pellets and in the reddish brown crude lysate, indicating soluble, heme-bound P450 enzyme. The crude lysate of Del54CYP141A1 was subjected to nickel

affinity chromatography in order to isolate pure protein for enzyme characterization. Del54CYP141A1 protein starts to elute at 40 mM imidazole and is also found in the fractions containing 100, 200 and 300 mM imidazole. Analysis by SDS-PAGE (Figure 2.3A) showed that each of the elution fractions from 100-300 mM imidazole contain mainly the Del54CYP141A1 protein (with a band at 40 kDa) without any significant protein contaminants. The 40 mM imidazole fraction has lower amounts of Del54CYP141A1 and a contaminant of similar concentration at ~70 kDa. Reasonably pure Del54CYP141A1 was obtained after nickel affinity chromatography in higher imidazole concentrations, as indicated by SDS-PAGE analysis, suggesting that the protein might be amenable to protein crystallization. Therefore, further purification steps were employed to improve protein purity for crystallization attempts. Anion exchange chromatography using Q-Sepharose column appeared to result in some protein aggregation, as shown by high MW bands in the SDS-PAGE gel (Figure 2.3B). Nonetheless, the following gel filtration chromatography step using Superdex-200 column was able to remove the high MW contaminants, producing highly purified Del54CYP141 (Figure 2.3B).

The assessment of the purity of Del54CYP141A1 through comparisons of the R_z values after different stages of purification did not indicate any substantial changes, suggesting that protein purity was not substantially enhanced and that these steps served mainly to remove minor contaminant species. As shown by the UV-visible absorption spectrum of Del54CYP141A1 after each purification step (Figure 2.3C), the Rz values remain fairly similar to each other after successive purification steps. These results likely indicate that Del54CYP141A1 heme incorporation is sub-stoichiometric during protein expression, but that the heme bound is stably retained in the protein. In efforts to improve heme content, different concentrations of the heme precursor ΔALA, were used in further expression trials. As shown in Figure 2.4, protein expression trials using 0.1 mM ΔALA produced the highest R_z value of approximately 0.6. However, the R_z values were similar to each other (between 0.4-0.6) at ΔALA concentrations of 0.1, 0.25 and 0.5 mM, and clear trends in heme content were not evident.

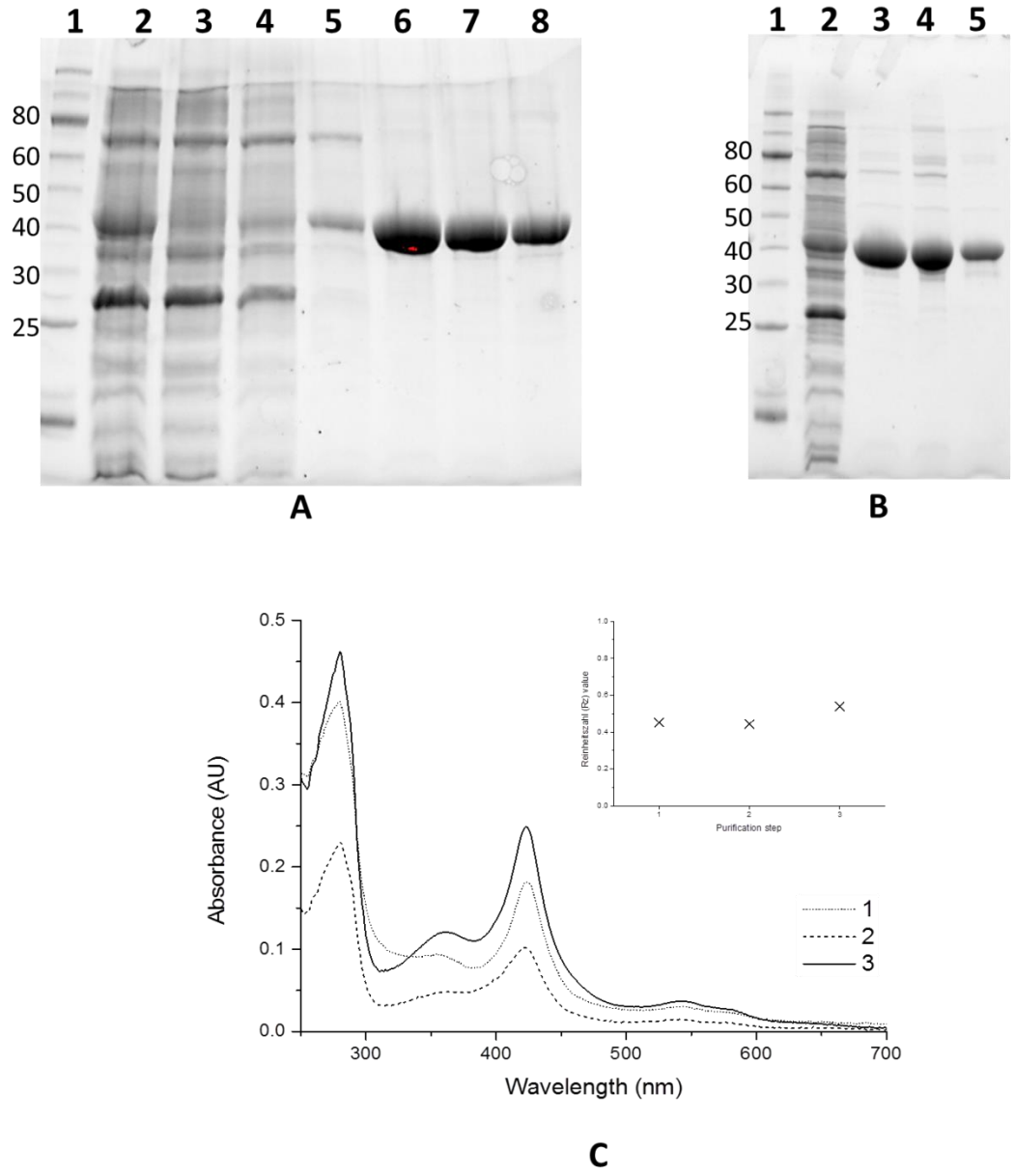


Figure 2.3 Purification of Del54CYP141A1.
(A) SDS-PAGE gel analysis of Del54CYP141A1 after nickel affinity chromatography. Lane 1: NEB Unstained Protein Ladder, Broad range marker, bands labelled in kDa; Lane 2: Crude lysate; Lane 3: Flow-through fraction; Lane 4: Wash fraction; Lanes 5-8: Elution fractions using 40, 100, 200 and 300 mM imidazole, respectively. (B) SDS-PAGE gel analysis of Del54CYP141A1 showing complete (3-step) purification. Lane 1: NEB Unstained Protein Ladder, Broad range marker, bands labelled in kDa; Lane 2: Crude lysate; Lane 3: Nickel affinity chromatography; Lane 4: Anion exchange chromatography (Q-Sepharose column); Lane 5: Gel filtration chromatography (Superdex-200 column). (C) UV-visible absorption spectrum of the oxidized form of Del54CYP141A1 after nickel affinity (dotted line), anion exchange (dashed line) and gel filtration (solid line) chromatography. The absorption maximum is at 424 nm throughout purification possibly indicating that this is the natural Soret peak wavelength for Del54CYP141A1. The inset shows the Rz values for Del54CYP141A1 protein samples each purification step. The values are between 0.4-0.6, indicating incomplete heme incorporation.

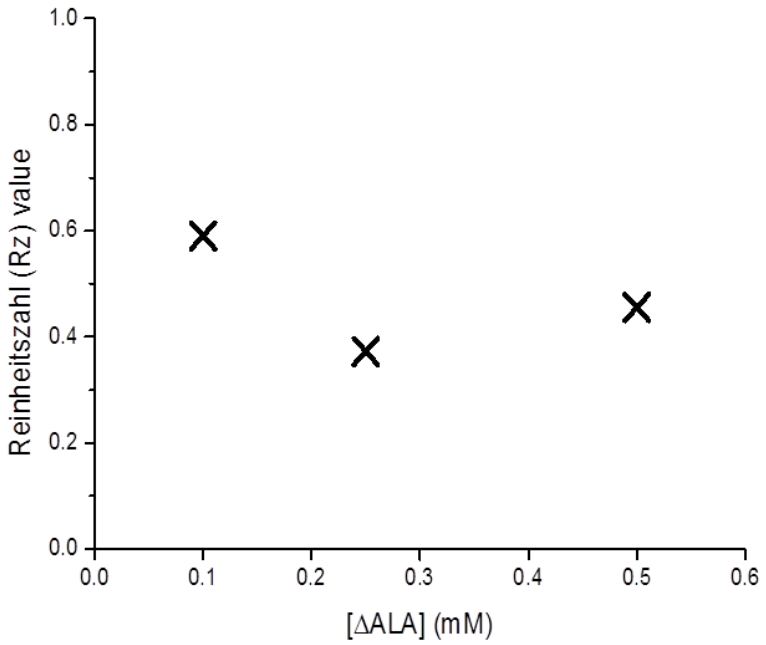


Figure 2.4 Plot of purified Del54CYP141A1 protein Rz values against increasing concentrations of ΔALA used for protein production.
Each Rz value was calculated from the UV-visible absorption spectrum of the oxidized form of pure Del54CYP141A1 proteins produced using different concentrations of ΔALA during expression. Del54CYP141A1 heme content is sub-stoichiometric under all conditions.

2.4.2 Spectroscopic characterisation of truncated CYP141A1

In order to confirm that the truncated CYP141A1 enzyme retains fundamental properties of P450s, the purified enzyme was subjected to reduction and binding to carbon monoxide (CO). The amount of pure Del30CYP141A1 recovered was insufficient and therefore only pure Del54CYP141A1 was further characterized. The spectral properties of Del54CYP141A1 for the oxidized, reduced and CO-bound enzyme forms are shown in Figure 2.5. The oxidized Del54CYP141A1 displays a Soret peak with an absorption maximum at 424 nm and has its heme α and β bands located at 573 and 544 nm, respectively. The Soret maximum is at a much longer wavelength than is typical for the P450s, many of which have their low-spin Soret maxima at ~418 nm. On reduction with sodium dithionite, the Soret peak was maintained at 424 nm but the peak broadens and a single feature is observed at 557 nm in the Q-band region of the spectrum. The CO-bound Del54CYP141A1 exhibits a Soret peak at 424 nm with a shoulder at 440 nm. Under conditions used, the 424 nm feature is predominant, suggesting that the majority of the

enzyme is proximally coordinated by a cysteine in its thiol state, with the minority species (at 440 nm) being cysteine thiolate-coordinated.

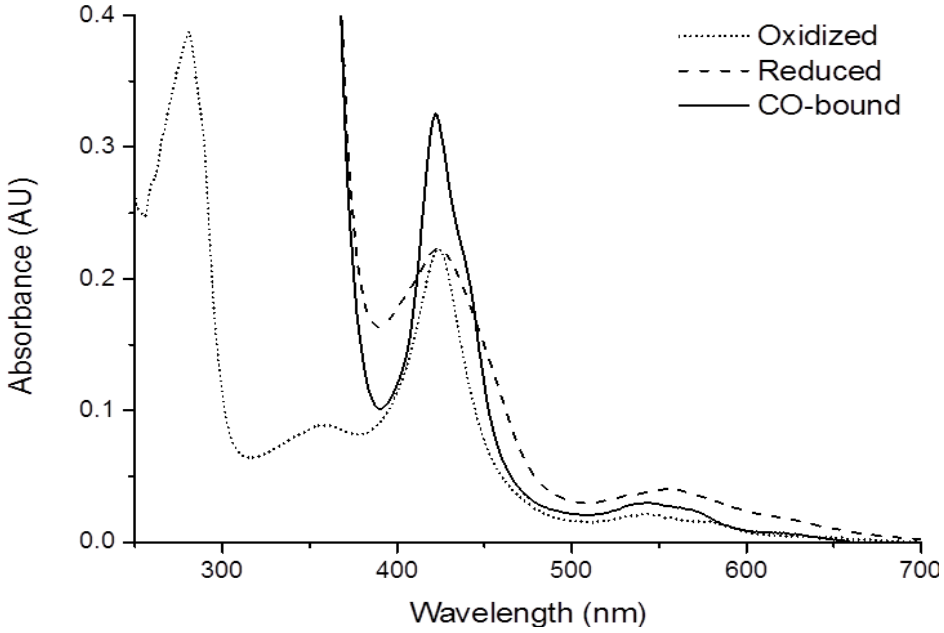


Figure 2.5 UV-visible absorption spectra for Del54CYP141A1.
The UV-visible absorption spectrum for Del54CYP141A1 (~2 μM) is shown in the oxidized form (dotted line), the dithionite-reduced form (dashed line) and the reduced/carbon monoxide-bound form (solid line). The spectra show the oxidized Del54CYP141A1 with an absorption maximum at 424 nm and alpha/beta bands at 573/544 nm, indicating that the ferric heme iron is in a low-spin form. On addition of sodium dithionite, the Soret peak is maintained at 424 nm, with a single Q-band feature at 557 nm. The CO-bound Del54CYP141A1 has a Soret peak maximum at 424 nm, with a shoulder at ~440 nm.

2.4.3 Crystallization trials of Del54CYP141A1

Prior to protein crystallization trials, Del54CYP141A1 was subjected to SEC-MALS analysis in order to determine its oligomeric state, homogeneity and monodispersity. The SEC-MALS chromatogram of the pure Del54CYP141A1 in 50 mM Tris HCl, 0.25 M NaCl, 1 mM DTT, pH 8.5 is shown in Figure 2.6. Del54CYP141A1 eluted as a broad peak with the dominant feature at approximately 15.8 ml. The molar mass of the protein fraction was calculated to be 40.7 kDa, similar to its theoretical mass of 40.2 kDa, and the polydispersity value is 1.001. This polydispersity value indicates that the peak is homogeneous with respect to molar mass. Therefore, the pure Del54CYP141A1 was determined to be monomeric, homogeneous and monodisperse in solution and these properties are preferred

for crystallization. However, crystallization trials of Del54CYP141A1 using a total of 288 different crystallization conditions did not produce any crystals. After 3 weeks, phase separation phenomena were seen in most of the crystallization drops, as shown in Figure 2.7.

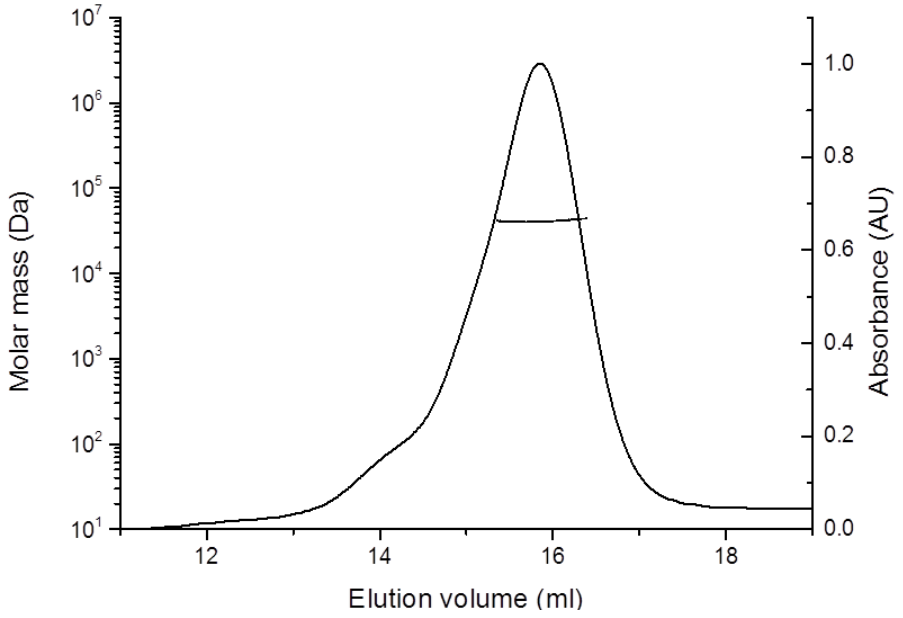


Figure 2.6 SEC-MALS analysis of pure Del54CYP141A1.
The SEC-MALS output displays one major peak at an elution volume of ~15.8 ml, corresponding to a molecular mass of ~40 kDa and indicating that the Del54CYP141A1 sample is predominantly monomeric.

Figure 2.7 Del54CYP141A1 crystallization trial outcome.
Phase separation was observed in most crystallization drops, indicating unsuccessful nucleation and crystallization under conditions used.

2.5 Discussion

N-terminal truncation of the CYP141A1 amino acid sequence was performed with the aim of creating shortened protein constructs amenable to crystallization and obtaining high-quality diffracting crystals. The main motivation for choosing this method was to establish whether success in CYP144A1 truncation (i.e. retention of Del30CYP144A1 spectroscopic characteristics similar to the full length CYP144A1 protein) could be used to produce protein crystals for this species and/or could be applied in engineering studies of other Mtb and bacterial P450s to facilitate their crystallization and structural analysis. (18) The first CYP141A1 truncated form generated was a construct named Del30CYP141A1 following the removal of the first 30 N-terminal residues of CYP141A1, similar to work done previously with CYP144A1. (18) The second construct was Del54CYP141A1 with the first 54 N-terminal residues removed based on Mtb H37Rv transcriptomic studies. (17) In this study, a second, internal, transcriptional start site (TSS) was detected at genome position 3486623, adjacent to a -10 consensus motif (TAGAGT, 3486612-3486617). The transcript generated from this internal TSS would have the potential to initiate translation at residue Val55, hence the removal of the first 54 N-terminal residues in the Del54CYP141A1 construct. Moreover, both constructs remove disordered regions predicted by various bioinformatics tools as shown in Table 2.3 below. The removal of the disordered region was hypothesized to improve protein stability/integrity, and enable better crystal packing which results to the production of protein crystals that diffract to high resolution required for detailed structural analysis. In addition, bioinformatics tools were also employed to predict the probability of the truncated CYP141A1 to be expressed efficiently and to crystallize, as shown in Table 2.4.

Table 2.3 The N-terminal end disordered region of CYP141A1, as predicted by various bioinformatics tools.

Bioinformatics web-server	N-terminal end disordered region
PONDR (19,20)	1-MTSTSIPTFPFDRPVPTEPSPMLSELRNSCPVA-33
GlobPlot (21,22)	1-MTSTSIPTFP FDRPVPTEPS PMLSEL-26
IUPred (23,24)	1-MTSTSIPTFP FDRPVPTEPS -20
Metadisorder (25,26)	1-MTSTSIPTFP FDRPVPT-17

Table 2.4 Outcomes of bioinformatics analyses on the expression and crystallization propensity for truncated CYP141A1 enzymes.
Values in parentheses are the probability that expression or crystallization will occur.

Bioinformatics tool	Full-length	Truncated CYP141A1	
	CYP141A1	Del30CYP141A1	Del54CYP141A1
Expression propensity			
PPCpred (27,28)	Fail (0.459)	Fail (0.381)	Fail (0.367)
Crysalis (33,34)	Fail (0.435)	Fail (0.454)	Fail (0.446)
Crystallization propensity			
PPCpred (27,28)	Fail (0.271)	Success (0.711)	Success (0.707)
XANNpred (31,32)	Fail (0.192)	Fail (0.293)	Fail (0.240)
Crysalis (33,34)	Fail (0.436)	Fail (0.435)	Fail (0.471)

For protein expression, the prediction was consistent with the experimental result for Del30CYP141A1, but not for Del54CYP141. The Del54CYP141A1 construct was expressed as a highly soluble protein that was purified as described in the Methods section. However, despite apparent purity by SDS-PAGE analysis, UV-visible spectroscopy indicated that the purified protein was heme deficient. An explanation for this phenomenon is that heme incorporation into the Del54CYP141A1 protein is not efficient, but once incorporated remains

stably bound. Altering amounts of the heme precursor ΔALA in *E. coli* cells expressing the Del54CYP141A1 enzyme did not result in more extensive heme incorporation, suggesting that poor heme incorporation does not result from inadequate heme supply. In the study of P450s (and other cofactor-binding proteins) the terms for cytochromes P450 lacking a heme cofactor and containing a heme cofactor are apo-cytochrome and holo-cytochrome, respectively. It is likely that the purified Del54CYP141A1 protein is a mixture of both apo- and holo-cytochrome which cannot be differentiated by SDS-PAGE (since the heme is not covalently attached) and thus present as a single band of highly pure protein. This possibility is consistent with the UV-visible absorption spectrum of purified Del54CYP141A1. The protein absorbance (at 280 nm) for purified Del54CYP141A1 is much higher than the absorbance peak specific for its heme cofactor (at 424 nm), confirming its sub-stoichiometric heme content and its composition as a mixture of apo-Del54CYP141A1 and holo-Del54CYP141A1, with the apo-Del54CYP141A1 form being predominant.

The holo-Del54CYP141A1 retains typical P450 spectroscopic characteristics, with the ferric heme iron predominantly in a low-spin state and apparently tethered to the protein by a conserved cysteine thiolate as its proximal axial ligand. In addition, CO-binding to the reduced-Del54CYP141A1 produced a major species with Soret absorption maximum at 424 nm, indicating that the larger proportion of the CO-bound Del54CYP141A1 has the cysteine ligand in a protonated (thiol) form. The minor species, evident as a shoulder with an absorption maximum at 440 nm, appears uncommon among other characterized P450 enzymes. Typically, a ferrous-CO P450 complex will exhibit a Soret peak at approximately 450 nm, although different P450s can have maxima either side of this wavelength. To date, only CYP141A1 and CYP170A1 from *Streptomyces coelicolor* A3(2) have been reported to exhibit a Soret peak at 440 nm in the ferrous-CO state. (37) The structural origins of this unusual spectral feature remain obscure. However, the truncated (Del54CYP141A1) form of CYP141A1 shares the “P440” characteristic of the full length CYP141A1 enzyme and thus the 54 amino acid truncation apparently does not influence the axial ligation state in the Del54CYP141A1 mutant.

In terms of protein crystallization, the prediction by XANNpred and Crysalis bioinformatics tools were consistent with the experimental results obtained, in that the Del54CYP141A1 and Del30CYP141A1 were not successfully crystallized in this study (Table 2.4). However, despite the predictions from the Crysalis and PPCpred programs indicating that these proteins should “fail” in “expression propensity”, both these constructs and the intact CYP141A1 can be expressed in *E. coli* .(38) UV-visible spectroscopic analysis of Del54CYP141A1 shows that both apo- and holo- forms of this protein exist in the purified protein mixture. The presence of both apo- and holo- forms of Del54CYP141A1 in the sample is considered as ‘internal’ heterogeneity which may hinder the packing of the protein to form regular crystals amenable to X-ray diffraction studies. Although SEC-MALS analysis showed that Del54CYP141A1 was predominantly monomeric, homogeneous and monodisperse in solution, the internal heterogeneity of Del54CYP141A1 cannot be detected by this method. However, the unsuccessful crystallization outcome for Del54CYP141A1 highlights the importance of having a homogeneous protein sample for successful crystallization.

2.6 Summary

Studies presented in this paper demonstrate that CYP141A1, one of the larger *M. tuberculosis* P450s, can be produced in forms in which either 30 amino acids (Del30CYP141A1) or 54 amino acids (Del54CYP141A1) are deleted from the N-terminus. While expression levels are poor for the Del30CYP141A1 enzyme, larger amounts of protein are produced for the Del54CYP141A1 protein in *E. coli* in a construct predicted to mimic a form of CYP141A1 predicted to arise from an internal transcriptional start site. The Del54CYP141A1 protein has sub-stoichiometric heme content (*ca* 30-40% incorporation) but heme appears to remain stably bound to the protein once incorporated. The heterogeneous nature of the Del54CYP141A1 protein (as heme-bound and heme-free forms) is prejudicial to successful crystallization of the protein, but other studies confirm that the ligation of the heme iron is through cysteine thiolate in the resting form of this protein, suggesting that the holoprotein form of Del54CYP141A1 can be functional in substrate oxidation. In view of the

inability to crystallize the Del54CYP141A1 protein (consistent with outcomes from predictive algorithms), further studies in this area should involve more detailed analysis of the CYP141A1 protein sequence and design of more constructs that may enhance its propensity to crystallize in order to determine its three-dimensional structure.

2.7 References

1. Betts J, Lukey P, Robb L, McAdam R. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Nature. 2002;104(3):79–83.

2. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol. 2003;48(1):77–84.

3. Kendall SL, Rison SCG, Movahedzadeh F, Frita R, Stoker NG. What do microarrays really tell us about *M. tuberculosis*? Trends Microbiol. 2004 Dec;12(12):537–44.

4. McLean KJ, Carroll P, Lewis DG, Dunford AJ, Seward HE, Neeli R, Cheesman MR, Marsollier L, Douglas P, Smith WE, Rosenkrands I, Cole ST, Leys D, Parish T, Munro AW. Characterization of active site structure in CYP121: A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem. 2008;283(48):33406–16.

5. Capyk JK, Kalscheuer R, Stewart GR, Liu J, Kwon H, Zhao R, Okamoto S, Jacobs WR, Eltis LD, Mohn WW. Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27 steroids. J Biol Chem. 2009;284(51):35534–42.

6. Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR, Montellano PRO De. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(49):20687–92.

7. Driscoll MD, McLean KJ, Levy C, Mast N, Pikuleva IA, Lafite P, Rigby SEJ, Leys D, Munro AW. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: Evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. J Biol Chem. 2010;285(49):38270–82.

8. Belin P, Le Du MH, Fielding A, Lequin O, Jacquet M, Charbonnier J-B, Lecoq A, Thai R, Courçon M, Masson C, Dugave C, Genet R, Pernodet J-L, Gondry M. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(18):7426–31.

9. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. Methods. 2004 Nov;34(3):354–63.

10. Derewenda ZS. Application of protein engineering to enhance crystallizability and improve crystal properties. Acta Crystallogr Sect D Biol Crystallogr. 2010 May 1;66(5):604–15.

11. Johnson EF, Stout CD. Structural diversity of eukaryotic membrane cytochrome P450s. J Biol Chem. 2013;288(24):17082–92.

12. McLean KJ, Marshall KR, Richmond A, Hunter IS, Fowler K, Kieser T, Gurcha SS, Besra GS, Munro AW. Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes. Microbiology. 2002;148(10):2937–49.

13. McLean KJ, Cheesman MR, Rivers SL, Richmond A, Leys D, Chapman SK, Reid GA, Price NC, Kelly SM, Clarkson J, Smith WE, Munro AW. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. J Inorg Biochem. 2002;91:527–41.

14. Ouellet H, Podust LM, Ortiz De Montellano PR. *Mycobacterium tuberculosis* CYP130: Crystal structure, biophysical characterization, and interactions with antifungal azole drugs. J Biol Chem. 2008;283(8):5069–80.

15. Driscoll MD, McLean KJ, Cheesman MR, Jowitt TA, Howard M, Carroll P, Parish T, Munro AW. Expression and characterization of *Mycobacterium tuberculosis* CYP144: Common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. Biochim Biophys Acta - Proteins Proteomics. 2011;1814(1):76–87.

16. Chenge JT, Le D Van, Swami S, McLean KJ, Kavanagh ME, Coyne AG, Rigby SEJ, Cheesman MR, Girvan HM, Levy CW, Rupp B, von Kries JP, Abell C, Leys D, Munro AW. Structural characterization and ligand/inhibitor identification provide functional insights into the *Mycobacterium tuberculosis* cytochrome P450 CYP126A1. J Biol Chem. 2016;292(1):1310–29.

17. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. Cell Rep. 2013;5(4):1121–31.

18. Chenge J, Kavanagh ME, Driscoll MD, McLean KJ, Young DB, Cortes T, Levy CW, Rigby SEJ, Leys D, Abell C, Munro AW. Structural characterization of CYP144A1 – a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*. Sci Rep. 2016;6(26628):1–12.

19. Garner E, Romero P, Dunker A, Brown C, Obradovic Z, Li X, Rani M, Uversky V, Gillespie J, Fink A, Vucetic S, Radivojac P, Peng K. PONDR®: Predictor of Natural Disordered Regions [Internet]. Molecular Kinetics, Inc. 2003, [updated 2007 Jan 12; cited 2017 May 20]. Available from: http://www.pondr.com/

20. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins Struct Funct Genet. 2003;53:566–72.

21. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot 2: Intrinsic protein disorder, domain and globularity prediction Version 2.3 [Internet]. EMBL, [updated 2006 Aug 28; cited 2017 May 20]. Available from: http://globplot.embl.de/

22. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003 Jul 1;31(13):3701–8.

23. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: Prediction of intrinsically unstructured proteins [Internet]. Institute of Enzymology, Hungarian Academy of Sciences, [updated 2005 Aug 15; cited 2017 May 20]. Available from: http://iupred.enzim.hu/

24. Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005 Aug 15;21(16):3433–4.

25. Kozlowski LP. GeneSilico Metadisorder service: Prediction of intrinsically unstructured proteins (protein disorder) from amino acid sequence only [Internet]. International Institute of Molecular and Cell Biology, Warsaw, [updated 2012 May 24; cited 2017 May 20]. Available from: http://iimcb.genesilico.pl/metadisorder/predict_protein_disorder_by_metadisorder.html

26. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics. 2012;13(1):111.

27. Mizianty MJ, Kurgan L. PPCpred webserver [Internet]. Virginia Commonwealth

University, [updated 2011 Jul 1; cited 2017 Jun 1]. Available from: http://biomine.cs.vcu.edu/servers/PPCpred/

28. Mizianty MJ, Kurgan L. Sequence-based prediction of protein crystallization, purification and production propensity. Bioinformatics. 2011 Jul 1;27(13):i24–33.

29. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred-RF: Prediction of protein crystallizability [Internet]. Sanford Burnham Prebys Medical Discovery Institute, [updated 2015 May 15; cited 2017 Jun 1]. Available from: http://ffas.burnham.org/XtalPred-cgi/xtal.pl

30. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. Bioinformatics. 2007 Dec 15;23(24):3403–5.

31. Overton IM, van Niekerk CAJ, Barton GJ. XANNpred: SSPF crystallization propensity predictors [Internet]. The Barton group, University of Dundee, [updated 2011 Apr; cited 2017 Jun 1]. Available from: http://www.compbio.dundee.ac.uk/xtal/cgi-bin/xannpred_in.pl

32. Overton IM, van Niekerk CAJ, Barton GJ. XANNpred: Neural nets that predict the propensity of a protein to yield diffraction-quality crystals. Proteins Struct Funct Bioinforma. 2011 Apr;79(4):1027–33.

33. Wang H, Feng L, Zhang Z, Webb GI, Lin D, Song J. Crysalis: an integrated server for computational analysis and design of protein crystallization [Internet]. College of Chemistry And Chemical Engineering, Xiamen University and Faculty of Medicine, Monash University,[updated 2016 Feb 24; cited 2017 Jun 1]. Available from: http://nmrcen.xmu.edu.cn/crysalis/

34. Wang H, Feng L, Zhang Z, Webb GI, Lin D, Song J. Crysalis: an integrated server for computational analysis and design of protein crystallization. Sci Rep. 2016 Feb 24;6(1):21383.

35. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, Mcwilliam H, Remmert M, Sö Ding J, Thompson JD, Higgins DG. Clustal Omega Multiple Sequence Alignment EMBL-EBI [Internet]. EMBL-EBI, [updated 2011 Oct 11; cited 2017 Jun 1]. Available from: http://www.ebi.ac.uk/Tools/msa/clustalo/

36. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, Mcwilliam H, Remmert M, Sö Ding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7(539).

37. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, Cane DE. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem. 2008 Mar 28;283(13):8183–9.

38. Le Van D. Analysis of novel drug target enzymes in *Mycobacterium tuberculosis*. 2011.

CHAPTER 3

CRYSTALLIZATION OF CYP141A1, AN ORPHAN CYTOCHROME P450 FROM *MYCOBACTERIUM TUBERCULOSIS* H37RV

Nur Dayana Nisbar[1], Colin W. Levy[2], Kirsty J. Mclean[1] and Andrew W. Munro[1]

[1]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom,

[2]Manchester Protein Structure Facility, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

3.1 Abstract

The CYP141A1 P450 encoded by the *Rv3121* gene from the pathogenic bacterium *Mycobacterium tuberculosis* H37Rv is absent from the genome of the vaccine strain *Mycobacterium bovis* BCG and this suggests that it may be relevant for *M. tuberculosis* virulence and infectivity towards the human host. A CYP141A1 three-dimensional structure will unveil its active site architecture, which may allow this P450 to be exploited as novel therapeutic target. CYP141A1 was cloned and expressed in *Escherichia coli*, and the protein was purified. Pure, oxidized CYP141A1 exhibits an absorbance spectrum indicative of a low spin ferric heme iron ligated by cysteine thiolate. Interestingly, the reduced/carbon monoxide-bound form displays a Soret peak at 440 nm rather than at ~450 nm, making this an atypical P450 enzyme. Size exclusion chromatography coupled to multi-angle light scattering revealed that the purified CYP141A1 was not homogeneous, hence hampering protein crystallization. However, the optimization of protein buffer with addition of reducing agent as well as adjustment of buffer ionic strength and pH was found to increase the proportion of monomeric CYP141A1, making the sample more homogeneous and monodisperse. Applying the optimized CYP141A1 buffer solution to size exclusion chromatography in the final step of purification resulted in the isolation of predominantly

homogeneous, monomeric CYP141A1. Consequently, protein crystallization trials using this highly pure, monodisperse CYP141A1 generated crystals reproducibly. The best crystal diffracted to 2.7 Å resolution and belonged to space group $P6_122$ with unit-cell parameters of a=b= 95.7 Å, c= 265.26 Å and α=β= 90°, γ= 120°. A weak anomalous signal arising from the intrinsic iron of the heme was detected during data collection which generated the phase for the diffraction data via experimental phasing method and consequently produced an initial model structure of CYP141A1.

3.2 Introduction

According to the World Health Organization (WHO), tuberculosis caused by the pathogen *Mycobacterium tuberculosis* (Mtb) is one of the top 10 causes of death worldwide. Consequently, ending the tuberculosis epidemic is among the health targets of the newly adopted Sustainable Development Goals. (1) Due to the emergence of drug resistant strains of *M. tuberculosis*, there is an urgent need for the development of novel anti-TB therapeutics. The identification of 20 cytochrome P450 enzymes in the *M. tuberculosis* H37Rv genome (2) indicates their likely important functions in the pathogen. (3) Cytochrome P450s are monooxygenases that catalyze the reductive scission of molecular oxygen which is bound to the iron atom of the heme cofactor located in their active site. This leads to insertion of an atom of oxygen into the substrate and production of a molecule of water from the other oxygen atom. One of the twenty Mtb P450 enzymes, CYP141A1, could have important roles in infectivity or virulence of *M. tuberculosis,* (4) as the *Rv3121* gene encoding CYP141A1 is absent from the genome of the vaccine strain *Mycobacterium bovis* BCG, and is located in the RD12 (Region of Difference) segment of the genome which is deleted from this bacterium. (5) In view of CYP141A1's potential as a novel drug target for combating TB, characterization of this enzyme was prioritized. Structural characterization of CYP141A1, in particular, could provide a framework on which to base the rational design of new therapeutic agents to treat tuberculosis.

X-ray crystallography is the most powerful method for determining three-dimensional structures of proteins, but obtaining diffraction-quality protein crystals is often the main rate-

limiting step in the process. A typical protein crystallization screening process usually employ sparse-matrix screens containing a set of limited crystallization conditions with varying salt, buffer, pH, precipitants and additives. (6) In addition to the chemical components of the crystal screens, the components of the protein buffer solution can also affect the success of a protein crystallization screening process. A protein can have complex properties and its behaviour depends on the buffer that provides the local solvent environment for the protein. The solvent environment of the protein critically controls protein homogeneity, stability and solubility, which are all factors expected to be highly correlated with the probability of a protein forming useful crystals. Generally, a standard purification buffer is used for preparing protein sample for crystallization. However, the chosen buffer condition may not necessarily be the best condition for crystallization. As a consequence, altering the composition of the buffer used during purification is often necessary, especially for proteins that are prone to aggregation. The buffer screening process thus involves selecting a superior buffer to that previously used during purification and testing several different additives in order to improve production of the monomeric form of the protein sample. Screening for buffer conditions that optimally maintain homogeneity of proteins in solution, and which reduce their aggregation and polydispersity can be done using various quality-control techniques. These include light scattering, mass spectrometry, differential scanning fluorimetry and analytical ultracentrifugation. (7) As a result, the optimization of the buffer solution is expected to improve protein crystallization outcome. The process of protein crystallization consists of different stages. Typically, the first stage is identifying crystallization condition that can generate reproducible crystals appropriate for X-ray data collection, followed by optimization of the crystallization conditions in order to produce crystals suitable for structure determination. Ideally, the production of a large and visually perfect crystal is preferable for the downstream X-ray crystallography process.

In this paper, we report on the protein crystallization strategy adopted for CYP141A1, which consists of three steps: (1) buffer screening and protein quality assessment, (2) sparse-matrix screening, and (3) improvement of crystal growth. This manuscript also presents the strategies used to collect X-ray diffraction data and to solve the crystallographic phase problem for CYP141A1. As a result, the elucidation of a new crystal structure of a *M.*

tuberculosis P450 enzyme was obtained. The availability of the CYP141A1 crystal structure allows the determination of its structural properties and would facilitate the application of a structure-based drug discovery approach.

3.3 Materials and methods

3.3.1 Molecular biology

The *Rv3121* gene encoding CYP141A1, a cytochrome P450 enzyme from *Mycobacterium tuberculosis*, was amplified from the cosmid (MTCY339) containing DNA from the *M. tuberculosis* H37Rv strain and provided by Professor Stewart Cole (EPFL, Lausanne, Switzerland). The primers used in the reaction were designed from the genomic sequence (2) and include an engineered restriction site for *Nde*I and the initiation codon ATG in the forward primer, and a *Bam*HI site and the stop codon TAG in the reverse primer. The primers are 5'-GCGGCAGCCAT**ATG**ACTACCGCCG-3' for forward primer and 5'-GCGCGGAGGC**TAG**GATCCGGCTGC-3' for reverse primer. The underlined letters in the forward and reverse primers indicate engineered *Nde*I and *Bam*HI restriction endonucleases sites, respectively. The bold letters indicate the start (ATG) and stop (TAG) codons, respectively. The gene amplification conditions used were 95 °C for 2 min; 30 cycles of 95 °C for 45 s, 62 °C for 30 s, and 72 °C for 1.5 min; and then a final polymerisation step of 72 °C for 10 min. The *Rv3121* PCR fragment was cloned into the pGEM-T vector (Promega, Madison WI, USA) and for protein expression this plasmid clone was digested with NdeI and BamHI and the *Rv3121* fragment was ligated into vector pET15b (Merck Ltd, Beeston, UK), pre-digested with the same restriction endonucleases. The resulting plasmid clone was transformed into *E. coli* BL21 (DE3) Gold cells and bacterial cultures grown overnight for plasmid extraction. The purified plasmid was sequenced in entirety in order to verify the amino acid sequence of CYP141A1 by Sanger sequencing (Source Bioscience, UK). This *CYP141A1* expression plasmid vector (pCYP141A1) was used for all protein expression/purification studies reported in this study.

3.3.2 Expression and purification of CYP141A1

The expression plasmid pCYP141A1 was transformed into *E. coli* strain BL21 Gold (DE3) to enable CYP141A1 protein production. A single plasmid-containing colony obtained from the transformation was inoculated into 5 ml of LB medium supplemented with 50 µg/ml carbenicillin and incubated overnight (37 °C, 200 rpm). A sub-culture was prepared the next day by inoculating 5 ml of the overnight culture into 300 ml of LB medium supplemented with 50 µg/ml carbenicillin, and the cells were grown at 37 °C, 200 rpm until the OD_{600} reach 0.5. This culture was used as the inoculum for a larger cell culture growth. CYP141A1 was typically expressed in 12 litre cultures of 2xYT medium. Each 2 litre flask containing 500 ml of 2xYT medium was sterilized and supplemented with 50 µg/ml carbenicillin before inoculating with 10 ml of the starter culture. The cells were grown at 37 °C with 200 rpm agitation until the OD_{600} reached 0.5. The incubation temperature was then reduced to 18 °C and the culture was allowed to equilibrate. When the OD_{600} reached 0.6-0.8, 0.25 mM of 5-aminolevulinic acid hydrochloride was added as a heme synthesis precursor. Subsequently, 0.1 mM IPTG was added to induce *CYP141A1* protein expression and the culture was incubated for 20-24 hours at 18 °C, 180 rpm. The cells were harvested by centrifugation using Beckman Coulter Avanti J-25 High Performance centrifuge and JA-10 fixed angle rotor (6000 rpm, 4 °C, 10 min) and the cell pellets were stored at -20 °C until lysed.

For CYP141A1 purification, the cells were thawed and resuspended in lysis buffer (50 mM potassium phosphate, 500 mM potassium chloride, 10 mM imidazole, 10% glycerol, pH 8.0) with addition of one protease inhibitor tablet per 100 ml of cell suspension (SIGMAFAST™ Protease Inhibitor Cocktail tablets, EDTA-free, Sigma-Aldrich, UK), 10 µg/ml lysozyme and 10 µg/ml DNase (Sigma-Aldrich, UK). The cells were kept on ice at all times and lysed by intermittent sonication 20-25 times with 20 second bursts at full power, and with 60 seconds cooling time between each burst using a Bandelin Sonopuls GM2600 sonicator. The lysate was then centrifuged (14500 rpm, 45 minutes, 4 °C) to remove cellular debris using Beckman Coulter Avanti J-26XP centrifuge and JLA-16.250 fixed angle rotor. The supernatant was collected and used as a crude CYP141A1 enzyme sample for further purification. The crude enzyme was mixed with nickel affinity resin (Nickel-IDA resin,

Generon, UK) pre-equilibrated with lysis buffer (50 mM potassium phosphate, 500 mM potassium chloride, 10 mM imidazole, 10% glycerol, pH 8.0) to facilitate binding to the resin. The CYP141A1-loaded resin was transferred to a column and washed with 20 column volumes of nickel buffer A (50 mM potassium phosphate, 500 mM potassium chloride, 10% glycerol, pH 8.0) containing 20 mM imidazole. The protein was then eluted using nickel buffer A with 200 mM imidazole. The eluted protein (containing the red CYP141A1 protein) was pooled and dialysed against two changes of buffer containing 5 litres of buffer Q (50 mM Tris-HCl, 1 mM EDTA, pH 7.5). Subsequently, the dialysed protein was loaded onto a Q-Sepharose column (Q-Sepharose Fast Flow resin, GE Healthcare Life Sciences, UK) pre-equilibrated with buffer Q. The column was washed with 2 column volumes of buffer Q and the protein was eluted using a linear gradient of buffer Q with 0-500 mM KCl. The selected eluted protein fractions were again dialysed against two changes of buffer containing 5 litres of buffer HA (25 mM potassium phosphate, pH 6.5). Following this step, the dialysed protein was loaded onto a hydroxyapatite column (Hydroxyapatite Type 1 resin, Sigma-Aldrich, UK) pre-equilibrated with buffer HA. The column was washed with 2 column volumes of buffer HA and the protein was eluted using buffer HB (500 mM potassium phosphate, pH 6.5). The selected eluted protein fractions were pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 20, MWCO 30000, GE Healthcare Life Sciences, UK) to ~1 ml prior to a final purification step using a gel filtration column (HiLoad 16/600 Superdex 200, GE Healthcare Life Sciences, UK) pre-equilibrated with selected buffer during buffer screening (10-50 mM potassium phosphate/ HEPES/ Tris-HCl, 0-500 mM KCl/ NaCl, 0-0.5 mM TCEP/ DTT). Fractions of highest purity were again pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 6, MWCO 30000, GE Healthcare Life Sciences, UK) to an appropriate protein concentration prior to protein crystallization.

3.3.3 Protein identification by trypsin digestion and mass spectrometry

Coomassie stained protein bands were excised from acrylamide gels. The gel slices were then diced into cubes of approximately 1 mm^3. The gel cubes were then hydrated

before the stain was extracted with alternate hydration and dehydration steps in 50% acetonitrile and 50 mM ammonium bicarbonate (dehydration/hydration/dehydration). Following the final dehydration step 20 µl of mass spectrometry grade trypsin (Promega or Sigma, UK), diluted in 50 mM ammonium bicarbonate as directed by the manufacturer was added to the gel cubes. The digestion was performed at 37 ^{o}C overnight. Following manual removal of the liquid from the digest, 40 µl of water was added to the gel cubes to obtain the maximum yield of the tryptic peptides. The combined extracts were centrifuged in a microfuge at 13000 rpm for 10 mins to pellet any particulate matter. 35 µl of the supernatant was placed in vials on the inline HPLC and 20 µl of this sample was used for each run.

Tryptic peptides were subjected to LC/MS/MS, separated on a nano-flow HPLC (Dionex, Thermo Fisher Scientific, UK) using a C18 column (LC Packings, Acclaim Pep Map 100 - a C-18 pre-column is also used). A nano-spray injection (fused silica emitter) was employed to introduce samples into a Brüker Esquire 3000plus ion trap mass spectrometer. The spectrometer was used in positive ion mode, the flow rate was 200 nl/min, 'Picotip' fused silica emitters (needles) with a 360 µm diameter and 10 µl tip diameter (Presearch Ltd FS360-20-10-D-20) were used for the nano-spray ionisation. The mass spectrometry data were then subjected to protein database searches using the Mascot MS/MS ion search software (Matrix Sciences, USA). Protein identification experiments were performed by Dr. Martin Read from Manchester Institute of Biotechnology Protein Identification services.

3.3.4 Assessment of CYP141A1 purity

The purity of CYP141A1 was assessed by using UV visible absorption spectroscopy, through the determination of its Reinheitszahl (Rz) value (the ratio of the low-spin ferric P450-specific absorption maximum at 420 nm compared to the protein-specific absorption at 280 nm), and by SDS-PAGE (on a 12% polyacrylamide denaturing gel) of protein samples taken at each stage of the CYP141A1 purification process. Due to the fact that CYP141A1 is not converted to a ferrous-carbon monoxide complex with an A_{max} at ~450 nm (instead with this feature at 440 nm), the concentration was more accurately determined from the Soret maximum of the oxidized enzyme (in an almost completely low-spin state) using an

extinction coefficient of 95 mM^{-1} cm^{-1}. (8) The Beer-Lambert law (A = ε.c.l) was used for the calculation of cytochrome P450 concentration, where A is the absorbance; ε is the Soret band extinction coefficient; c is the sample concentration and l is the path length of the sample cuvette.

3.3.5 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were performed using ÄKTA Purifier HPLC system (GE Healthcare Life Sciences, UK), coupled to a Wyatt DAWN HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX differential refractometer (Wyatt Technology, UK). Chromatographic separation was performed using Superdex-200 10/300 GL size-exclusion column (GE Healthcare Life Sciences, UK) with a 100 µl sample loop at a flow rate of 0.75 ml/min, in the selected running buffer (10-50 mM potassium phosphate/HEPES/Tris-HCl, 0-500 mM KCl/NaCl, 0-0.5 mM TCEP/DTT). The output was analysed by ASTRA V software (Wyatt Technology, UK). For CYP141A1, the value used for dn/dc was 0.183 ml/g. The light scattering system provides accurate estimates of molar mass for homogeneous samples and indicates where peaks are composed of mixtures of either different oligomers or conformers, as a well as providing estimation of polydispersity. SEC-MALS experiments were performed by Ms Hilda Diana Ruiz from the Biomolecular Analysis Facility (University of Manchester).

3.3.6 Crystallization of CYP141A1

Initial crystallization trials were performed using 96-well sitting drop crystallization plates containing different crystallization conditions in each well (also referred to as reservoir solution). Four different plates were used, with each plate containing 96 different crystallization condition from sparse-matrix screens PACT *premier*™ HT-96, JCSG-*plus*™ HT-96 and SG1™ Screen HT-96 (Molecular Dimensions, UK). Crystallization drops were dispensed using a robotic platform (mosquito® Crystal, TTP Labtech, UK) by mixing 300 nl CYP141A1 (40 mg/ml in buffer consisting of 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH

8.5) with a 300 nl reservoir solution. Initial CYP141A1 crystallization trials gave two positive hits in C10 and F4 drop of PACT *premier*™ HT-96 plate. Red crystals of CYP141A1 appeared after three days and continued to grow after 2 weeks. Improvement of CYP141A1 crystal size and quality was performed by increasing the size of crystallization drops and by grid screens of the initial crystallization conditions. Crystals of an appropriate size (50-200 µm) were selected for X-ray diffraction data collection.

3.3.7 X-ray diffraction data collection

A single red crystal of CYP141A1 was mounted on a 0.1-0.2 mm 20 micron Mounted CryoLoop™ (Hampton Research, USA), dipped briefly in glycerol as cryoprotectant and frozen in liquid nitrogen prior to diffraction studies. X-ray diffraction data were collected at the Diamond synchrotron beamlines I03 and I04-1 (Harwell UK). The data were indexed and processed with the *XDS* program package. (9)

3.4 Results

3.4.1 Expression of CYP141A1

By using conditions with bacterial transformant growth temperature at 37 °C, 200 rpm agitation and induction of gene expression with 1 mM IPTG in late log phase, CYP141A1 could be expressed at high levels under the control of a T7 RNA polymerase promoter system. However, cell lysis using sonication revealed that most of the CYP141A1 expressed was located in inclusion bodies and that only a small amount was found in a soluble form. Therefore, the growth conditions of the host strain were optimized by varying incubation temperature, induction time and IPTG concentration. As a result, improved expression conditions for producing higher levels of soluble CYP141A1 were established with induction during the mid-logarithmic phase of growth (OD_{600} = 0.6-0.8) using 0.1 mM IPTG at 18 °C, and by continuing growth for approximately 20-24 hours post induction. Attempts to increase CYP141A1 expression by extending the growth period up to 48 hours did not improve protein recovery, but instead caused increased expression of *E. coli* contaminant proteins.

Mass spectrometry analysis (Figure 3.1) identified these contaminants as glucosamine-6-phosphate aminotransferase (GlmS) and catalase (katE) which are stress-responsive proteins commonly expressed in *E. coli* host strains (10) These contaminants caused issues with isolation of pure CYP141A1, as they proved difficult to separate from the P450 enzyme. Therefore, CYP141A1 expression time was maintained at 20-24 hours post IPTG induction. Further improvement to the expression of CYP141A1 was achieved by the addition of 0.25 mM ΔALA as a heme precursor before IPTG induction.

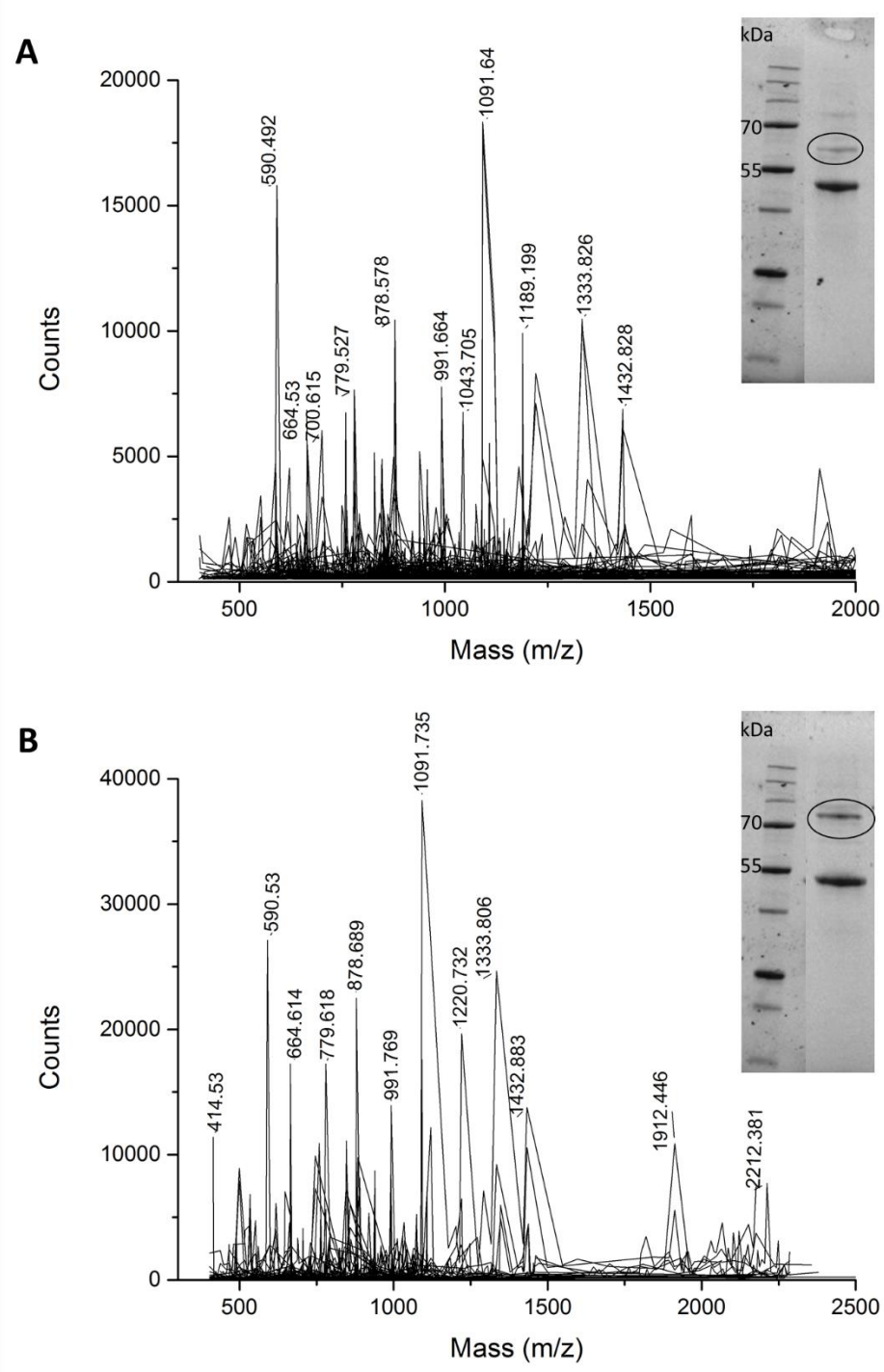


Figure 3.1 Mass spectrometry analysis of CYP141A1.
LC/MS/MS spectrograms of the protein of interest which resulted in the identification of *E. coli* contaminant proteins, (A) glucosamine-6-phosphate aminotransferase (GlmS) enzyme and (B) catalase (katE) enzyme, using the Mascot MS/MS ion search software (Matrix Sciences, USA). The insets show SDS-PAGE gels with the protein bands of interest (circled in black). The protein bands were excised and digested using mass spectrometry grade trypsin before subjected to LC/MS/MS for identification.

3.4.2 Purification of CYP141A1

The purity of CYP141A1 was examined after each purification step, as described in Section 3.3.2. A successful CYP141A1 purification strategy was established using column chromatography steps with Nickel-IDA, Q-Sepharose and Hydroxyapatite resins. A final polishing step using gel filtration with Superdex 200 column chromatography was added to produce highly purified CYP141A1 protein samples required for crystallization trials. The ratio of heme Soret-specific absorbance ($A_{\sim 420nm}$) to that of the total protein absorption (A_{280nm}), the Rz value, was measured after each step in the purification process, and the Rz value was seen to increase at each stage. The Rz value after the final step was usually between 1.8 and 2.0, indicating highly purified CYP141A1, as verified by SDS-PAGE. Figure 3.2 shows an SDS-PAGE gel demonstrating purity of CYP141A1 samples. The apparent molecular mass of CYP141A1, based on comparisons with protein standard MW markers, is approximately 46 kDa, consistent with the predicted mass from amino acid sequence of 45.9 kDa, including the N-terminal 6x-His-tag region.

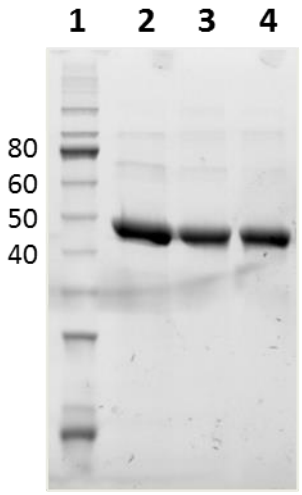


Figure 3.2 Purification of *M. tuberculosis* CYP141A1.
The SDS-PAGE gel (12 % polyacrylamide) shows purification of CYP141A1 expressed in *E. coli* BL21 (DE3) Gold cells. Lane 1: NEB Unstained Protein Ladder, Broad range marker, with bands labelled in kDa; Lane 2: Nickel-IDA elution fraction; Lane 3: Post-Q Sepharose fraction; Lane 4: Post-Hydroxyapatite fraction.

3.4.3 Optimization of buffer solution compatible with a monomeric CYP141A1 state

A common problem for purified proteins intended for structural studies by X-ray crystallography is obtaining conformationally homogeneous and monodisperse protein samples. Prior to CYP141A1 crystallization, the purified protein was concentrated and the presence of aggregates or oligomers was assessed using size exclusion chromatography with multi-angle light scattering (SEC-MALS). Consequently, optimisation of the CYP141A1 buffer solution was performed based on SEC-MALS analyses of samples. This involves screening for buffers than that previously used during purification and by testing different additives in order to enhance production of a monomeric form of the CYP141A1 protein sample.

Initially, the aggregation/oligomerisation state of CYP141A1 was analysed in different types of buffers (potassium phosphate, HEPES and Tris) at pH 7.5. Chromatograms of the purified CYP141A1 in these buffers are shown in Figure 3.3. In 100 mM potassium phosphate, pH 7.5 (black line), the CYP141A1 sample eluted broadly with two peaks at 11.9 ml and 13.3 ml, and the molar mass of the protein fraction was calculated to be between 90-200 kDa. In 10 mM HEPES, 0.5 M KCl, pH 7.5 (red line), the CYP141A1 sample eluted earlier with a single peak at 10.5 ml and the molar mass of the protein fraction calculated to be approximately 500 kDa. In 10 mM Tris HCl, 0.25 M KCl, pH 7.5 (blue line), the CYP141A1 sample eluted with a broad shoulder followed by one narrow peak at 15.2 ml and the molar mass for the protein fraction of the narrow peak was calculated to be between 45.3 kDa. The initial buffer screening showed that potassium phosphate and HEPES buffers caused CYP141A1 to aggregate and form higher oligomeric states. Therefore, the best buffer system for CYP141A1 identified at this stage was 10 mM Tris HCl, 0.25 M KCl, pH 7.5, as the protein is predominantly in the monomeric state in this buffer.

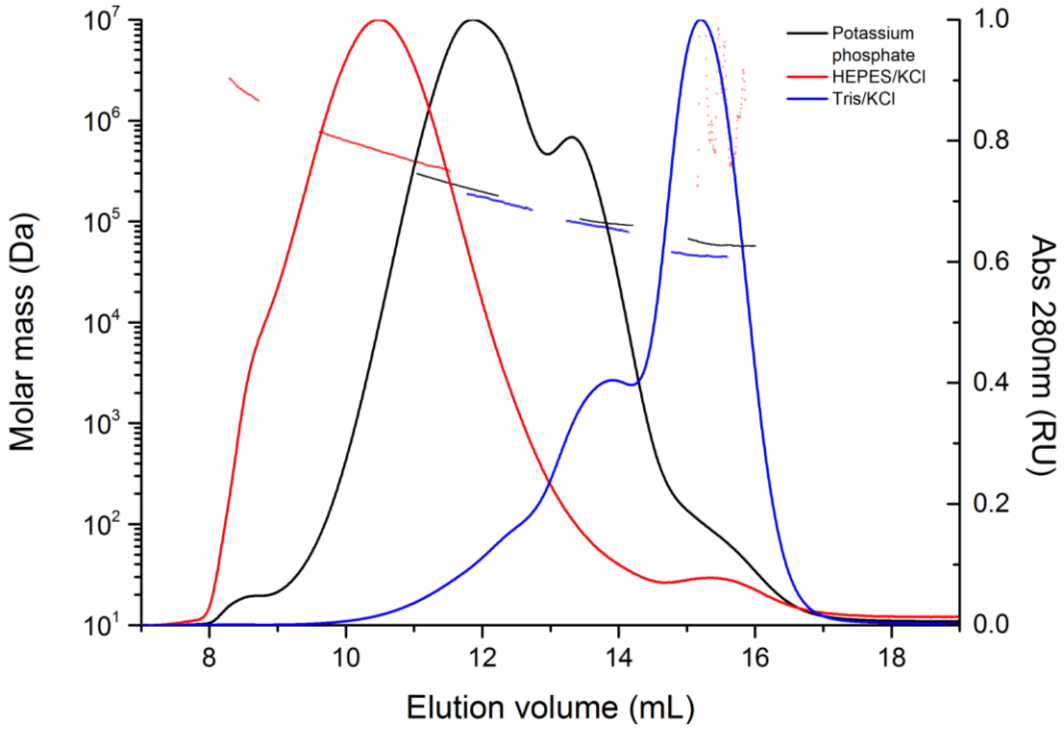


Figure 3.3 SEC-MALS analysis for initial CYP141A1 buffer optimisation.
An overlay is shown of size exclusion chromatography multi-angle light scattering experiments on the pure CYP141A1 in potassium phosphate (black line), HEPES (red line) and Tris HCl (blue line) buffer systems. The measured molar mass for the major peaks ranged between 50 and 500 kDa in these different conditions. The theoretical mass for CYP141A1 monomer is 45.9 kDa. This initial buffer screening indicates that CYP141A1 is predominantly monomeric only using the Tris HCl buffer system.

In efforts to further enhance occupancy of the monomeric state of CYP141A1, the reducing agents dithiothreitol (DTT) and tris(2-carboxyethyl)phosphine (TCEP) were added to the Tris HCl buffer system. As shown in Figure 3.4, in 10 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 7.5 (red line), the CYP141A1 sample eluted with a less extensive shoulder feature, followed by a narrower peak at 15.5 ml and with the molar mass for this major protein fraction calculated to be 46.2 kDa. The effects on monomer formation by adding either DTT or TCEP are similar (data not shown). By increasing the ionic strength and the pH of the buffer to 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5 (blue line), the CYP141A1 sample eluted with one narrow peak at 15.9 ml and the molar mass for the protein fraction was calculated to be 39.3 kDa. This optimized buffer solution (50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5) was thus used in subsequent studies for CYP141A1, as the protein was almost completely monomeric in this buffer.

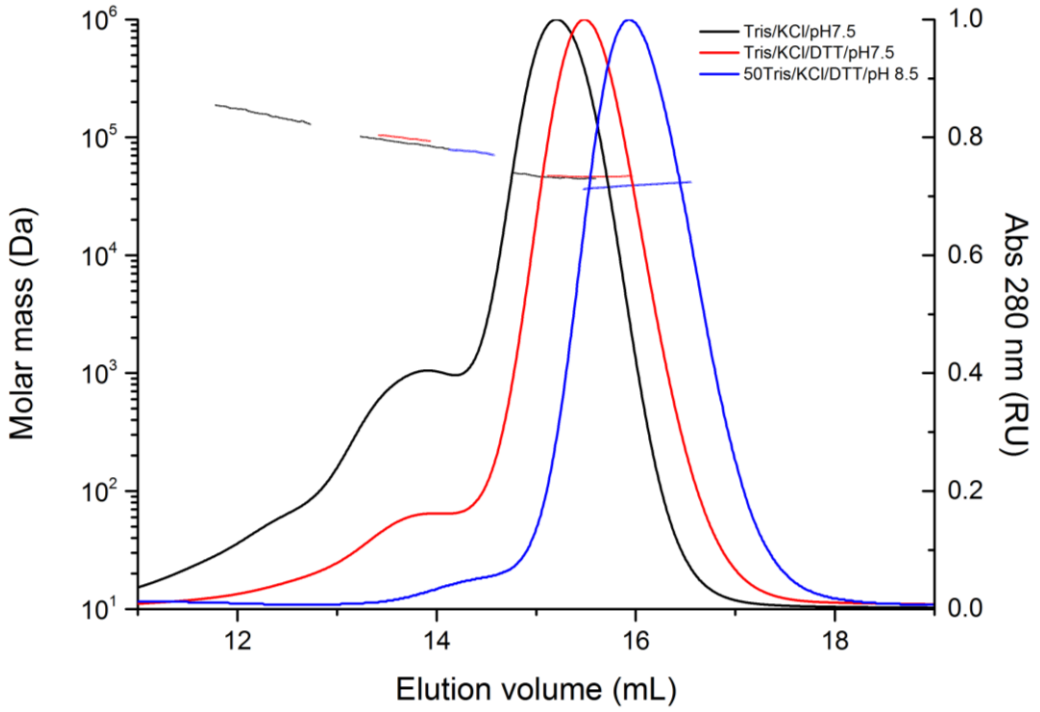


Figure 3.4 SEC-MALS analysis for final CYP141A1 buffer optimisation.
An overlay is shown of size exclusion chromatography-multi angle light scattering experiments of the pure CYP141A1 in 10 mM Tris HCl, 0.25 M KCl, pH 7.5 (black line), 10 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 7.5 (red line) and 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5 (blue line) buffer systems. The measured molar mass for the major peaks ranged between 40 and 50 kDa; the theoretical mass for the CYP141A1 monomer is 45.9 kDa. This analysis indicates that CYP141A1 is predominantly monomeric in the 50 mM Tris HCl, 0.25 M KCl, 1 mM DTT, pH 8.5 buffer conditions.

Based on these SEC-MALS results, several oligomeric states of the CYP141A1 were separated into different elution peaks and the molecular weights for these peaks were determined by using ASTRA V software (Wyatt Technology). The molecular weight calculated by MALLS was then correlated with those expected for monomer, dimer, trimer, higher oligomers and aggregates of the protein. Consequently, the results from these experiments could be categorized as follows: 1) Monodisperse and homogeneous samples (more than 90% monomer), 2) polydisperse and heterogeneous samples (50-90% monomer, protein elutes in non-void volume) and 3) aggregated samples (protein elutes in void volume). Based on this categorization, and as shown in Table 3.1, CYP141A1 was found to be monodisperse and homogeneous in Tris HCl buffer (pH 7.5-8.5), with 0.25 M KCl or NaCl, and with DTT or TCEP as a reducing agent. Therefore, following this finding, the CYP141A1 buffer system used during the final protein purification step was 50 mM Tris

HCl, 0.25 M NaCl, 1 mM DTT, pH 8.5. As a result, predominantly monomeric CYP141A1 was isolated during this final step of purification which was subjected to crystallization and gave a positive crystallization outcome. Similar analysis and conclusions were also drawn by Price et al. (11), which demonstrates the tremendous value of the SEC MALS analysis for predicting crystallization success of protein samples.

Table 3.1 A summary of SEC-MALS experiments indicating the proportions of monomeric and oligomeric states of CYP141A1 in each buffer system used.

BUFFER COMPONENTS				OLIGOMERISATION PERCENTAGE	
Buffer	**Salt**	**Additive**	**pH**	**Monomer**	**Others**
100 mM potassium phosphate			7.0	55	45
10 mM HEPES	0.5 M KCl		7.5		100
10 mM HEPES	0.5 M KCl	1 mM TCEP	7.5	50	50
10 mM Tris HCl	0.25 M KCl		7.5	60	40
10 mM Tris HCl	0.25 M KCl	1 mM TCEP	7.5	94	6
10 mM Tris HCl	0.25 M NaCl	1 mM DTT	7.5	90	10
10 mM Tris HCl	0.5 M NaCl	1 mM DTT	7.5	78	22
50 mM Tris HCl	0.25 M KCl	1 mM DTT	8.5	97	3
50 mM Tris HCl	0.25 M NaCl	1 mM DTT	8.5	100	

3.4.4 Crystallization, X-ray data collection and processing of CYP141A1

Out of the 288 conditions screened during crystallization trials, two conditions in the PACT *premier*™ HT-96 plate were identified to be conducive to CYP141A1 crystallization. However, only one of these conditions was able to generate reproducible crystals. Hexagonal bipyramidal-shaped crystals with dimensions of 100 x 100 x 100 µm were obtained with good reproducibility from 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 6.5, 20% w/v PEG 3350 conditions, as shown in Figure 3.5A. One such crystal diffracted to the highest resolution shell of approximately 3.8 Å. Even so, at this resolution, no structural solution could be found by using molecular replacement methods.

Therefore, optimization of the crystallization condition was performed in order to improve CYP141A1 resolution data. The first strategy employed was to increase the CYP141A1 crystal size in attempts to obtaining a higher quality and more complete set of diffraction data. The crystallization drop volume was increased to 1 µl with a 1:1 ratio of protein and reservoir solution (0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 6.5, 20% w/v PEG 3350). As a result, larger sized, yet similarly shaped, CYP141A1 crystals were produced with dimensions of 150 x 200 x 100 µm, as shown in Figure 3.5B. This larger CYP141A1 crystal form diffracted to an improved resolution of approximately 3.2 Å, but the structure again could not be solved using molecular replacement methods.

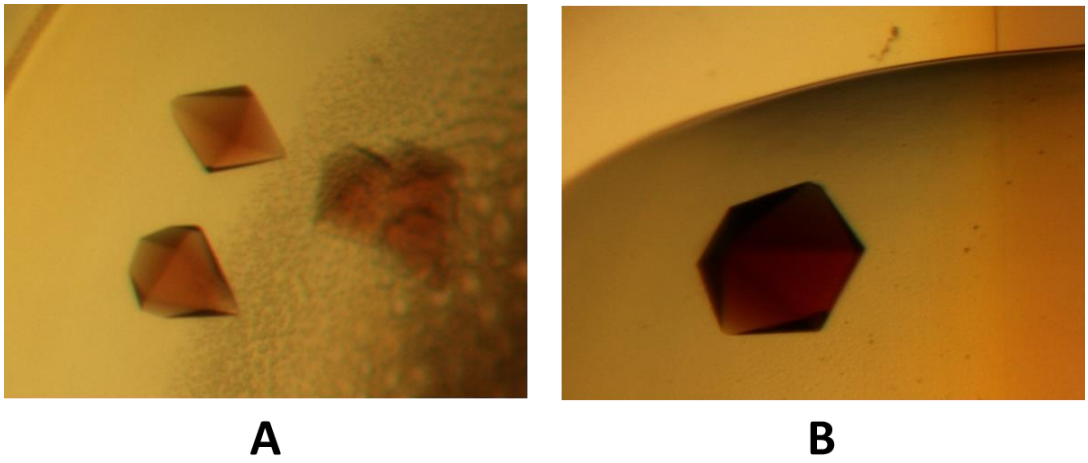


Figure 3.5 Crystallization of CYP141A1.
Hexagonal bipyramidal-shaped crystals of CYP141A1 were grown using 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 6.5, 20% w/v PEG 3350 as the reservoir solution. (A) CYP141A1 crystals with dimensions of 100 x 100 x 100 µm were grown in a 500 nl crystallization drop with a 1:1 ratio of protein and reservoir solution. (B) CYP141A1 crystals with dimensions of 150 x 200 x 100 µm were grown in a 1 µl crystallization drop with a 1:1 ratio of protein and reservoir solution.

As an alternative route to obtaining structural information, a different data collection strategy using a shorter diffraction beam wavelength was employed. CYP141A1 crystal was irradiated using a diffraction beam wavelength of 0.9282 Å. This diffraction produced higher resolution data of approximately 2.7 Å which the structure can be solved using experimental phasing method. The CYP141A1 X-ray data collection analysis is shown in Figure 3.6 and the statistics of data collection and processing are summarized in Table 3.2.

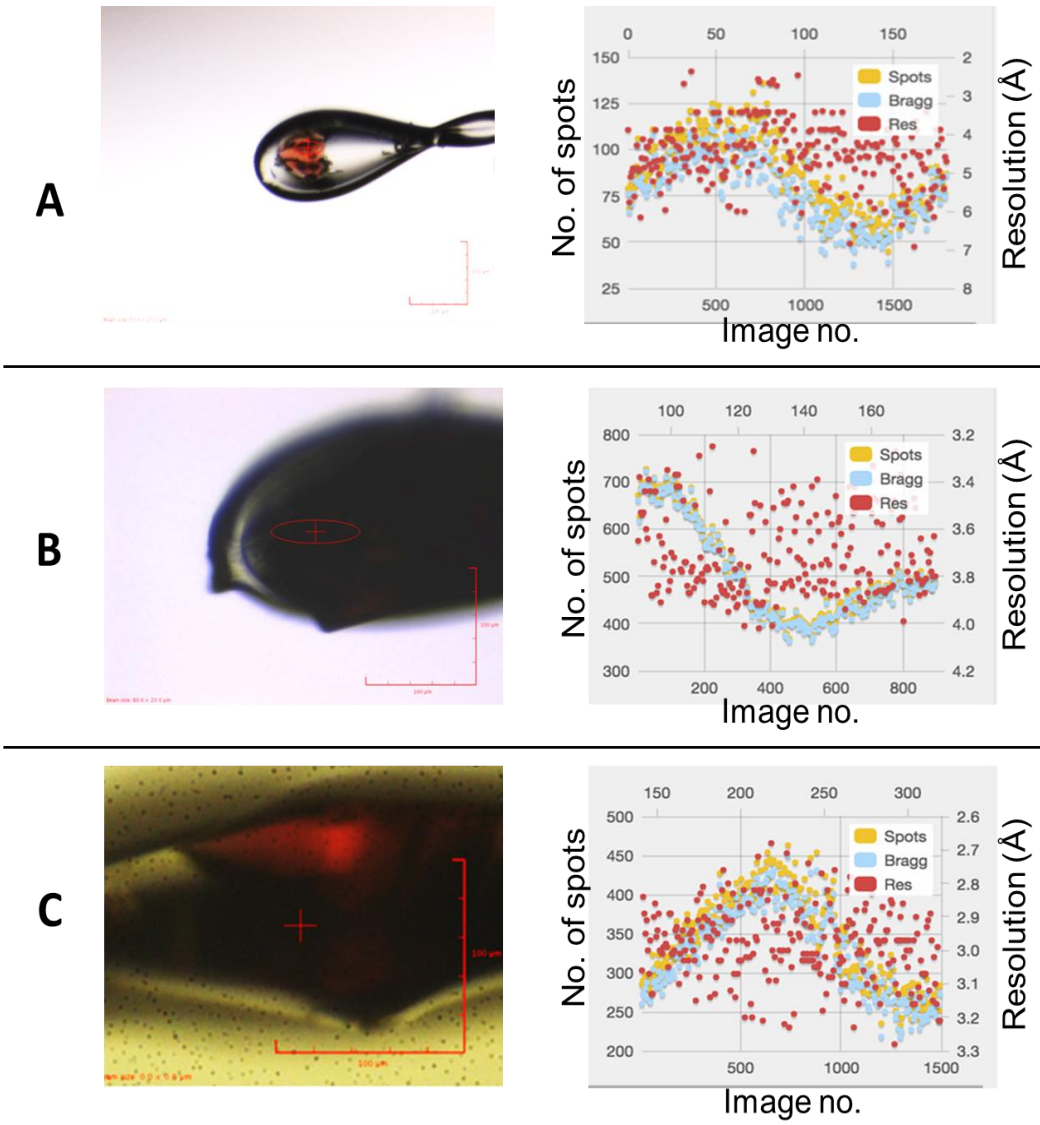


Figure 3.6 CYP141A1 X-ray diffraction data collection analysis.
In the left side panels, images are shown of CYP141A1 crystals mounted on a cryoloop for X-ray diffraction at the Diamond synchrotron beamline I03 and I04-1. In the right side panels, DISTL (12) plots for rapid analysis of X-ray diffraction patterns were collected from each CYP141A1 crystal. (A) CYP141A1 crystal with dimensions of approximately 100 x 100 x 100 µm that diffracted to a highest resolution of approximately 3.8 Å. (B) CYP141A1 crystal with dimensions of approximately 150 x 200 x 100 µm that diffracted to a highest resolution of approximately 3.2 Å. (C) CYP141A1 crystal with dimensions of approximately 100 x 150 x 200 µm that diffracted to a highest resolution of approximately 2.7 Å.

Table 3.2 Crystallization, data collection and data processing for CYP141A1.
Values in parentheses are for the highest resolution shell.

Parameters	CYP141A1 crystals		
	A	B	C
Crystallization			
Method	Sitting drop		
Protein concentration (mg ml^{-1})	40	20	20
Protein buffer solution	50 mM Tris HCl, 0.25 M NaCl, 1 mM DTT, pH 8.5		
Reservoir solution	0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 6.5, 20% w/v PEG 3350		
Volume and ratio of drop	500 nl, 1:1	1000 µl, 1:1	1000 µl, 1:1
Volume of reservoir (µl)	30	30	30
Data collection			
Diffraction source	I03, Diamond		I04-1, Diamond
Wavelength (Å)	0.9763	0.9763	0.9282
Data processing			
Space group	$P222_1$	P622	$P6_122$
a, b, c (Å)	47.3, 81.91, 262.12	95.6, 95.6, 265.77	95.7, 95.7, 265.26
α, β, γ (º)	90, 90, 90	90, 90, 120	90, 90, 120
Resolution range (Å)	262.12 – 3.77 (3.83 – 3.77)	79.04 – 3.17 (3.22 – 3.17)	265.26 – 2.67 (2.71 – 2.67)
Total no. of reflections	68464 (3563)	116429 (6070)	404648 (17683)
No. of unique reflections	10785 (532)	12989 (625)	21435 (1020)
Completeness (%)	98.1 (100)	99.3 (99.4)	100 (99.8)
Multiplicity	6.3 (6.7)	9.0 (9.7)	18.9 (17.3)
I/ σI	11.9 (2.3)	13.0 (2.3)	30.7 (2.2)
R_{meas} (%)	9.4 (89.6)	9.9 (72.7)	6.0 (136.4)
$CC_{1/2}$ (%)	100 (70)	100 (80)	100 (80)

3.4.5 CYP141A1 structure determination

An initial structure of CYP141A1 was obtained using the 2.7 Å resolution diffraction data set (CYP141A1 crystal C) and experimental phasing methods. Despite a very weak anomalous signal arising from the intrinsic iron of the heme, a novel phasing package implemented at Diamond Light Source known as Big EP was able to both phase the diffraction data and to produce an initial partial model of CYP141A1 (Figure 3.7). This partial model was used to search for structural similarity in the ePDB and resulted in a number of P450 structure hits. Each hit structure was in turn used as the input model for molecular replacement trials using Phaser. (13) However, no significant molecular replacement solutions were obtained and thus the weak experimental phases were used as the basis for model building and refinement.

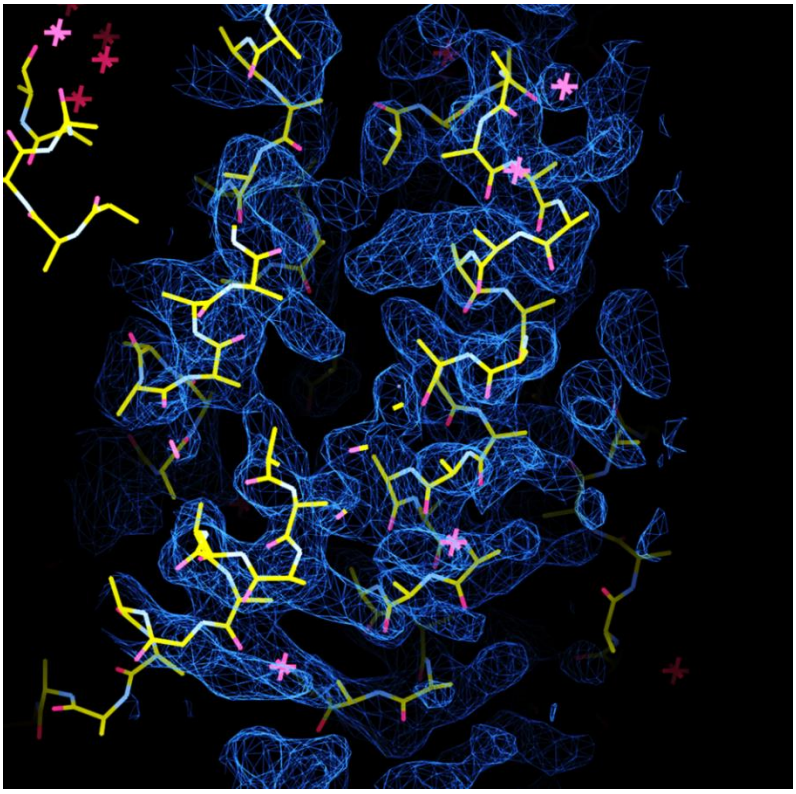


Figure 3.7 Electron density map of CYP141A1.
A phased map was obtained using AUTOBUILD, (14) a pipeline program in the Big EP phasing package. The phase obtained by AUTOBUILD is the best phase for building the initial partial model of CYP141A1. A DALI search using this model returned hits for P450s, and hence provide confidence to proceed with rebuilding and refinement of this model.

3.5 Discussion

In order to begin crystallization experiments, sufficient amount of soluble protein is needed. By using the T7 RNA polymerase system, soluble CYP141A1 enzyme was expressed when the temperature during induction is at 18 ^{o}C. Many of the well-characterised Mtb P450s have been expressed as soluble proteins through lowering temperature during induction, (8,15) presumably because slower rates of protein production allow newly transcribed recombinant protein time to fold properly.

Subsequently, the soluble protein needs to be highly pure prior to crystallization trials. The presence of N-terminal 6x-His-tag on CYP141A1 have assisted the purification process and after 3-step of chromatography, a reasonably pure CYP141A1 was obtained. However, assessment by SEC-MALLS indicate that the sample was heterogenous and polydisperse which can be a deterrent to crystallization. (16) One possible source of sample heterogeneity is the variation in oligomer state. Therefore, CYP141A1 buffer solution was optimised in order to obtain predominantly monomeric form of CYP141A1. Interestingly, CYP141A1 requires reducing agent such as DTT or TCEP in the buffer solution suggesting that heterogeneity of CYP141A1 is due to the formation of intra- or intermolecular disulphide bridge by free cysteines.(17) In addition, higher ionic strength and pH of the buffer is also required for crystallization. (18) As expected, the optimization of CYP141A1 buffer conditions resulted in the formation of reproducible CYP141A1 crystals which allows structure determination by X-ray crystallography.

As CYP141A1 crystals are readily available, these crystals were subjected to X-ray diffraction studies using a synchrotron source. Although the physical appearance of CYP141A1 crystals meets the requirement for data collection, the crystals did not produce high-resolution diffraction data. To date, CYP141A1 crystals have diffracted in the range of 2.7–4 Å. One possible reason for this moderate resolution diffraction data is the high solvent content and loose packing of the protein molecules in the crystals. Solvent content analysis of most of the CYP141A1 diffraction data is in the range of 60–70%, whereas the average value for solvent content in protein crystals is 43%. (19) It is hypothesized that the high

solvent content is due to the presence of large channels within the CYP141A1 crystal, which may be the substrate/product access/egress channels. However, since the current model structure of CYP141A1 still needs to be built and refine, this hypothesis cannot yet be confirmed.

3.6 Summary

The determination of the three-dimensional structure of CYP141A1 by X-ray crystallography was achieved following preparation of high quality P450 protein. Firstly, expression in *E. coli* and purification of CYP141A1 was optimised in order to obtain sufficient amount of highly pure enzyme for crystallization trials. A new protein buffer solution for CYP141A1 was identified through SEC-MALS experiments, and this buffer allowed the isolation of monomeric and homogeneous CYP141A1 protein samples. Following the identification of an "optimal" buffer solution for CYP141A1, crystallization trials using typical sparse-matrix crystallization screening processes gave a positive crystallization outcome. Further efforts to optimise the crystallization conditions were made, which led to improvements in CYP141A1 crystal quality and diffraction data. Finally, minor changes in the data collection strategy resulted in successful determination of an initial CYP141A1 model structure.

3.7 References

1. World Health Organization (WHO). Global tuberculosis report 2016. Geneva: WHO Press. www.who.int/tb/publications/global_report/en/ Date last accessed: July 7, 2017. Date last updated: 2016..

2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537–44.

3. Hudson SA, McLean KJ, Munro AW, Abell C. Cytochrome P450 enzymes: a cohort of novel TB drug targets. Biochem Soc Trans. 2012;40(3):573–9.

4. McLean KJ, Dunford AJ, Neeli R, Driscoll MD, Munro AW. Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems. Arch Biochem Biophys. 2007;464(2):228–40.

5. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci USA. 2003;100(13):7877–82.

6. Jancarik J, Kim SH, IUCr. Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Crystallogr. 1991 Aug 1;24(4):409–11.

7. Geerlof A, Brown J, Coutard B, Egloff MP, Enguita FJ, Fogg MJ, Gilbert RJC, Groves MR, Haouz A, Nettleship JE, Nordlund P, Owens RJ, Ruff M, Sainsbury S, Svergun DI, Wilmanns M. The impact of protein characterization in structural proteomics. Acta Crystallogr Sect D Biol Crystallogr. 2006;62(10):1125–36.

8. McLean KJ, Cheesman MR, Rivers SL, Richmond A, Leys D, Chapman SK, Reid GA, Price NC, Kelly SM, Clarkson J, Smith WE, Munro AW. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. J Inorg Biochem. 2002;91:527–41.

9. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr. 2010 Feb;66(Pt 2):133–44.

10. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta - Gen Subj. 2006;1760(9):1304–13.

11. Price WN, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J, Tong SN, Forouhar F, Swaminathan SS, Acton T, Xiao R, Luft JR, Lauricella A, DeTitta GT, Rost B, et al. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol. 2009 Jan 14;27(1):51–7.

12. Zhang Z, Sauter NK, Van Den Bedem H, Snell G, Deacon AM. Automated diffraction image analysis and spot searching for high-throughput crystal screening. J Appl Cryst. 2006;39:112–9.

13. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, IUCr, C. AS, A. P, M. N, J. RR, M. RL, T. S, L. WG. *Phaser* crystallographic software. J Appl Crystallogr. 2007 Aug 1;40(4):658–74.

14. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr Sect D Biol Crystallogr. 2007 Jan 1;64(1):61–9.

15. Driscoll MD, McLean KJ, Cheesman MR, Jowitt TA, Howard M, Carroll P, Parish T, Munro AW. Expression and characterization of *Mycobacterium tuberculosis* CYP144: Common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. Biochim Biophys Acta - Proteins Proteomics. 2011;1814(1):76–87.

16. Krauss IR, Merlino A, Vergara A, Sica F. An overview of biological macromolecule crystallization. Int J Mol Sci. 2013;14(6):11643–91.

17. van der Laan JM, Swarte MBA, Groendijk H, Hol WGJ, Drenth J. The influence of purification and protein heterogeneity on the crystallization of *p*-hydroxybenzoate hydroxylase. Eur J Biochem. 1989 Feb 15;179(3):715–24.

18. Gosavi RA, Mueser TC, Schall CA. Optimization of buffer solutions for protein crystallization. Acta Crystallogr Sect D Biol Crystallogr. 2008;64(5):506–14.

19. Matthews BW. Matthew coefficient.pdf. J Mol Biol. 1968;33:491–7.

CHAPTER 4

SPECTROSCOPIC AND BIOPHYSICAL CHARACTERISATION OF *MYCOBACTERIUM TUBERCULOSIS* CYTOCHROME P450, CYP141A1

Nur Dayana Nisbar[1], Le Van Duyet, Madeline E. Kavanagh[2], Kirsty J. McLean[1], Hazel M. Girvan[1], Colin W. Levy[3] and Andrew W. Munro[1]

[1]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

[3]Manchester Protein Structure Facility, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

4.1 Abstract

CYP141A1 from *Mycobacterium tuberculosis* was expressed, purified and characterised by various biophysical techniques. The CYP141A1 UV-visible absorption spectrum displays typical features for a cytochrome P450 at resting state, but with mixed spin-state ferric heme iron. An unusually distinct Soret shift was observed when reduced CYP141A1 is bound to carbon monoxide. Instead of having a Soret peak at 450 nm, ferrous carbon monoxide-bound CYP141A1 exhibits a Soret peak at 440 nm. In addition, electron paramagnetic resonance and magnetic circular dichroism spectra identify the spin state of the CYP141A1 ferric heme iron to be predominantly low spin with a small proportion of high spin heme iron. Moreover, CYP141A1 contains a *b*-type heme iron ligated by cysteine and a water molecule at its proximal and distal face, respectively. Binding of CYP141A1 to various azole antifungal drugs was also characterised by UV-visible spectroscopy and electron paramagnetic resonance spectroscopy, with affinity for clotrimazole, econazole and miconazole being among the highest among the drugs tested. The tight binding of these three azoles highlights the potential of CYP141A1 as a drug target. Preliminary fragment-based

screening also gives some compound hits that can be the starting point for the development of potent, isoform-specific inhibitors. Circular dichroism, differential scanning calorimetry and light scattering analysis of the purified CYP141A1 provide insights on the CYP141A1 tertiary fold and structural composition. A three-dimensional structure of CYP141A1 was obtained by X-ray crystallography, which is crucial for structure-based drug design for this P450.

4.2 Introduction

In 2017, tuberculosis (TB) continues to be a major cause of death worldwide with over 95% of TB deaths occurring in low- and middle-income countries. (1) Although *Mycobacterium tuberculosis* (Mtb) was identified as the causative pathogen of the airborne disease about 100 years ago, (2) efforts to completely eradicate the spread of this bacterium have still not been successful. The emergence of multidrug-resistant (MDR) and extensively-drug resistant (XDR) strains has complicated the problem further. Drug resistant strains typically arise due to poor administration or inadequate intake of the standard anti-TB treatments, especially in poorer regions and in those where there is limited accessibility to the drugs. Currently, only 52% of MDR-TB and 28% of XDR-TB patients are successfully treated. (1) Therefore, there is an urgent need to find solutions for the treatment of TB (all types thereof) and the best approach is through the identification of novel Mtb drug targets, as well as the development of new compounds with novel modes of action.

The determination of the genome sequence of the pathogenic *M. tuberculosis* H37Rv strain has revealed a number of novel genes and has suggested how their gene products might be targeted to kill the bacterium. A large proportion of the genome encodes for proteins involved in lipid synthesis and metabolism. (3) Mtb contains a diverse set of lipophilic molecules, including unusually long chain, extensively substituted mycolipids that form a waxy coat around the bacterium and that are likely important in preventing entry of molecules into the cell. (4) An interesting observation relating to lipid metabolising enzymes is the presence of 20 cytochrome P450 enzymes in Mtb H37Rv. These P450s are heme-containing monooxygenases known for their catalytic activity in metabolism of fatty acids,

steroids, and other lipophilic molecules. (5) This suggests that there may be important roles for these enzymes in Mtb, hence making them potential novel drug targets.

Characterisation of various Mtb P450 enzymes has confirmed novel and/or unexpected functions such as sterol demethylation, cholesterol oxidation, unusual secondary metabolite production and regulation of oxidation of respiratory menaquinone. (6–10) As the 20 Mtb P450 genes are widely dispersed across the Mtb genome and in most cases do not have high amino acid identity with one another, there is an apparent lack of clues from the genetic context to inform on the biochemical and physiological roles of other uncharacterised Mtb P450s. (11) In order to find determinants for virulence in Mtb, the genomic differences between Mtb H37Rv and the vaccine strain *M. bovis* BCG were analysed. (12) There are a number of deleted genomic regions called regions of difference (or RD regions) in the BCG strain, (13) and two of Mtb P450-encoding genes are located in this region. The *CYP141A1* and *CYP130A1* genes are present in RD12 and 13, respectively and may be relevant for Mtb virulence and infectivity toward the human host. (14) The potential of CYP141A1 and CYP130A1 as determinants of Mtb virulence is the major reason for their enzymatic characterisation.

Characterisation of CYP130A1 was already reported by Ortiz de Montellano's group, (15) and therefore we took on the task of characterising CYP141A1. To understand the properties of CYP141A1, we investigated the nature of its heme ligation state and the spin state of the ferric heme iron using a variety of spectroscopic techniques. We also examined the binding interactions of CYP141A1 with azole antifungal drugs, as well as with small molecules (fragments) identified to bind this enzyme through drug discovery fragment screening methods. In addition, we applied various biophysical techniques to explore CYP141A1 properties prior to the determination of CYP141A1's structural features using X-crystallography.

In this paper, we report the cloning and expression of the *Rv3121* (*CYP141A1*) gene from Mtb H37Rv, leading to the purification and biophysical characterisation of the CYP141A1 enzyme. CYP141A1 exhibits typical features of P450 enzymes, but reveals interesting spectroscopic properties in the Fe^{II}-CO-bound state. CYP141A1 also displays

high affinity binding for azole antifungal drugs, suggesting that the enzyme has the potential as a viable drug target. Data from CYP141A1 binding to small molecules was obtained from fragment screening studies, and the availability of a CYP141A1 crystal structure should provide a good starting point for the development of potent isoform-specific inhibitors using a fragment-based drug discovery approach.

4.3 Materials and methods

4.3.1 CYP141A1 cloning, expression and purification

The gene encoding CYP141A1 (*Rv3121*) from the *M. tuberculosis* H37Rv strain was amplified by PCR from a *M. tuberculosis* H37Rv chromosomal cosmid DNA library (provided by Professor Stewart Cole at the École polytechnique Fédérale de Lausanne [EPFL]). The bacterial artificial chromosome (BAC) clones containing CYP141A1 (*Rv3121*) were prepared using standard PCR amplification protocols with *Pfu* Turbo DNA polymerase (Agilent, UK) used to amplify the gene. The *M. tuberculosis* H37Rv genome sequence was used to design primers to generate the *Rv3121* gene for insertion into the pET15b plasmid vector using the forward primer 5'-GCGGCAGCCAT**ATG**ACTACCGCCG-3' and the reverse primer 5'-GCGCGGAGGC**TAG**GATCCGGCTGC-3'. The underlined letters in the forward and reverse primers indicate engineered *Nde*I and *Bam*HI restriction endonucleases sites, respectively. The bold letters indicate the start (ATG) and stop (TAG) codons, respectively. The gene amplification conditions used were 95 °C for 2 min; 30 cycles of 95 °C for 45 s, 62 °C for 30 s, and 72 °C for 1.5 min; and then a final polymerisation step of 72 °C for 10 min. The PCR product was then digested using *Nde*I/*Bam*HI and cloned into pET15b using a Quick Ligation Kit (New England Biolabs, UK) to generate the pCYP141A1 construct.

The production of CYP141A1 protein was achieved by transforming *E. coli* strain BL21 Gold (DE3) (Agilent, UK) with the pCYP141A1 plasmid construct. Protein was produced in *E. coli* strain BL21 Gold (DE3) (typically 12 litres, grown in 2xYT medium) by isopropyl β-D-thiogalactopyranoside (IPTG) (0.1 mM) induction in the presence of the heme precursor δ-aminolevulinic acid (0.25 mM) at OD_{600} = 0.6-0.7, 18 °C. The transformant cells were grown

for further 24 h and then cells were harvested by centrifugation using a Beckman Coulter Avanti J-25 High Performance centrifuge and a JA-10 fixed angle rotor (6000 rpm, 4 °C, 10 min), and the cell pellets were stored at -20 °C until required for protein purification.

Cell pellets were resuspended in 300 ml of lysis buffer (50 mM potassium phosphate, 500 mM potassium chloride, 10 mM imidazole, 10% glycerol, pH 8.0) with addition of a protease inhibitor tablet (SIGMAFAST™ Protease Inhibitor Cocktail tablets, EDTA-free, Sigma-Aldrich, UK), 10 µg/ml lysozyme and DNase (Sigma-Aldrich, UK). The cells were kept on ice at all times and lysed by intermittent sonication 20-25 times with 20 second bursts at full power, and with 60 seconds cooling time between each burst using a Bandelin Sonopuls GM2600 sonicator. The lysate was then centrifuged using a Beckman Coulter Avanti J-26XP centrifuge and JLA-16.250 fixed angle rotor (14500 rpm, 45 minutes, 4 °C) and the supernatant was collected.

Binding of the supernatant with Nickel-IDA resin (Generon, UK) pre-equilibrated with lysis buffer was performed using a batch method. The loaded column was washed with 20 column volumes of nickel buffer A (50 mM potassium phosphate, 500 mM potassium chloride, 20 mM imidazole, 10% glycerol, pH 8.0). The protein was eluted using nickel buffer A with 200 mM imidazole. The eluted proteins were pooled and dialysed against buffer Q (50 mM Tris-HCl, 1 mM EDTA, pH 7.5). Subsequently, the dialysed protein was loaded onto a Q-Sepharose column (Q Sepharose Fast Flow resin, GE Healthcare Life Sciences, UK) pre-equilibrated with buffer Q. The column was washed with 2 column volumes of buffer Q and the protein was eluted using a linear gradient of 0-500 mM KCl in buffer Q. The selected eluted protein was again dialysed against buffer HA (25 mM potassium phosphate, pH 6.5). Following this step, the dialysed protein was loaded onto a hydroxyapatite column (Hydroxyapatite Type 1 resin, Sigma-Aldrich, UK) pre-equilibrated with buffer HA. The column was washed with 2 column volumes of buffer HA and the protein was eluted using a linear gradient of buffer HB (500 mM potassium phosphate, pH 6.5). The selected eluted protein fractions were pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 20 MWCO 30000, GE Healthcare Life Sciences, UK) to < 1 ml prior to the final purification step using a gel filtration column (HiLoad 16/600 Superdex 200, GE

Healthcare Life Sciences, UK) with 50 mM Tris-HCl, 0.25 M NaCl, 1 mM DTT or TCEP (pH 8.5). CYP141A1 purity was determined by SDS-PAGE and UV-visible spectroscopy. The purest fractions were pooled and concentrated as previously described (to ~200 µM) and stored at -80 °C until further use.

4.3.2 Spectroscopic characterisation

4.3.2.1 UV-visible spectroscopy

UV-visible absorption spectra of the oxidized, reduced and reduced/CO-bound forms of CYP141A1 were recorded on a Cary 50 UV-visible scanning spectrophotometer (Varian, UK) using a 1 cm pathlength quartz cuvette and spectra were recorded between 250-800 nm. Typically, purified CYP141A1 at ~2-10 µM in 100 mM potassium phosphate, pH 7.0 at room temperature was used for experiments. Binding of carbon monoxide (CO) to the CYP141A1 protein was performed after the reduction of the oxidized protein using a small amount of solid sodium dithionite. 30 to 60 bubbles of CO were added slowly into the ferrous P450 solution inside a fume cupboard and spectra were recorded immediately afterwards.

4.3.2.2 Mass spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was performed on CYP141A1 (40 µM) using a Xevo G2-S Q-TOF UPLC instrument (Waters, Elstree UK) coupled to an Acquity UPLC system. Samples were passed through an Acquity UPLC BEH300 C4 column (1.7 µm 2.1 x 50 mm) using a mobile phase of Solvent A: water with 0.1% formic acid, and Solvent B: 95% acetonitrile containing 0.01% formic acid. The elution gradient was run using 95% Solution A for 5.21 minutes, 100% Solution B for 1 minute, and 100% Solution A for 1 minute at a flow rate of 0.2 ml/min over a total run time of 7.29 minutes. The electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 40 V. Nitrogen was used as the desolvation gas at a total flow of 850 litres per hour. Data acquisition and processing was performed using Micromass MassLynx v4.1 software with total mass spectra reconstructed from the ion series using the pre-installed MaxEnt algorithm.

4.3.2.3 Electron paramagnetic resonance (EPR) spectroscopy

Electron Paramagnetic Resonance (EPR) spectra were collected in X-band (~9.4 GHz) mode on a Bruker ER-300D series electromagnet which has a microwave source interfaced with a Bruker EMX control unit (Oxford Instruments, UK). A 10 K temperature was applied and maintained by using a helium flow cryostat attached with an ITC 503 controller (Oxford Instruments). A microwave power of 0.5 mW with a modulation frequency of 100 KHz and modulation amplitude of 5 G were used to collect the EPR spectra of the P450 samples. The oxidised, ligand-free CYP141A1 samples were prepared at ~200 µM using 100 mM potassium phosphate (pH 7.5), and azole-bound CYP141A1 samples were prepared with the addition of 100-500 µM of each azole to CYP141A1. The software packages supplied with the EPR instrument were used to calculate the *g*-values for all samples. EPR spectra were collected by Dr. Kirsty McLean (University of Manchester).

4.3.2.4 Magnetic circular dichroism (MCD) spectroscopy

The protein samples were prepared to a concentration of 200 µM, and the spectra for ligand-free CYP141A1 were recorded on a Jasco circular dichrograph model J-810 for the UV-visible and model J730 for near-infrared regions. MCD spectra in the UV-visible region were recorded using a 0.1 cm pathlength quartz cuvette. Spectra in the near-infrared region were recorded using a 0.2 cm pathlength quartz cuvette with the same sample concentration. MCD signals were recorded at a magnetic field of 8 Tesla and at ambient temperature. Magnetic circular dichroism spectra were collected by Dr. Justin Bradley and Dr. Myles Cheesman (Department of Chemistry, University of East Anglia).

4.3.3 Ligand binding titration by UV-visible spectroscopy

Ligand binding titrations were carried out on a Cary 50 UV-visible scanning spectrophotometer (Varian, UK) using a 1 cm pathlength quartz cuvette and at 25 ºC. The concentration of CYP141A1 was from 2-10 µM in 50 mM Tris-HCl (pH 8.5). For binding studies, the following compounds were used: (i) azole antifungal drugs; (ii) fragment compounds and (iii) FDA-approved drugs. Compounds were prepared as stock solutions

(0.1 - 100 mM) in dimethyl sulfoxide (DMSO) and titrated (in 0.1-0.2 µl aliquots) into a 1 ml cuvette containing the CYP141A1 solution up to the point where the titration was clearly completed, or until the compound became insoluble and caused turbidity. Spectra were recorded between 250 to 800 nm after each addition of ligands. Difference spectra were generated by subtraction of the ligand-free CYP141A1 spectrum from each ligand-bound spectrum. Binding constants (K_d values) were determined by plotting the absorbance changes calculated from each difference spectrum (ΔA_{peak} - ΔA_{trough}) against the concentration of the ligand used. Data were fitted using a standard hyperbolic (Michaelis-Menten) equation (Equation 1), the Morrison equation (Equation 2) or the sigmoidal Hill equation (Equation 3) depending on the affinity of the ligand and type of binding curve observed. All data fitting was performed using Origin software (OriginLab, MA). All titrations were repeated in triplicate, and data from representative experiments are reported.

Equation 1: $A_{obs} = (A_{max} \times L\ )/(K_D + L)$

Equation 2: $A_{obs} = (A_{max}/2E) \times \left(\left((L + E + K_D) - ((L + E + K_D)^2) - (4 \times L \times E)\right)^{0.5}\right)$

Equation 3: $A_{obs} = A_{max}(L^n/(K_D + L^n))$

A_{obs} is the observed change in absorbance, A_{max} is the maximum absorbance change at apparent ligand saturation, E is the total enzyme concentration, L is the concentration of ligand and n is the apparent number of binding sites/extent of cooperativity in the case of the Hill equation.

4.3.4 Fragment-based screening for substrate/inhibitor-like ligands

Fragment-based screening studies were conducted in collaboration with Professor Chris Abell's lab at the University of Cambridge. Approximately 1 ml of purified CYP141A1 (~200 µM) was sent to the Abell lab for the initial fragment screening. CYP141A1 was assayed against heme-focused and NMR fragment libraries using a UV-visible spectrophotometric heme absorbance-shift assay and a ClarioStar plate reader. In addition, the fragment libraries were screened in cocktails of 3 fragments for binding to CYP141A1 by

relaxation edited, ligand-observed NMR spectroscopy, using a Carr-Purcell-Meiboom Gill (CPMG) pulse sequence. A reduction in the intensity of ligand proton signals of >80% compared to samples that did not contain protein was used to select fragments as hits. Following this process, the binding interactions of the fragment hits and CYP141A1 were characterised using ligand binding titrations by UV-visible spectroscopy (as described in Section 4.3.3).

4.3.5 Biophysical characterisation of CYP141A1

4.3.5.1 Circular dichroism (CD)

CD spectra were recorded on a JASCO J-810 spectropolarimeter at 25 °C. Far-UV spectra were recorded over the wavelength range of 190-260 nm in a 0.1 cm pathlength quartz cuvette with a scan rate of 10 nm min^{-1}. Near UV and visible CD spectra were recorded over the wavelength ranges 260-320 nm and 320-600 nm, respectively, using a 0.5 cm pathlength quartz cuvette and with a scan rate of 20 nm min^{-1}. The CD spectrum of the buffer (10 mM potassium phosphate, pH 7.5) was then subtracted from the measured spectrum to give the corrected spectrum. Spectra were recorded in duplicate and averaged. The protein concentrations of CYP141A1 used were 2 µM (far UV) and 20 µM (near UV-visible). The secondary structural content was analyzed by fitting using the SELCON program. (16) The CD spectra in both the far-UV and near UV-visible regions were plotted using Origin software (OriginLab corporation, Northampton USA).

4.3.5.2 Differential scanning calorimetry (DSC)

DSC was performed using CYP141A1 (8 µM) in 10 mM potassium phosphate, pH 7.5. A baseline scan was collected by loading the DSC sample and reference cells with degassed buffer using an appropriate temperature and scan rate (typically 20-90 °C, 60 °C/h). The DSC scans were repeated several times using the same parameters to establish reversibility and reproducibility. This repetition minimises baseline artefacts that can be induced by the thermal shock involved in loading the sample or reference cells. The protein solution was added to the DSC sample cell after final cooling and scanned using the same

program as in the reference scan. All thermal transitions were record by a VP-DSC instrument (MicroCal Inc., Amherst, MA). Data were analysed using Origin software.

4.3.5.3 Size exclusion chromatography with Multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were performed using an ÄKTA Purifier HPLC system (GE Healthcare Life Sciences, UK), coupled to a Wyatt DAWN HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX differential refractometer (Wyatt Technology, UK). Chromatographic separation was performed using Superdex-200 10/300 GL size-exclusion column (GE Healthcare Life Sciences, UK) with a 100 µl sample loop at a flow rate of 0.75 ml/min, in 50 mM Tris-HCl, 0.25 M NaCl, 1 mM DTT or TCEP (pH 8.5) buffer. The output was analysed by ASTRA V software (Wyatt Technology, UK). For CYP141A1, the value used for dn/dc was 0.183 ml/g. The light scattering system will provide accurate estimates of molar mass for homogeneous samples and will indicate where peaks are composed of mixtures either of different oligomers or conformers, as well as estimation of polydispersity. Polydispersity is calculated by dividing the weight-average molar mass (M_w) with the number average molar mass (M_n). SEC-MALS experiments were performed by Ms Hilda Diana Ruiz from the Biomolecular Analysis Facility, University of Manchester.

4.3.6 Crystallization and structure determination of CYP141A1

Crystals of CYP141A1 were obtained by mixing 400 nl of 40 mg ml^{-1} CYP141A1 with 400 nl of reservoir solution containing 0.2 M potassium thiocyanate, 0.1 M Bis-tris propane pH 6.5, 20% w/v PEG 3350, using the sitting drop method and by incubating at 4 °C. Reddish crystals of CYP141A1 appeared after three days and continued to grow for 2 weeks. A single reddish crystal of CYP141A1 was mounted on a 0.1-0.2 mm cryoloop, dipped briefly in glycerol as cryoprotectant and frozen in liquid nitrogen prior to diffraction. X-ray diffraction data were collected at the Diamond synchrotron beamline I04 (Harwell UK). The data were indexed and processed with the XDS program package. (16) The initial structure of ligand-free CYP141A1 was solved by an experimental phasing method. Manual model rebuilding was performed with COOT. (17) Structure refinement and validation were

performed with phenix.refine (18) and MolProbity (19) from the PHENIX package (20), respectively. All molecular graphics were prepared using PyMOL (21) or CCP4mg (22).

4.4 Results

4.4.1 Purification of CYP141A1

CYP141A1 was purified using a series of column chromatography steps involving Nickel-IDA, Q-Sepharose, Hydroxyapatite and Superdex 200 resin. Using the Reinheitszahl (Rz) (A_{420}/A_{280}) value as a guide to CYP141A1 protein purity, the four step purification strategy typically resulted to pure CYP141A1 with a Rz value of 1.6 and an approximately 8-10-fold purification from the cells. In addition, analysis by SDS-PAGE also demonstrates high purity of CYP141A1 by the appearance of a single major band with an apparent molecular mass of ~46 kDa. This is consistent with the predicted mass from the amino acid sequence of 45.9 kDa, including the N-terminal 6x-His tag region. The mass of the His_6-tagged full-length CYP141A1 after the removal of the N-terminal methionine was confirmed by LCMS analysis, with two major features observed at 45790 and 45968 Da. The former (45790 Da) likely corresponds to the mass of the intact protein minus the initiator methionine (predicted as 45763 Da), while the second species (45968 Da) has a slightly higher molecular weight, possibly due to a covalent modification. Furthermore, SEC-MALS analysis showed that the pure CYP141A1 was predominantly monomeric with a major elution peak at 16 ml that corresponds to a molecular mass of approximately 40 kDa. A small amount of a higher molecular weight species is also seen at ~14 ml, and is assigned to a minor proportion of the CYP141A1 dimer.

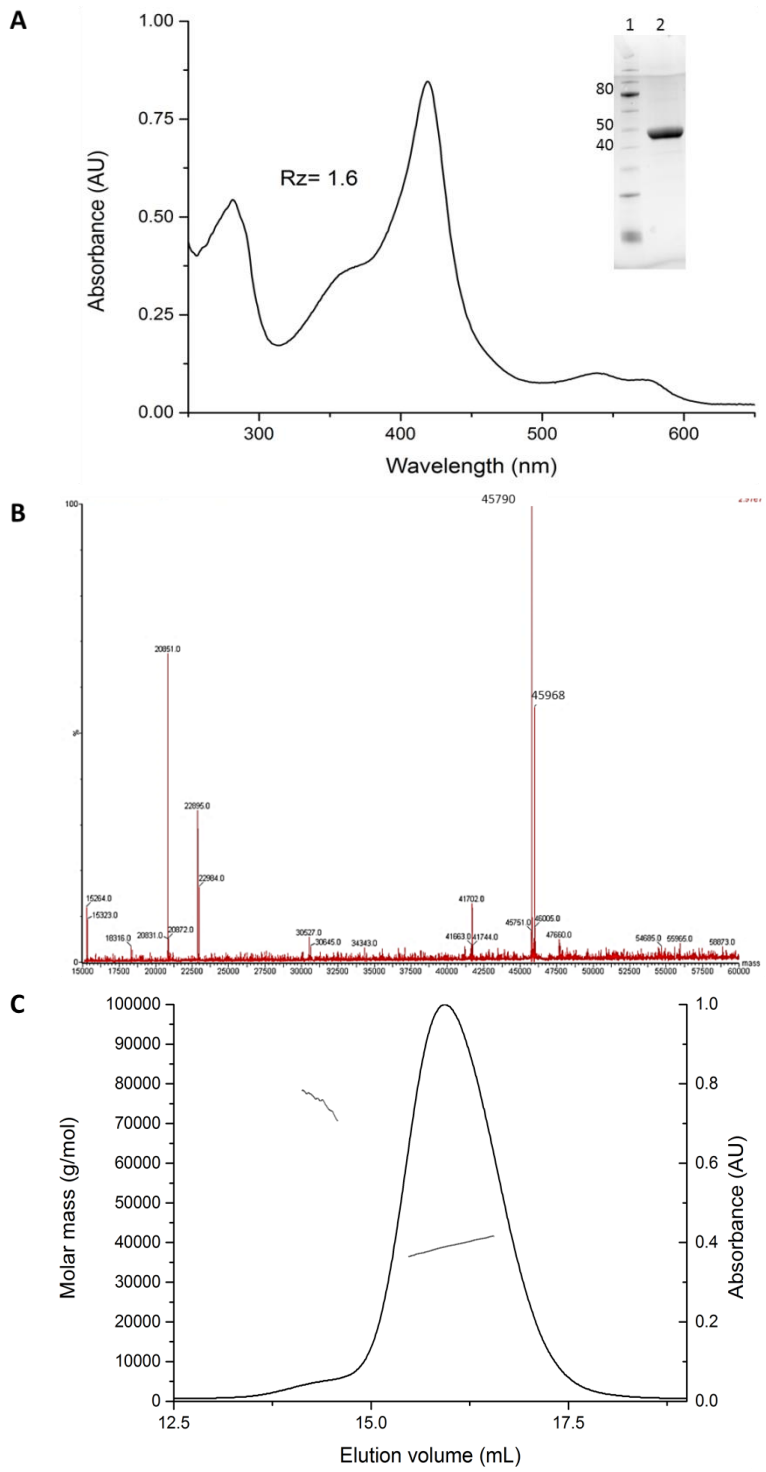


Figure 4.1 Purification of CYP141A1.

(A) The ratio of heme-specific to total protein absorption (i.e. A_{418}/A_{280}) gives a measure of the P450 purity, with a value of 1.6 indicating homogeneous CYP141A1, as verified by SDS-PAGE (inset). The inset shows an SDS-polyacrylamide gel with molecular mass markers (sizes in kDa) in lane 1 and a sample of purified CYP141A1 in lane 2. (B) The LCMS spectrum of CYP141A1 displays two major peaks at (i) 45790 Da that is consistent with the predicted mass of the His-tagged CYP141A1 apoprotein following the removal of the initiator methionine, and (ii) a slightly large species with mass of 45968 Da. (C) SEC-MALS analysis displays one major peak at an elution volume of ~16 ml that corresponds to a molecular mass of ~40 kDa, indicating that CYP141A1 is predominantly monomeric. A minor species at ~14 ml is likely a dimeric form of CYP141A1.

4.4.2 Spectroscopic characterisation of CYP141A1

4.4.2.1 UV-visible spectroscopic properties of CYP141A1

A fundamental property of P450s is the binding of carbon monoxide (CO) to ferrous heme iron to give a spectral species with a maximum near 450 nm. The UV-visible spectrum of oxidised (Fe^{3+}) CYP141A1 has its major heme absorbance band at ~418 nm (Soret) with a shoulder at ~390 nm, and with smaller α and β bands, and a cysteine thiolate-to-high spin ferric heme iron charge-transfer band located at 574, 537 and 650 nm, respectively. This indicates that CYP141A1 ferric heme iron is in a mixture of high spin and low spin states. Reduction of CYP141A1 with sodium dithionite results in a Soret band shift to 406 nm and development of a Q-band feature at 556 nm. Binding of carbon monoxide to reduced CYP141A1 produces a Fe^{II}-CO complex spectrum with a Soret peak at 440 nm, a shoulder at 420 nm and a Q-band feature at 549 nm. The spectral features of the ferrous-CO complex of CYP141A1 are consistent with the retention of cysteine thiolate as the proximal ligand to the P450 heme iron, but the shoulder at 420 nm likely indicates that a small proportion of CYP141A1 is in cysteine thiol-coordinated state. The appearance of this spectrum is reproducible for each batch of purified CYP141A1.

There is a clear dependence of the Soret peak position in relation to the solution pH. Previous studies showed that the ferric, ligand-free CYP141A1 absorbance maximum shifts from ~418 nm to shorter wavelengths as the solution pH is decreased from 7.5 to 6.0 (Duyet, LV, PhD Thesis, The University of Manchester, 2011). Figure 4.2 below shows CYP141A1 with a Soret maximum at ~418 nm in 100 mM potassium phosphate (pH 7).

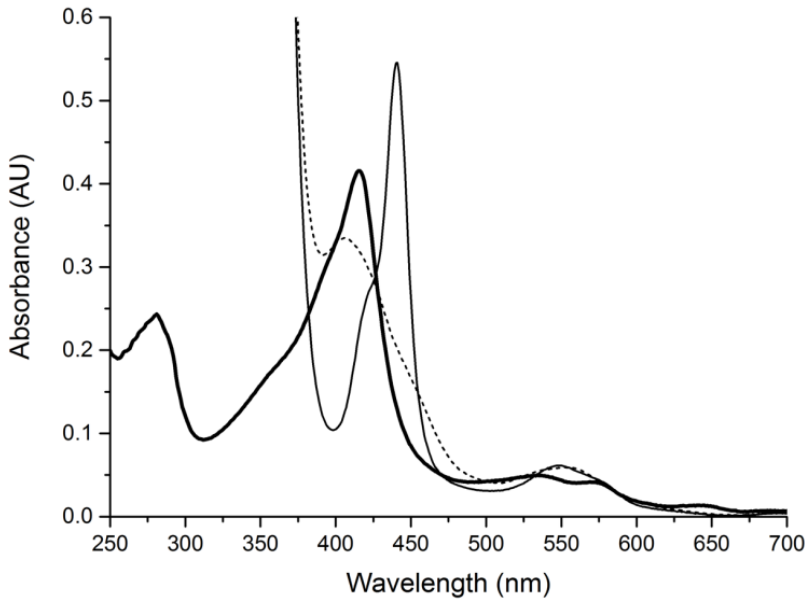


Figure 4.2 UV-visible spectral properties of CYP141A1.
The ferric form of CYP141A1 (~5 µM) has absorbance maxima at 418 nm (Soret), 574 nm (α), 537 nm (β) and 650 nm (charge-transfer species) (thick solid line). On addition of sodium dithionite, the Soret peak shifts to 406 nm (dashed line) and binding of carbon monoxide to the reduced CYP141A1 results in a Soret shift to 440 nm with a shoulder at 420 nm and a Q-band feature at 549 nm (thin solid line).

4.4.2.2 Electron paramagnetic resonance (EPR) spectroscopy of CYP141A1

X-band EPR spectroscopy was performed to report on the spin state of the CYP141A1 ferric heme iron and on the nature of the heme iron ligation. Ligand-free CYP141A1 shows major signals typical for low spin (LS) ferric heme iron, with *g*-values of 2.45 (g_z), 2.26 (g_y) and 1.91 (g_x), and another feature at 1.88 (potentially another g_x component). There and also minor signals from high spin (HS) ferric heme iron, indicated by *g*-values of 7.98 (g_z) and 3.59 (g_y). These data are consistent with the ligand-free CYP141A1 UV-visible absorption spectrum which shows a mixture of high spin and low spin states of the ferric heme iron. In addition, the *g*-values are also consistent with a low spin, ferric CYP141A1 having cysteine thiolate-coordinated P450 ferric heme iron with a distal water ligand in the sixth position. The high spin spectrum is consistent with the loss of the distal water ligand in a proportion of the enzyme.

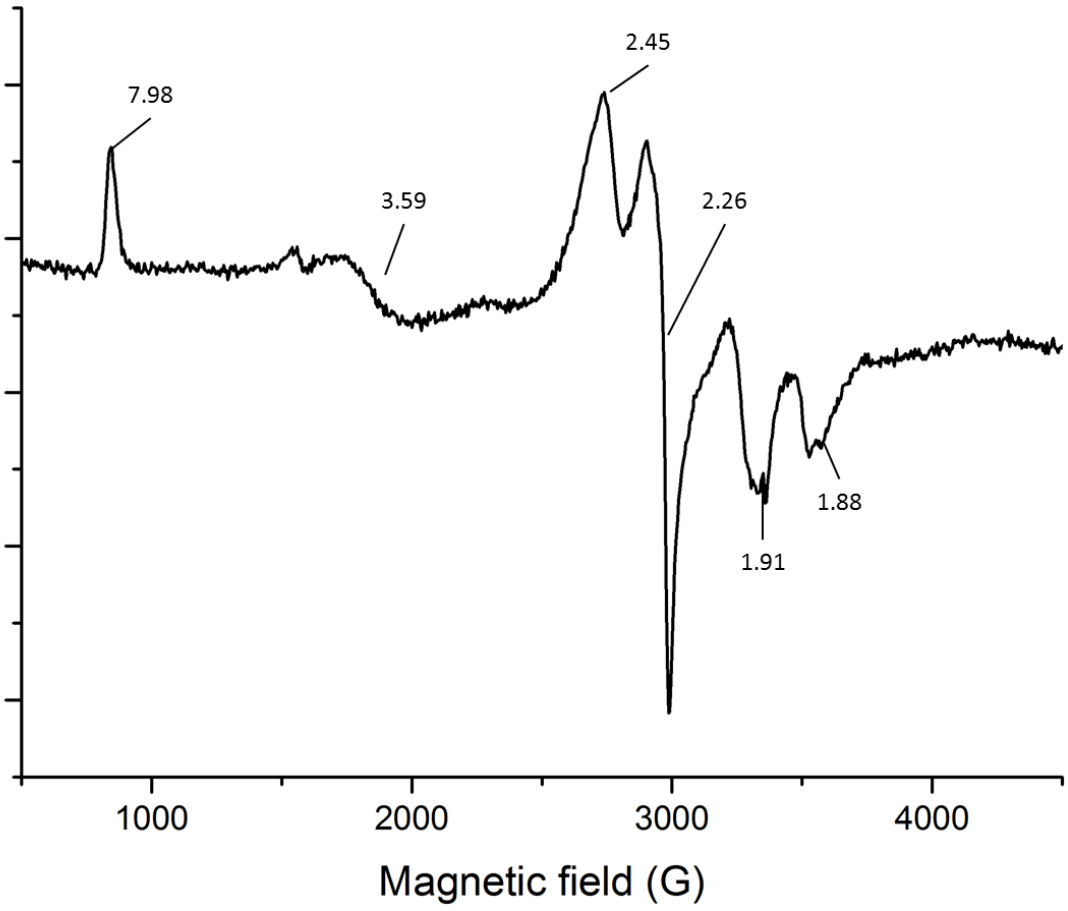


Figure 4.3 EPR spectrum of CYP141A1.
EPR spectrum for CYP141A1 in its as-purified, ligand-free form. The *g*-values for the LS (2.45/2.26/1.91 (1.88)) and HS (7.98/3.59) forms of CYP141A1 are shown in the spectrum.

4.4.2.3 Magnetic circular dichroism (MCD) spectroscopy of CYP141A1

MCD spectroscopy was also performed to further probe the spin state of the CYP141A1 ferric heme iron and to characterise its environment. The UV-visible MCD spectrum of CYP141A1 (Figure 4.4A) displays major features in the heme region located at 361 nm (-6.6), 412 nm (+34.6) and 430.5 nm (-25.9), with a crossover at 422 nm (0) that is close to the Soret peak position recorded in the UV-visible absorption spectrum. Additional features in the heme Q-band region are located at 525 nm (+7.9), 559 nm (+5.4), 567.5 nm (0) and 578.5 nm (-11). These features correspond to the low spin ferric heme iron form of the P450 with cysteine thiolate ligation. Moreover, high spin bands were also observed at 391 nm (+7.8) and at 650 nm (-0.2). These data are consistent with the presence of a proportion of high spin heme iron in CYP141A1, as observed in the UV-visible absorption and EPR spectrum of CYP141A1. A CYP141A1 MCD spectrum was also collected in the

near-infrared region (Figure 4.4B) to report on the position of a porphyrin-to-ferric heme iron charge transfer (CT) transition that can characterise the nature of the heme iron ligation in the ferric form of a heme protein. The CYP141A1 CT-band was observed at 1168 nm (+0.65) providing evidence for cysteinate-water ligation to the ferric heme iron and consistent with values obtained for other P450s. In addition, a weak positive band at ~900 nm (+0.15) was also observed that corresponds to the high spin heme iron proportion of the sample.

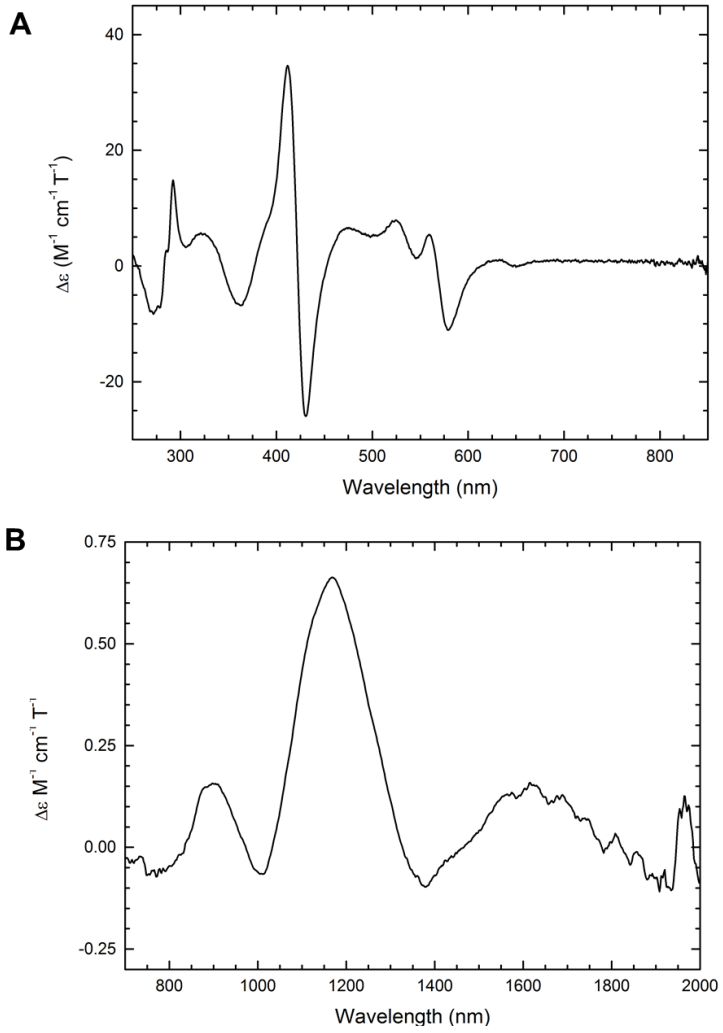


Figure 4.4 MCD spectra of CYP141A1.
(A) The UV-visible MCD spectrum for CYP141A1 (200 µM) is shown in the wavelength range from 250-850 nm. Protein was in 100 mM potassium phosphate, pH 8.5. (B) The near-infrared MCD spectrum of CYP141A1. Protein concentration was again 200 µM. Both spectra were recorded at room temperature. Spectral features are consistent with a predominantly low spin P450 ferric heme iron with cysteine thiolate coordination. Data were collected by Dr. Myles Cheesman and Dr. Justin Bradley (University of East Anglia, UK).

4.4.3 Azole binding characteristics of CYP141A1

4.4.3.1 UV-visible spectroscopy

CYP141A1 binds to various azoles antifungal drugs, exhibiting a type II (red) spectral shift of the ferric heme Soret peak. In the subsequent studies of ligand binding to CYP141A1, data were collected in 50 mM Tris HCl (pH 8.5). This buffer was used in view of studies that indicated that CYP141A1 was more stable/less prone to aggregation under these conditions, which are also conducive to CYP141A1 crystallization. The higher pH used resulted in a CYP141A1 Soret shift to ~421 nm. Addition of azole compounds induced a spectral shift of the ferric heme Soret peak from ~421 nm to ~423 nm, with small changes to the Soret absorbance intensity and modest spectral shifts in the Q-band region upon azole binding (Figure 4.5A). These spectral changes are consistent with the interaction of the CYP141A1 heme iron with a heterocyclic nitrogen atom. Almost all of the azole antifungals tested was found to bind tightly to CYP141A1. Clotrimazole, econazole and miconazole bound very tightly to CYP141A1 with K_d values of 0.22 ± 0.06, 1.59 ± 0.62 and 0.58 ± 0.10 µM, respectively. In addition, weaker binding was observed for ketoconazole and 1-phenylimidazole with K_d values of 1.98 ± 0.18 and 49.4 ± 6.1 µM, respectively. Negligible binding to CYP141A1 was observed for fluconazole, itraconazole and voriconazole, with no significant spectral changes observed upon addition of these azole drugs. All of these azoles are triazoles, whereas the azoles that do induce Soret shifts on binding are imidazoles, suggesting a strong preference for CYP141A1 towards the imidazole drugs. The K_d values for the binding of these compounds to CYP141A1 in comparison to previously reported values for their binding to other Mtb P450s are given in Table 4.1. CYP141A1 was shown to have a preference for binding azole drugs with imidazole rather than triazole groups. The same preference is seen for CYP121A1, CYP126A1 and CYP144A1, (17–19) while CYP51B1 binds both imidazole and triazole antifungal drugs without any clear preferences (Table 4.1). (20)

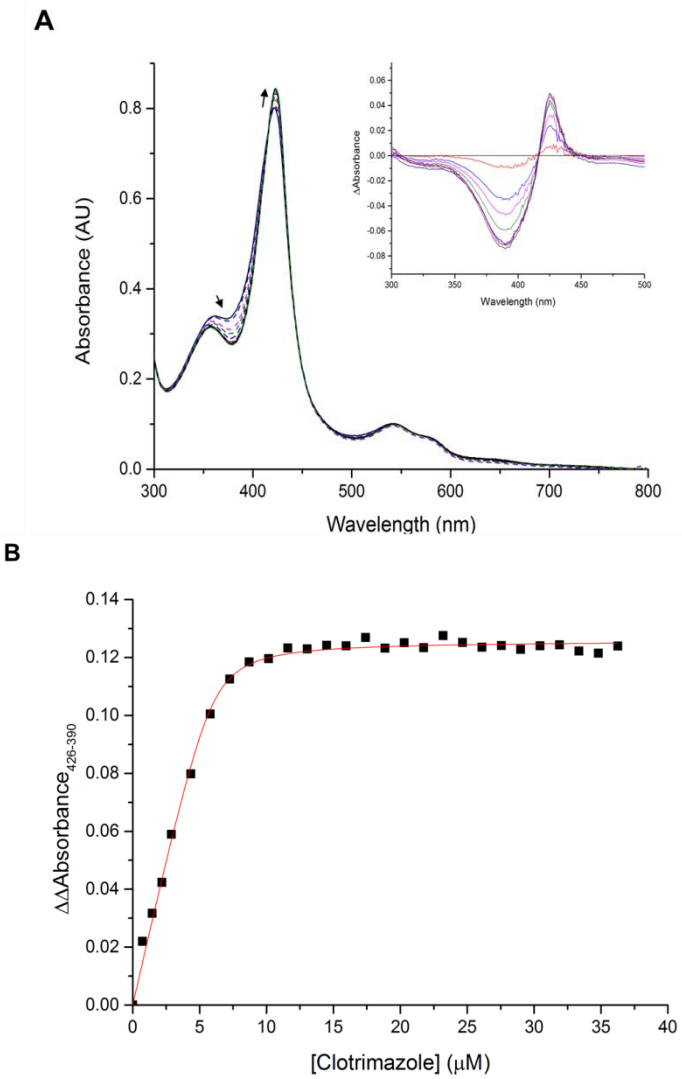


Figure 4.5 Binding of the azole inhibitor clotrimazole to CYP141A1.
(A) CYP141A1 (~8.5 µM) titration with the azole antifungal drug clotrimazole in 50 mM Tris HCl (pH 8.5). Clotrimazole binds to the ferric heme iron through its imidazole group and induces a type II red shift of the Soret peak from 421 to 423 nm. Isosbestic points are located at 416 and 444 nm. Arrows indicate directions of heme absorbance change at different wavelengths during the progression of the titration. The inset shows the difference spectra resulting from the titration of CYP141A1 with clotrimazole in the concentration range between 0 to ~40 µM. The maxima and minima in the difference spectra are located at 426 nm and 390 nm, respectively. (B) Plot of clotrimazole-induced absorption change ($\Delta A_{426nm} - \Delta A_{390nm}$) versus clotrimazole concentration, with data fitted using a tight binding (Morrison) equation to give a K_d of 0.22 ± 0.06 µM.

Table 4.1 Binding constants for azole antifungals drugs with CYP141A1 and other *M. tuberculosis* cytochrome P450 enzymes.
The dissociation constants (K_d values) for the binding of several antifungal drugs to CYP141A1 and to other *M. tuberculosis* cytochrome P450 enzymes were determined by UV-visible absorbance titration. Data for the values are taken from the following references: CYP141A1 (this work), CYP51B1 (21) CYP121A1 (20), CYP125A1 (22), CYP130A1 (15) and CYP142A1 (23). NB, no significant P450 spectral changes indicative of binding were observed upon addition of the azole drug. ND, there are no published data for the interaction of the relevant azole drug with the indicated P450 enzyme.

Azole inhibitor	Dissociation constant, K_d (μM)					
	CYP141 A1	CYP51 B1	CYP121 A1	CYP125 A1	CYP130 A1	CYP142 A1
Bifonazole	0.46 ± 0.08	ND	ND	ND	ND	ND
Clotrimazole	0.22 ± 0.06	0.18 ± 0.02	0.073 ± 0.008	5.3 ± 0.6	13.3 ± 0.6	3.8 ± 0.9
Econazole	1.59 ± 0.62	0.31 ± 0.04	0.024 ± 0.004	11.7 ± 0.7	1.93 ± 0.03	4.6 ± 0.2
Ketoconazole	1.98 ± 0.18	3.57 ± 0.25	3.44 ± 0.31	27.1 ± 0.9	48 ± 2	21 ± 4
Miconazole	0.58 ± 0.10	0.20 ± 0.04	0.136 ± 0.021	4.6 ± 0.4	1.7 ± 0.2	4 ± 1
Fluconazole	NB	5.82 ± 0.12	8.61 ± 0.21	43.1 ± 3.8	ND	860 ± 108
Voriconazole	NB	2.10 ± 0.16	ND	NB	ND	ND

4.4.3.2 EPR spectroscopy

EPR spectra for CYP141A1 in complex with azole antifungal drugs revealed perturbations that were consistent with changes in distal heme iron coordination. For CYP141 bound to the tight-binding azole compounds, the observed *g*-values change to mainly LS signals with a decrease or disappearance of the high spin signal. On addition of clotrimazole (the tightest binding azole), only the low spin signal with *g*-values of of 2.50 (g_z),

2.27 (g_y) and 1.87 (g_x) (2.50/2.27/1.87) were recorded, which is consistent with imidazole nitrogen directly ligating to the heme iron, and displacing the heme iron distal water. Econazole and miconazole binding to CYP141A1 also generated similar EPR spectra with a major low spin signals with *g*-values at 2.50/2.27/1.87 and 2.48/2.26/1.89, respectively. However, a small proportion of high spin heme iron was still retained by these azole-bound CYP141A1 enzymes, with a very small high spin signal (g_y = 7.98) likely reflecting a small proportion of CYP141A1 in which these azoles did not bind to the heme iron. The coordination of the imidazole nitrogen of these three azoles to the heme iron is likely to be a direct coordination with a displacement of the relatively weak axial water ligand by the tighter ligating azoles. However, given the relatively small g_z changes on econazole and miconazole binding (from 2.45 to 2.48 and 2.47, respectively), the indirect coordination of the heme iron via distal water molecule cannot be absolutely be ruled out. For CYP141A1 bound to the azole compounds, changes of the observed set of *g*-values for the major low spin signal and minor high spin signal were clearly seen between the ligand-free and azole-bound forms of CYP141A1 (Figure 4.6). Upon CYP141A1 binding to ketoconazole, a single major low spin species with *g*-values of 2.47/2.26/1.89 and a minor high spin signal with a g_z value of 7.98 were observed. As expected for the non-binding itraconazole, the EPR spectrum generated was similar to the ligand-free CYP141A1 with a major LS signal of 2.43/2.25/1.91 and a minor HS signal with *g*-values of 8.02 (g_z) and 3.51 (g_y).

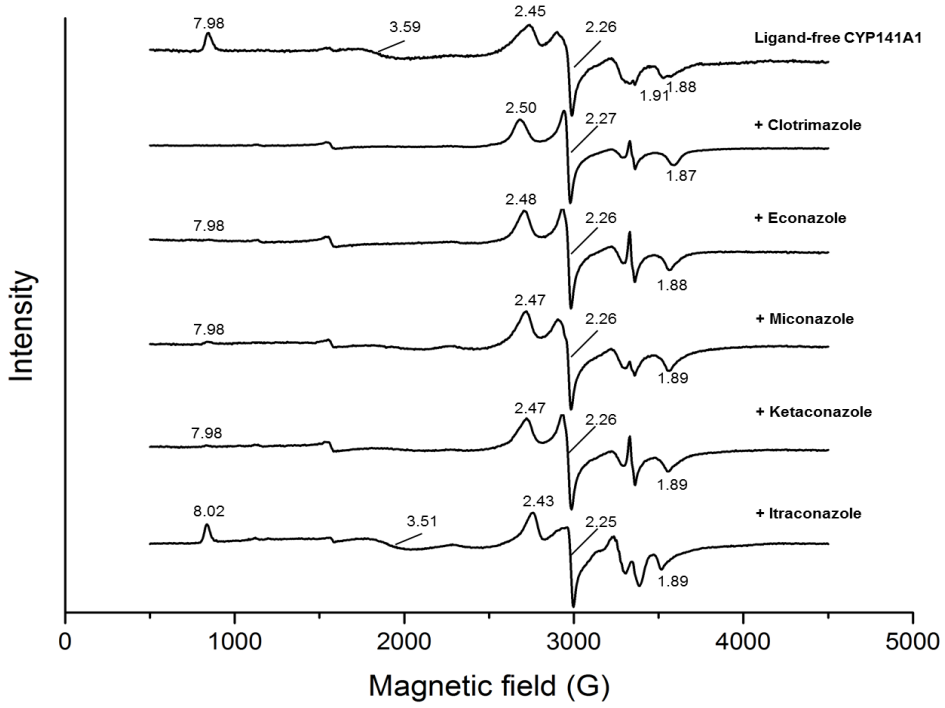


Figure 4.6 EPR spectra for CYP141A1.
EPR spectra are shown for CYP141A1 in its ligand-free form and for each of its complexes with azole antifungal drugs. The *g*-values for the low spin and high spin forms are indicated in each spectrum, with high spin values at g_z = 7.98, except for itraconazole (g_z =8.02), and also for g_y = 3.59 (ligand-free) and 3.51 (itraconazole-bound).

4.4.4 Preliminary fragment-based screening for CYP141A1

Preliminary fragment-based screening of heme-focused and NMR fragment libraries by UV-visible heme binding assay and CPMG (Carr-Purcell-Meiboom-Gill sequence) ligand-observed NMR spectroscopy generated 15 fragment hits. UV-visible binding assays identified a number of fragment hits that gave a Soret shift to a shorter wavelength, consistent with substrate-like properties. These fragment hits and the Soret shifts they induced are shown in Figure 4.7 below. All of these fragment were also found to be hits in the NMR screening process. Hit validation studies for some of the fragments were then carried out by ligand binding titrations using UV-visible spectroscopy. A Soret shift to a shorter wavelength was observed for fragment hits NMR032, NMR170 and NMR195 during titrations, consistent with the earlier heme binding assay data and indicating a partial displacement of the distal water ligands to the CYP141A1 heme iron. Dissociation constants for these fragments were calculated to be in the range between 215 - 350 µM. In addition,

compound NMR623 induced a 1.5 nm red shift, indicative of ligation to the heme iron, likely through the available nitrogen atom in this molecule (4-Bromo-1H-imidazole).

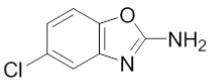
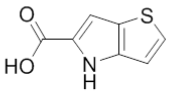
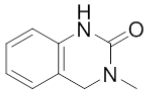
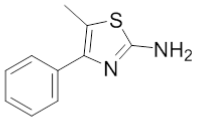
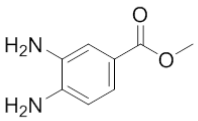
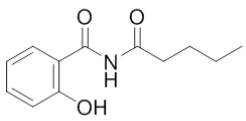
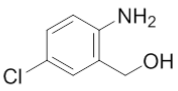
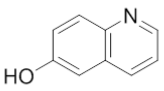
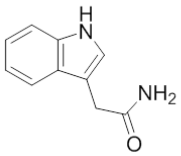
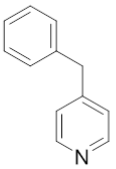
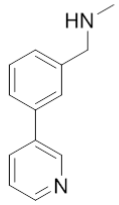
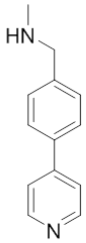
422 to 418 nm	 NMR032
422 to 419 nm	 NMR057  NMR013
422 to 420.5 nm	 NMR007  NMR008  NMR010  NMR033  NMR021  NMR006
No obvious change in Soret but weak type I difference spectra	 NMR170  NMR195  NMR415
422 to 423.5 nm	 NMR623

Figure 4.7 Fragment hits for CYP141A1.
Fragment hits are shown from initial fragment screening studies using UV-visible spectrophotometric heme absorbance-shift assays and CPMG ligand-observed NMR spectroscopy. The fragments are grouped based on their Soret shifts induced. The majority of the molecules are type I ligands, with NMR623 being the only type II (inhibitory) ligand.

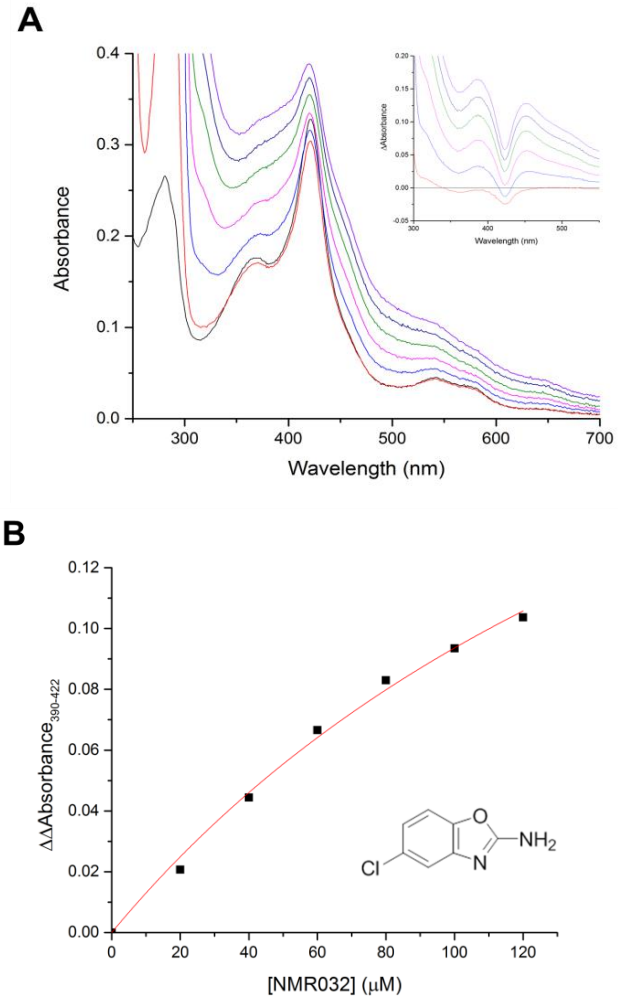


Figure 4.8 Binding of NMR032 fragment to CYP141A1.
(A) CYP141A1 (~3.3 µM) titration with the fragment NMR032 in 50 mM Tris HCl (pH 8.5). This ligand binds in the CYP141A1 active site, causes displacement of the distal water ligand to the heme iron and induces a high spin shift of the Soret peak from 421 to ~419 nm. The inset shows the difference spectra resulting from the titration of CYP141A1 with NMR032 in the concentration range between 0 to 120 µM. The maxima and minima in the difference spectra are located at 390 nm and 422 nm, respectively. (B) Plot of the NMR032-induced absorption change ($\Delta A_{390nm} - \Delta A_{422nm}$) versus NMR032 concentration, with data fitted using the Michaelis-Menten equation to give a K_d value of 212.6 ± 55.2 µM. The spectral data are affected to some extent by turbidity induced on binding the ligand, but the data can still be fitted with reasonable accuracy through mitigating effects of turbidity using peak minus trough measurements from the difference spectra.

4.4.5 Biophysical characterisation of CYP141A1

4.4.5.1 Circular dichroism (CD)

CD spectroscopy is useful for establishing protein integrity and structural content consistent with normal protein folding that is important for crystallization. It is also a rapid

method for determining for establishing protein secondary structure content in solution prior to X-ray crystallography studies. Secondary structural content can be derived from characteristic CD spectra collected in the far UV region (190-260 nm), where structural elements including alpha-helices, beta-sheets and random coil segments have different CD spectral properties. (24) The far UV CD spectrum of CYP141A1 exhibited minima at 222 and 208 nm, and a strong maximum at 190 nm that corresponds to mainly α-helical content, as expected for a cytochrome P450 enzyme. The SELCON prediction program estimated that 60% of CYP141A1 protein content is α-helical, 15% is β-sheet and around 3% is random coil. (16) In the near UV region (260-320 nm), a signal due to the aromatic amino acid side chains is observed, and is also characteristically observed in other P450 isoforms. In the visible region (320-600 nm), the CD spectrum of CYP141A1 exhibits a weak feature with negative ellipticity at ~350 nm in the region of the heme delta band. There is also a more intense spectral minimum at 410.5 nm (with a shoulder at ~385 nm) that reports on the high spin and low spin ferric forms of the CYP141A1 heme iron, respectively.

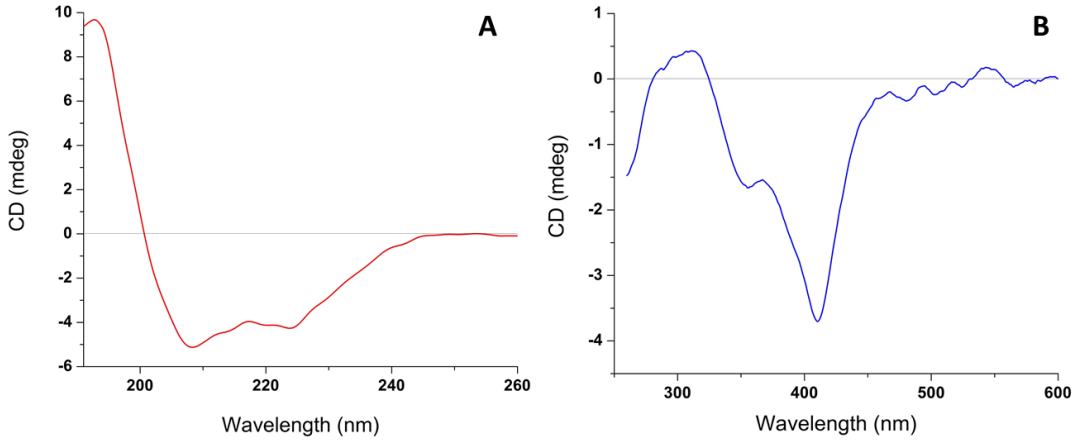


Figure 4.9 CD spectra for CYP141A1.
(A) Far-UV spectrum of CYP141A1 (2 µM) in 10 mM potassium phosphate, pH 7.5. CYP141A1 exhibits CD minima at 222 and 208 nm, and a strong CD maximum at 190 nm, consistent with a mainly alpha-helical protein. (B) Near UV-visible spectrum of CYP141A1 (20 µM) in 10 mM potassium phosphate, pH 7.5. CYP141A1 exhibits weak negative ellipticity at ~350 nm (in the region of the heme delta band) and a strong spectral minimum at 410.5 nm, reporting on the low spin (dominant) and high spin components of the Soret band of the ferric CYP141A1 heme iron. CD data are from Duyet LV, PhD thesis, University of Manchester, 2011

4.4.5.2 Differential scanning calorimetry (DSC)

DSC was performed to ensure proper protein folding and to determine protein stability, particularly the thermodynamic stability of CYP141A1 for crystallization. The global thermal stability of CYP141A1 is represented by its thermal denaturation midpoint or melting temperature (T_m). The DSC thermogram showed that CYP141A1 thermal unfolding occurs in the range of ~45-65 °C in a single broad transition. The data were best fitted to two individual transitions at 53.2 °C and 57.2 °C (T_m 1 and T_m 2, respectively), with a melting transition peak at 56.6 °C. These data likely correspond to the two different unfolding transitions associated with the smaller, beta sheet-rich beta domain (T_m 1) and the larger, alpha-helix rich alpha domain (T_m 2). These data are consistent with those observed for a number of other soluble P450s, hence validating a P450-type protein fold and tertiary structure of CYP141A1.

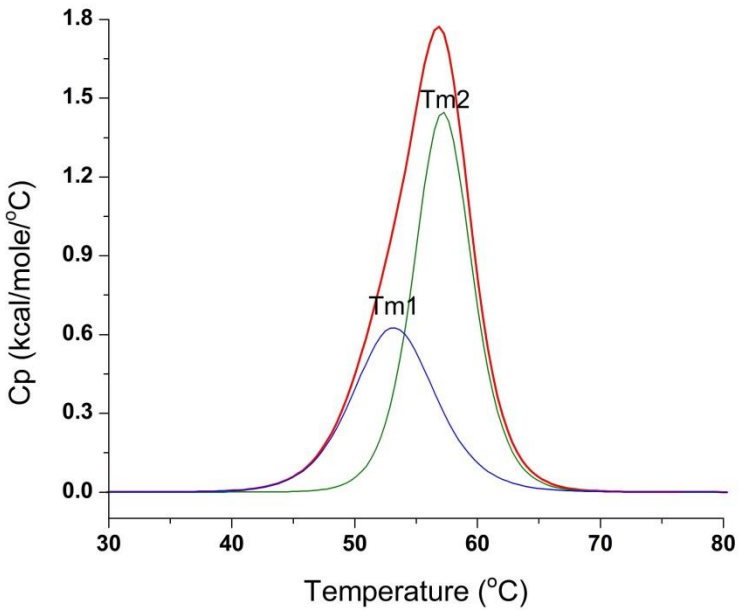
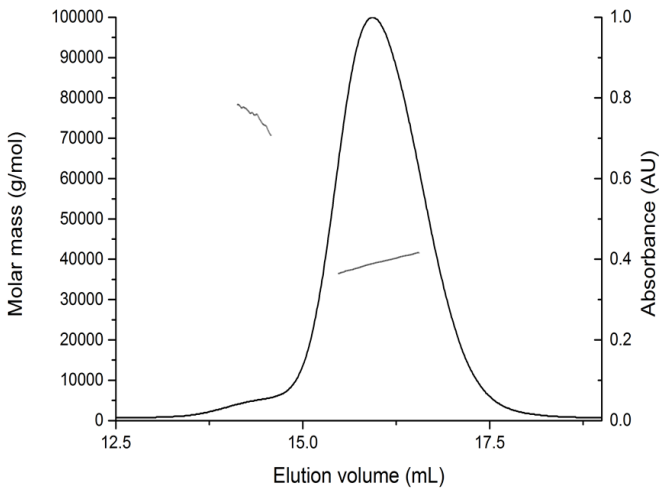


Figure 4.10 DSC spectrum of CYP141A1.
CYP141A1 (8 µM) in 10 mM potassium phosphate, pH 7.5 was heated from 20 °C to 90 °C with a scan rate of 60 °C/h. The two T_m values are indicated by T_m1 and T_m2 with values of 53.2 °C and 57.2 °C, respectively. Data are from Duyet LV, PhD thesis, University of Manchester, 2011.

4.4.5.3 Size exclusion chromatography with Multi-angle light scattering (SEC-MALS)

SEC-MALS experiments can provide information on the protein homogeneity of CYP141A1 in solution that is crucial for successful crystallization. A homogeneous sample

contains only one type of molecule of defined MW. Thus, the average mass is independent of the averaging method and its polydispersity value will equal to 1. The SEC-MALS chromatogram output of CYP141A1 exhibited a single major peak at 15.9 ml elution volume. The molar mass for the protein fraction was calculated to be approximately 40 kDa and the polydispersity value was 1.001. This polydispersity value indicates that the peak is homogeneous with respect to molar mass.



Peak Results	
Masses	
Calculated Mass (µg)	941.74
Mass Fraction (%)	97.0
Molar mass moments (g/mol)	
Mn	3.912×10^4 (±1.359%)
Mp	3.894×10^4 (±0.635%)
Mw	3.917×10^4 (±1.367%)
Mz	3.922×10^4 (±3.065%)
Mz+1	3.927×10^4 (±4.952%)
M(avg)	3.878×10^4 (±0.073%)
Polydispersity	
Mw/Mn	1.001 (±1.928%)
Mz/Mn	1.003 (±3.353%)

Figure 4.11 SEC-MALS analysis of CYP141A1.
The light scattering system provides accurate estimates of molar mass for homogeneous sample and indicates where peaks are composed of mixtures of different oligomers or conformers. For a monodisperse sample, only one molecular weight is observed with both the Mw (the average molecular weight by mass) and Mn (the average molecular weight by number) giving the same value. The ratio of Mw/Mn is the polydispersity index. The SEC-MALS chromatogram output for CYP141A1 exhibits one major peak at 15.9 ml elution volume. The Mn and Mw values within the peak are 39120 and 39170 Da, respectively, and the polydispersity index value is 1.001, indicative of a monodisperse sample.

4.4.6 Structure determination of CYP141A1 by X-ray crystallography

As described in Chapter 3, hexagonal bipyramid-shaped crystals with average dimensions of 100 x 150 x 200 µm were obtained with good reproducibility using 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 6.5, 20% w/v PEG 3350 as precipitant. The best diffracting crystal diffracted X-rays up to 2.7 Å resolution and belonged to space group $P6_122$ with unit-cell parameters of a = b = 95.7 Å, c = 265.26 Å and α = β = 90°, γ = 120°. The crystallographic phase problem for CYP141A1 diffraction data was solved using the weak anomalous signal arising from the intrinsic iron of the heme. Taking advantage of a novel phasing package implemented at the Diamond Light Source and known as Big EP, the

phase of the diffraction data was determined, consequently producing an initial model structure of CYP141A1. Iterative cycles of manual building in COOT (14) and refinement in phenix.refine (15) resulted in a current model of the CYP141A1 structure with R and R-free values of 30.7% and 36.5%, respectively. Further efforts to complete the model and to enhance the refinement are ongoing.

The current model resembles a cytochrome P450 structure, displaying the general shape of a triangular prism and containing two major structural domains that sandwich the heme, as shown in Figure 4.12A. However, only the α helical elements can be built confidently at this stage while the β sheet elements are still missing and need refinement. Structural comparison of the current model of CYP141A1 with the Mtb cyclodipeptide oxidase CYP121A1 and a computational model of CYP141A1 are shown in Figure 4.12B, which indicates the structural elements and their organization that is expected for the final model of CYP141A1.

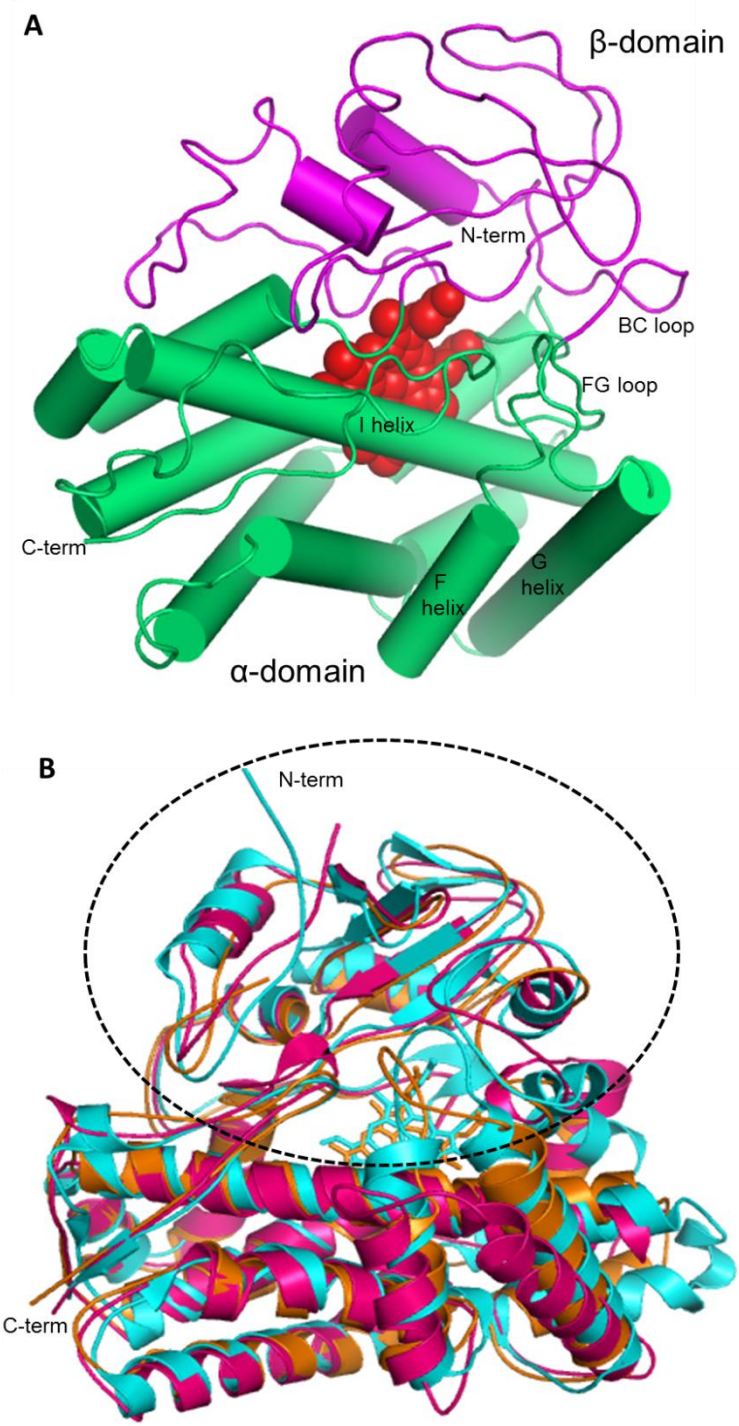


Figure 4.12 Three-dimensional structure of CYP141A1 obtained by X-ray crystallography.
(A) The overall fold of CYP141A1 resembles a trigonal prism with the heme (red, spacefill) sandwiched between the two major domains of the structure: a large α helix-rich domain (green) and the smaller β sheet-rich domain (magenta). (B) Structural comparison of CYP141A1 with CYP121A1 and the computational model of CYP141A1. Shown is the overlay of CYP141A1 (orange), CYP121A1 (cyan) and the computational model of CYP141A1 (pink). The r.m.s.d values between CYP141A1 and CYP121A1 and the computational model of CYP141A1 are 3.045 Å and 2.355 Å, respectively. The region circled needs to be rebuilt and refined in order to resolve the structural elements and their organization to produce the final model of CYP141A1. Molecular graphics were generated using Pymol graphic software. (25)

As in other P450 enzymes, the current model of CYP141A1 show that a cysteine (Cys366) provides the proximal axial ligand to the heme iron, with a water molecule as the distal axial ligand, shown in Figure 4.13A. The element known as the β-bulge segment, where the cysteine proximal ligand is located, is usually conserved and rigid. The heme binding motif of CYP141A1 and those of other structurally characterized *M. tuberculosis* cytochromes P450 were aligned using the Clustal Omega program, as shown in Figure 4.13B. Four amino acids; Phe359, Gly362, His364 and Cys366 are found conserved in all the Mtb P450 enzymes in this "heme-binding" region. In addition, the long I-helix which runs across the distal face of the heme and several conserved amino acids around the heme were identified in the current CYP141A1 model, as shown in Figure 4.13C. In the CYP141A1 structure, conserved residues such as Thr259 (implicated in proton delivery to iron-oxo species in the catalytic cycle) and Phe359 (known to be important in controlling the reduction potential of the heme iron in P450s) are likely to be important for CYP141A1 catalysis, as shown by studies on P450cam and P450 BM3. (26,27)

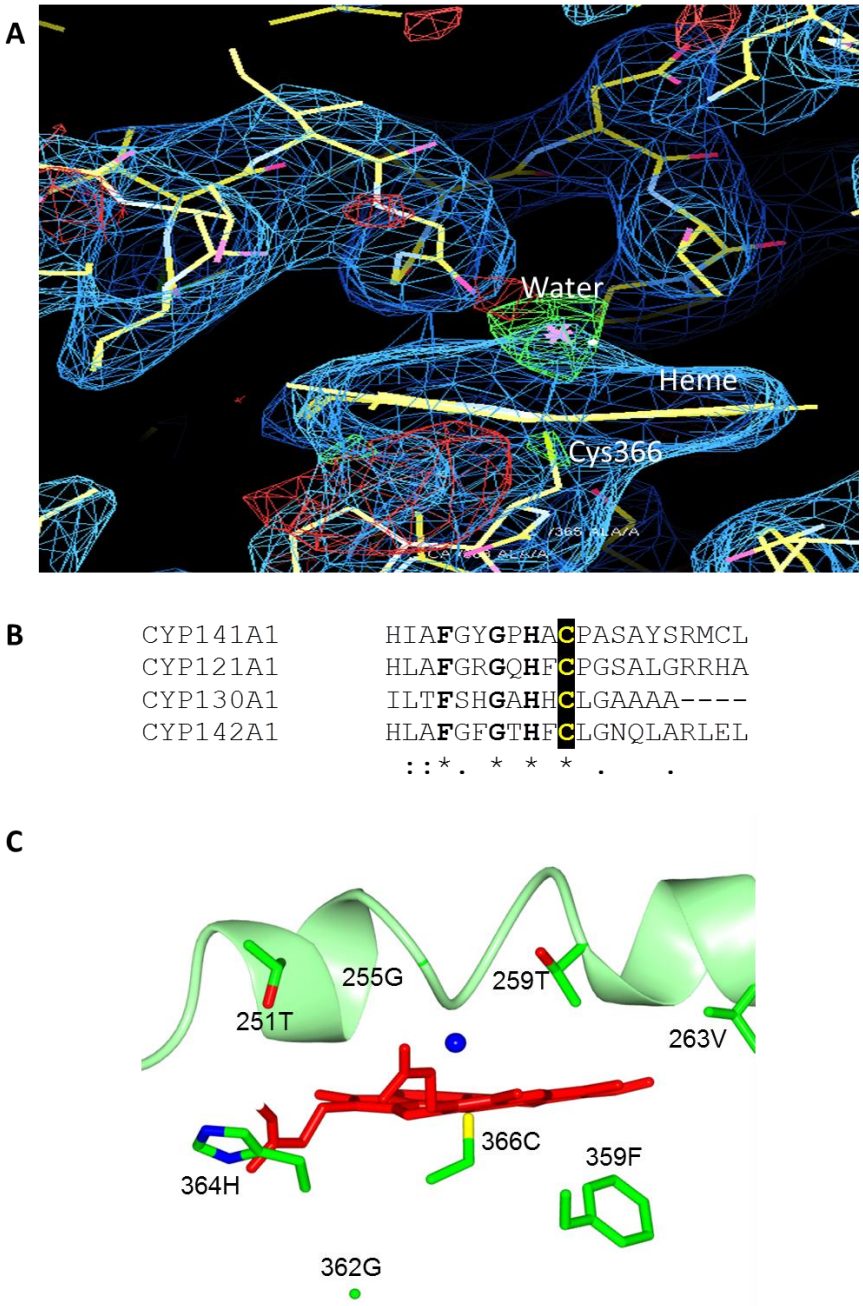


Figure 4.13 Residues surrounding the heme of CYP141A1.
(A) CYP141A1 heme binding region in atomic detail. The heme iron is ligated by a cysteine residue (Cys366) as a 5th proximal axial ligand and a water molecule as the 6th distal axial ligand. The corresponding 2FoFc electron density is contoured at 1.3 σ in blue as shown using the COOT model building program. (B) Heme binding motif of CYP141A1 aligned against other *M. tuberculosis* cytochromes P450 using the Clustal Omega program (EMBL-EBI). Four amino acids are completely conserved and are indicated in bold letters and by an asterisk at the base of the alignment. The cysteine residue that acts as the proximal ligand to the CYP141A1 heme iron and which is almost completely conserved in P450s is highlighted in yellow against a black background. Other positions where the amino acid class is retained in each P450 are indicated by dots. (C) Residues surrounding the active site of CYP141A1 within a 4 Å radius of the heme cofactor. The heme cofactor is in red, the cysteine-thiolate ligand is in yellow and the sixth (distal) water ligand is shown as a blue sphere. Shown in green are the side chains of the selected residues. Above the plane of the heme are the residues commonly found in the long I helix and below the plane of the heme are the residues shown conserved or identical in the alignment of the heme binding motif.

4.5 Discussion

Many of the well-characterised *Mycobacterium tuberculosis* (Mtb) P450s have been expressed as soluble proteins by slowing *E. coli* cell growth through lowering temperature and by decreasing IPTG concentration. (17,19,28) The same strategy was applied for the production of soluble CYP141A1 in an *E. coli* expression system that allows the isolation of pure CYP141A1 enzyme. Mass spectrometry indicates that the purified CYP141A1 is intact and SEC-MALS showed that it is predominantly monomeric (Figures 4.1B and 4.1C).

Various spectroscopic characterisation of oxidised CYP141A1 indicates that the ferric heme iron of CYP141A1 is in a mixture of a predominantly low spin state with a small proportion of the high spin state (Figures 4.2, 4.3 and 4.4). Other Mtb P450s characterised by UV-visible spectroscopy have predominantly low spin ferric heme iron with a major, narrow Soret peak located at 419 nm (CYP51B1), 416.5 nm (CYP121A1) and 418 nm (CYP142A1). An exception is CYP125A1, which is purified predominantly in the high spin ferric form with a major Soret peak located at 393 nm and a low spin Soret shoulder at 416 nm. (17,19,21,22) In addition, UV-visible, EPR and MCD spectroscopy also indicate that CYP141A1 has typical P450 cysteinate and water ligation to the ferric heme iron. The *g*-values for the LS proportion of CYP141A1 obtained from EPR are similar to those reported for other well characterised P450s, such as CYP51B1 (2.44/2.25/1.91), CYP121A1 (2.47/2.25/1.90) and P450cam (2.46/2.26/1.91). (17,21,29)

However, CO-binding to CYP141A1 displayed a Soret peak located at 440 nm instead of at ~450 nm for the ferrous-CO complex of this enzyme. This is a near-unique spectroscopic characteristic of CYP141A1 compared to other P450s, and one that may be due to altered properties of the cysteine thiolate ligand or to altered electronic properties of the heme itself. To date, only CYP170A1 from *Streptomyces coelicolor* A3(2) was reported to have similar UV-visible spectroscopic characteristics as CYP141A1 (30) and the structural basis for the observed Soret peak at 440 nm in the ferrous-CO complex remains unknown.

Consistent with a number of the Mtb P450 enzymes investigated to date, several azole antifungal drugs bind to CYP141A1 with typical coordination of a nitrogen atom in the

azole ring to the ferric heme iron, producing type II (red) shifts of the Soret peak and low spin EPR spectral features (Figures 4.5 and 4.6). Various studies have shown that azole antifungal drugs such as clotrimazole, econazole and miconazole are potent inhibitors of bacterial growth in mycobacteria and streptomycetes, but do not have significant inhibitory effects on *E. coli.* (20) This suggests that these azole drugs may target Mtb P450s *in vivo*, as *E. coli* does not have any P450s. In addition, clotrimazole and econazole have been reported to be effective against persistent and multidrug-resistant strains of Mtb, and econazole was used successfully to remove *M. tuberculosis* infection in a murine model. (31–33) Our data have shown that clotrimazole, econazole and miconazole bind tightly to CYP141A1, thus making it a putative azole drug target in Mtb. Moreover, as other Mtb P450s were also reported to have high affinity for azole drugs, these findings suggest that these azoles may target several important P450s. As a consequence, the risk of developing resistance to these azole drugs may be minimised due to their interacting with multiple cellular targets.

Due to the antimycobacterial potency of azoles and their potential to have multiple targets, various novel azole derivative compounds were developed as anti-TB drugs. (34,35) Preliminary fragment-based screening studies on CYP141A1 have provided a good starting point for the development of such compounds, as most of the fragment hits used in this study contain imidazole, triazole or pyridine groups that have the potential to coordinate the P450 heme iron (Figure 4.7). It is inferred that these fragments will interact with different binding modes within the CYP141A1 active site cavity. Therefore, at this stage, it is crucial to provide evidence for how these fragment molecules binds to the target protein.

Such evidence can clearly be provided using X-ray crystallography. Biophysical characterisation methods such as CD, DSC and SEC-MALS that were applied to CYP141A1 provided useful insights into the structure, stability and oligomerisation state of CYP141A1 in advance of crystallographic studies. Prior to crystal structure determination, CYP141A1 was expected to exhibit the general topology of a cytochrome P450 fold that consists of two major domains comprising a larger α-helix rich domain and smaller β-sheet rich domain. (28) This expectation was met by the crystal structure of CYP141A1. However, the resolution of

the structure at 2.7 Å was too low for a detailed structural analysis. A structural resolution of around 2.5 Å or better is desired for fragment characterisation. The first successful rational design of a Mtb cytochrome P450 inhibitor was based on fragment-bound CYP121A1 crystal structures solved at 1.2 to 1.5 Å resolution.(36) The four fragments were 3-((1H-1,2,4-triazol-1-yl)methyl)aniline, 2-methylquinolin-6-amine, 4-oxo-4,5,6,7-tetrahydrobenzofuran-3-carboxylate and 4-(1H-1,2,4-triazol-1-yl)phenol, which bind to CYP121A1 with K_d values of 1.6 mM, 400 ± 100 µM, 3 mM and 1.7 mM, respectively. All of the fragments bound within the large water-filled CYP121A1 distal active site and a superimposition of the fragment-bound structures reveals that the fragments form two overlapping groups that occupy a near-continuous stretch between the heme iron and the top of the active site. As initial structural determination of the ligand-free CYP141A1 has revealed features of its active site cavity and identified critical residues surrounding the heme iron, it is hypothesised that this region of the enzyme can be exploited for structure-based drug design in the future.

4.6 Summary

This manuscript presents spectroscopic and biophysical characterisation of the *M. tuberculosis* CYP141A1. A distinct characteristic of CYP141A1 compared to other Mtb P450 enzymes is a Soret band shift to 440 nm on binding CO, instead the typical P450 shift to approximately 450 nm. In addition, UV-visible, EPR and MCD spectroscopy showed that CYP141A1 exists naturally as a mixture of high spin and low spin ferric heme iron states. CD analysis indicates a typical bacterial P450 structure with predominantly alpha helical structure, a conclusion confirmed by X-ray crystallography which shows that the CYP141A1 structure consists of two major structural domains (alpha helix-rich and beta sheet-rich) as expected for a cytochrome P450 enzyme. DSC indicated two unfolding transitions at 53.2 °C and 57.2 °C, likely reflecting the T_m values for the beta and alpha domains, respectively. CYP141A1 binds to a range of azole compounds as potential inhibitors *in vivo*, and the development of specific inhibitors for this novel P450 is ongoing using a fragment-based drug discovery approach. However, future studies are also aimed at identifying substrate(s) for this novel P450 through compound screening methods.

4.7 References

1. World Health Organization. Global tuberculosis report 2016. Geneva: WHO Press. www.who.int/tb/publications/global_report/en/ Date last accessed: July 7, 2017. Date last updated: 2016.

2. Koch R. Die Aetiologie der Tuberculose (Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März gehaltenem Vortrage). Berliner klin Wochenschr. 1882;19(15):221–30.

3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537–44.

4. Brennan PJ. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis. 2003;83(1–3):91–7.

5. Denisov IG, Shih AY, Sligar SG. Structural differences between soluble and membrane bound cytochrome P450s. J Inorg Biochem. 2012;108:150–8.

6. Bellamine A, Mangla AT, Nes WD, Waterman MR. Characterization and catalytic properties of the sterol 14alpha-demethylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 1999;96(16):8937–42.

7. Ouellet H, Guan S, Johnston JB, Chow ED, Kells PM, Burlingame AL, Cox JS, Podust LM, De Montellano PRO. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. Mol Microbiol. 2010;77(3):730–42.

8. Capyk JK, Kalscheuer R, Stewart GR, Liu J, Kwon H, Zhao R, Okamoto S, Jacobs WR, Eltis LD, Mohn WW. Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27 steroids. J Biol Chem. 2009;284(51):35534–42.

9. Belin P, Le Du MH, Fielding A, Lequin O, Jacquet M, Charbonnier J-B, Lecoq A, Thai R, Courçon M, Masson C, Dugave C, Genet R, Pernodet J-L, Gondry M. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(18):7426–31.

10. Dhiman RK, Mahapatra S, Slayden RA, Boyne ME, Lenaerts A, Hinshaw JC, Angala SK, Chatterjee D, Biswas K, Narayanasamy P, Kurosu M, Crick DC. Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence. Mol Microbiol. 2009;72(1):85–97.

11. McLean KJ, Clift D, Lewis DG, Sabri M, Balding PR, Sutcliffe MJ, Leys D, Munro AW. The preponderance of P450s in the *Mycobacterium tuberculosis* genome. Trends Microbiol. 2006;14(5):220–8.

12. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. 1999 May 28;284(5419):1520–3.

13. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci USA. 2003;100(13):7877–82.

14. McLean KJ, Dunford AJ, Neeli R, Driscoll MD, Munro AW. Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems. Arch Biochem Biophys. 2007;464(2):228–40.

15. Ouellet H, Podust LM, Ortiz De Montellano PR. *Mycobacterium tuberculosis* CYP130: Crystal structure, biophysical characterization, and interactions with antifungal azole drugs. J Biol Chem. 2008;283(8):5069–80.

16. Sreerama N, Venyaminov SY, Woody RW. Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. Protein Sci. 1999 Feb;8(2):370–80.

17. McLean KJ, Cheesman MR, Rivers SL, Richmond A, Leys D, Chapman SK, Reid GA, Price NC, Kelly SM, Clarkson J, Smith WE, Munro AW. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. J Inorg Biochem. 2002;91:527–41.

18. Chenge JT, Le D Van, Swami S, McLean KJ, Kavanagh ME, Coyne AG, Rigby SEJ, Cheesman MR, Girvan HM, Levy CW, Rupp B, von Kries JP, Abell C, Leys D, Munro AW. Structural characterization and ligand/inhibitor identification provide functional insights into the *Mycobacterium tuberculosis* cytochrome P450 CYP126A1. J Biol Chem. 2016;292(1):1310–29.

19. Driscoll MD, McLean KJ, Cheesman MR, Jowitt TA, Howard M, Carroll P, Parish T, Munro AW. Expression and characterization of *Mycobacterium tuberculosis* CYP144: Common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. Biochim Biophys Acta - Proteins Proteomics. 2011;1814(1):76–87.

20. McLean KJ, Marshall KR, Richmond A, Hunter IS, Fowler K, Kieser T, Gurcha SS, Besra GS, Munro AW. Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes. Microbiology. 2002;148(10):2937–49.

21. McLean KJ, Warman AJ, Seward HE, Marshall KR, Girvan HM, Cheesman MR, Waterman MR, Munro AW. Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. Biochemistry. 2006;45(27):8427–43.

22. McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D, Munro AW. The structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection. J Biol Chem. 2009;284(51):35524–33.

23. Driscoll MD, McLean KJ, Levy C, Mast N, Pikuleva IA, Lafite P, Rigby SEJ, Leys D, Munro AW. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: Evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. J Biol Chem. 2010;285(49):38270–82.

24. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem. 2000 Dec;287(2):252–60.

25. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.8. 2015.

26. Poulos TL, Finzel BC, Howard AJ. High resolution crystal structure of Cytochrome P450cam. J Mol Biol. 1987;195:687–700.

27. Ost TWB, Miles CS, Munro AW, Murdoch J, Reid GA, Chapman SK. Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3. Biochemistry. 2001;40(45):13421–9.

28. McLean KJ, Belcher J, Driscoll MD, Fernandez CC, Le Van D, Bui S, Golovanova M, Munro AW. The *Mycobacterium tuberculosis* cytochromes P450: physiology, biochemistry & molecular intervention. Future Med Chem. 2010;2(8):1339–53.

29. Tsai R, Yu CA, Gunsalus IC, Peisach J, Blumberg W, Beinert H. Spin-state changes in cytochrome P450cam on binding of specific substrates. Proc Natl Acad Sci. 1970;66(4):1157–63.

30. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, Cane DE. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem. 2008 Mar 28;283(13):8183–9.

31. Ahmad Z, Sharma S, Khuller GK. The potential of azole antifungals against latent/persistent tuberculosis. FEMS Microbiol Lett. 2006 May;258(2):200–3.

32. Ahmad Z, Sharma S, Khuller GK, Singh P, Faujdar J, Katoch VM. Antimycobacterial activity of econazole against multidrug-resistant strains of *Mycobacterium tuberculosis*. Int J Antimicrob Agents. 2006;28(6):543–4.

33. Ahmad Z, Sharma S, Khuller G. Azole antifungals as novel therapeutic agents against murie tuberculosis. FEMS Microbiol Lett. 2006;261(2):181–6.

34. Podust LM, Von Kries JP, Eddine AN, Kim Y, Yermalitskaya L V., Kuehne R, Ouellet H, Warrier T, Alteköster M, Lee JS, Rademann J, Oschkinat H, Kaufmann SHE, Waterman MR. Small-molecule scaffolds for CYP51 inhibitors identified by high-throughput screening and defined by X-ray crystallography. Antimicrob Agents Chemother. 2007;51(11):3915–23.

35. Castagnolo D, Radi M, Dessì F, Manetti F, Saddi M, Meleddu R, De Logu A, Botta M. Synthesis and biological evaluation of new enantiomerically pure azole derivatives as inhibitors of *Mycobacterium tuberculosis*. Bioorg Med Chem Lett. 2009;19(8):2203–5.

36. Hudson SA, McLean KJ, Surade S, Yang YQ, Leys D, Ciulli A, Munro AW, Abell C. Application of fragment screening and merging to the discovery of inhibitors of the *Mycobacterium tuberculosis* cytochrome P450 CYP121. Angew Chemie - Int Ed. 2012;51(37):9311–6.

CHAPTER 5

SPECTROSCOPIC AND STRUCTURAL CHARACTERISATION OF CYP143A1, AN ORPHAN CYTOCHROME P450 FROM *MYCOBACTERIUM TUBERCULOSIS* H37RV

Nur Dayana Nisbar[1], Shalini Swami[1], Madeline E. Kavanagh[2], Kirsty J. McLean[1], Hazel M. Girvan[1], Colin W. Levy[3] and Andrew W. Munro[1]

[1]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

[3]Manchester Protein Structure Facility, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

5.1 Abstract

Tuberculosis is caused by the human pathogen *Mycobacterium tuberculosis* (Mtb). Genome sequencing of the virulent strain *M. tuberculosis* H37Rv revealed the presence of 20 cytochrome P450 (CYP) enzymes, many of which are essential for bacterial viability, survival and pathogenicity in the human host, thus making them viable drug targets. CYP143A1, encoded by the *Rv1785c* gene, is an orphan enzyme with no known function. CYP143A1 is located chromosomally adjacent to the *Rv1786* gene encoding a 3Fe-4S ferredoxin, which is potentially the redox partner for CYP143A1. Spectroscopic analyses of CYP143A1 have provided fundamental information on the nature of its heme binding site and its interactions with azole drugs. The CYP143A1 ferric heme iron resting state was found predominantly in a low spin state, ligated by a cysteine proximal ligand (in the thiolate state) and by a distal water molecule as the sixth ligand. The determination of a high-resolution crystal structure of CYP143A1 has enabled elucidation of the architecture of its active site and allowed for the application of a fragment-based screening approach to facilitate inhibitor development. Various fragment hits were identified and their interactions with CYP143A1 were characterised by spectroscopic techniques to reveal their novel ligand binding modes within the CYP143A1 active site cavity. The application of fragment-based screening approaches

should provide the basis for the development of *M. tuberculosis* P450 isoform-specific inhibitors, as well as here providing insights into the molecular structure, catalytic function and physiological role of CYP143A1 in *M. tuberculosis*.

5.2 Introduction

Tuberculosis (TB), caused by the pathogenic bacterium *Mycobacterium tuberculosis* (Mtb) continues to infect millions of the human population and has claimed more lives worldwide than any other infectious disease. (1) The standard front-line anti-TB medications that are administered today resulted from a “golden age” of antibiotic drug discovery between the 1940s and 1960s. After the approval of rifampicin in 1967, there was a gap of over 40 years until the approval of a new anti TB drug, bedaquiline in 2012. (2) In this 40 year gap, the growing epidemic of multidrug-resistant (MDR; resistance to at least isoniazid and rifampicin); and extensively-drug resistant (XDR; resistance to at least isoniazid and rifampicin, as well as to any fluoroquinolone and to any of the three second-line injectable drugs – amikacin, capreomycin or kanamycin) strains of Mtb have complicated the treatment of TB and have made efforts to stop the spread of this disease more challenging. There is a crucial need for shorter tuberculosis treatment regimens (shorter than 6 months for drug-sensitive TB and shorter than 18 months for MDR TB) to improve patient adherence, improve treatment outcomes and reduce treatment cost. (3,4)

High-throughput screening (HTS) approaches have identified a number of compound hits that show high potency against *M. tuberculosis in vitro*, but not *in vivo*, due to issues associated with penetrating the bacterial cell wall. (5) Over the past 10 years, the field of TB drug discovery has seen a shift to high-throughput whole cell phenotypic screening from target-based screening in the search for potent inhibitors. In both approaches, poor efficiency of identifying new TB drugs has been linked to the limited chemical diversity within pharmaceutical library collections. (6) Hence, a different approach that allows exploration of more diverse chemical space is the way forward for developing new anti-TB drugs. Fragment-based drug discovery (FBDD) is a powerful and now widely used approach that begins with screening a library consisting of small molecules (150-300 Da) against a target

protein, using a variety of biochemical, biophysical and structural biology methods to identify fragment hits and their binding modes. Fragment hits provide good starting points for rational compound elaboration that explores protein “hotspots” for ligand interactions. In addition, the good physiochemical properties of fragment libraries are relevant for TB drug discovery, in which several existing anti-TB drugs are fragment-sized molecules and not drug-like in their properties. Due to the complex life cycle of the pathogen and the progression of the disease, existing antimicrobials for TB include molecules that are much larger and more lipophilic than those defined as drug-like, as well as other anti-TB drugs that are much smaller and less lipophilic.

In recent years, there have been a number of novel inhibitors of Mtb target proteins developed through fragment-based drug discovery campaigns targeting several Mtb P450s. (6-10) We have applied a similar methodology to CYP143A1, an orphan cytochrome P450 from *M. tuberculosis* H37Rv. CYP143A1 is encoded by the *Rv1785c* gene and is located alongside a ferredoxin-encoding gene (*Rv1786*) in the Mtb genome, as shown in Figure 5.1. The ferredoxin’s function is inferred to play a role in transferring electrons to CYP143A1 thereby suggesting an important role of the P450 in Mtb.

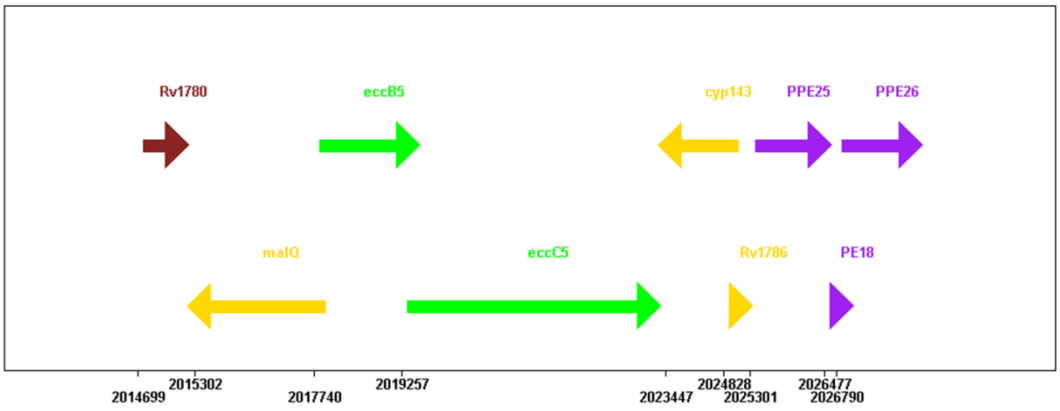


Figure 5.1 Gene environment of CYP143A1 in the *M. tuberculosis* H37Rv genome.
CYP143A1 is encoded by the *Rv1785c* gene and is located at 2023.45 kb alongside a ferredoxin-encoding gene (*Rv1786*) in the Mtb genome. Figure is adapted from TubercuList (http://tuberculist.epfl.ch/).

The major aims of this study are to apply a fragment-based screening approach to CYP143A1 in order to develop isoform-specific inhibitors of this enzyme as novel anti-TB

drug candidates, and also to discover the likely natural substrate for CYP143A1, hence determining the functional role of the enzyme. In this paper, we report on the structural and spectroscopic characterisation of ligand-free, azole-bound and fragment-bound forms of CYP143A1. The binding of azole antifungal drugs and diverse fragment molecules to the P450 provides new insights into CYP143A1 ligand preferences and ligand binding interactions within its active site cavity, providing data which are beneficial for structure-based drug design.

5.3 Materials and methods

5.3.1 Cloning, protein expression and purification of CYP143A1

The *Rv1785c* gene encoding CYP143A1 was amplified by PCR from a *M. tuberculosis* H37Rv chromosomal cosmid DNA library (provided by Professor Stewart Cole, EPFL, Lausanne, Switzerland). The *Rv1785c*-containing bacterial artificial chromosome clones were prepared by using standard PCR amplification protocols and *Pfu* Turbo DNA polymerase (Agilent, UK) to amplify the *Rv1785c* gene. The genome sequence of *M. tuberculosis* H37Rv was used to design primers for *Rv1785c* gene amplification. The primers used were: forward primer 5'-GGTACCAGGATCCG**ATG**ACGACCCCGG-3', and reverse primer 5'-GCGCTGGTCG**TGA**GAATTCTGTACAGG-3'. The underlined sequences in the forward and reverse primers indicate engineered *Bam*HI and *Eco*RI restriction endonuclease sites, respectively. The boldface letters indicate the start (ATG) and stop (TAG) codons, respectively. The gene amplification conditions used were 95 °C for 2 min; 30 cycles of 95 °C for 45 s, 62 °C for 30 s and 72 °C for 1.5 min; and then a final polymerisation step of 72 °C for 10 min. The PCR product was then digested using *Bam*HI/*Eco*RI and cloned into the pET47b vector using Quick Ligation™ Kit (New England Biolabs, UK) between the *Bam*HI and *Eco*RI sites to generate the pCYP143A1 construct.

The production of the CYP143A1 protein was achieved by transforming *E. coli* strain BL21 Gold (DE3) (Agilent, UK) with the pCYP143A1 plasmid construct using the heat-shock method. Protein was produced in *E. coli* strain BL21 Gold (DE3) (typically 12 litres, grown in

2xYT medium supplemented with 30 mg/ml kanamycin) using isopropyl β-D-thiogalactopyranoside (IPTG, 0.2 mM) induction in the presence of the heme precursor δ-aminolevulinic acid (0.25 mM) at OD_{600} = 0.5-0.7, with growth at 25 °C. The transformant cells were grown for further 24 hours at 25 °C and then cells were harvested by centrifugation using a Beckman Coulter Avanti J-25 High Performance centrifuge and a JA-10 fixed angle rotor (6000 rpm, 4 °C, and 10 min).

Cell pellets were resuspended in 300 ml of lysis buffer (50 mM potassium phosphate, 500 mM KCl, 10 mM imidazole, 10% glycerol, pH 8.0) with addition of protease inhibitor tablets (SIGMAFAST™ Protease Inhibitor Cocktail Tablets, EDTA-free, Sigma-Aldrich, UK), 10 µg/ml lysozyme and 10 µg/ml DNase (Sigma-Aldrich, UK). The cells were kept on ice at all times and lysed by intermittent sonication 20-25 times with 20 second bursts at full power, and with 60 seconds cooling time between each burst using a Bandelin Sonopuls GM2600 sonicator. The lysate was then centrifuged (14500 rpm, 45 minutes, 4 °C, as described above) and the supernatant was collected.

The supernatant was loaded onto a Nickel-IDA resin (Generon, UK) pre-equilibrated with lysis buffer. The loaded column was washed with 20 column volumes of nickel buffer A (50 mM potassium phosphate, 500 mM KCl, 20 mM imidazole, 10% glycerol, pH 8.0). The protein was eluted using nickel buffer A containing 200 mM imidazole. The eluted protein samples were pooled and dialysed against buffer Q (50 mM Tris-HCl, 1 mM EDTA, pH 7.5). Subsequently, the dialysed protein was loaded onto a Q-Sepharose column (Q-Sepharose Fast Flow resin, GE Healthcare Life Sciences, UK) pre-equilibrated with buffer Q. The column was washed with 2 column volumes of buffer Q and the protein was eluted using a linear gradient of 0-500 mM KCl in buffer Q. Selected eluted protein fractions were pooled and concentrated using Vivaspin protein concentrator spin columns (Vivaspin 20 MWCO 30000, GE Healthcare Life Sciences, UK) to < 1 ml prior to a final purification step using a gel filtration column (HiLoad 16/600 Superdex 200, GE Healthcare Life Sciences, UK) and 10 mM Tris-HCl, 150 mM KCl buffer, pH 7.5. CYP143A1 purity was determined by SDS-PAGE and UV-visible spectroscopy. The purest CYP143A1-containing fractions were pooled and concentrated to ~200 µM, and then stored at -80 °C until required for use.

5.3.2 Mass spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was performed on CYP143A1 (40 µM) using a Xevo G2-S Q-TOF UPLC instrument (Waters, Elstree UK) coupled to an Acquity UPLC system. Samples were passed through an Acquity UPLC BEH300 C4 column (1.7 µm 2.1 x 50 mm) using a mobile phase of Solvent A: water with 0.1% formic acid, and Solvent B: 95% acetonitrile containing 0.01% formic acid. The elution gradient was run using 95% Solution A for 5.21 minutes, 100% Solution B for 1 minute, and 100% Solution A for 1 minute at a flow rate of 0.2 ml/min over a total run time of 7.29 minutes. The electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 40 V. Nitrogen was used as the desolvation gas at a total flow of 850 litres per hour. Data acquisition and processing was performed using Micromass MassLynx v4.1 software with total mass spectra reconstructed from the ion series using the pre-installed MaxEnt algorithm.

5.3.3 UV-visible spectroscopy

UV-visible absorption spectra of the oxidized, reduced and reduced/CO-bound forms of CYP143A1 were recorded on a Cary 50 UV-visible scanning spectrophotometer (Agilent, UK) using a 1 cm pathlength quartz cuvette. Spectra were recorded between 250-800 nm. Typically, purified CYP143A1 was used at ~2-5 µM in 100 mM potassium phosphate, pH 7.0 at room temperature. Binding of carbon monoxide (CO) to the CYP143A1 protein was performed after the reduction of the oxidized protein using a small amount of solid sodium dithionite. 30 to 60 bubbles of CO were added slowly into the ferrous P450 solution inside a fume cupboard to achieve full CYP143A1 complexation with the CO.

5.3.4 Fragment-based screening for substrate- and inhibitor-like ligands

Fragment screens was carried out on a CLARIOstar microplate reader (BMG Labtech, Germany) using UV-star 96-well microplates (Greiner Bio-one, UK) in absorbance mode, according to a previously reported method (9). Heme-focused library fragments (1 mM) were prepared as stock solutions (1-100 mM) in deuterated dimethyl sulfoxide (DMSO-d_6) and

added to solutions of CYP143A1 (4-6 µM) in the appropriate buffer, or to buffer alone, to achieve a final concentration of 1% v/v DMSO-d_6. Any interference from the inherent absorbance of small molecules was removed by subtracting spectra from buffer control cuvettes from those of samples containing CYP143A1. The difference in maximum wavelength of the P450 Soret band ($\Delta\lambda_{max}$) observed in the presence of fragments compared to that of a ligand-free control (1% v/v DMSO-d_6) was used to identify fragments as type I and type II heme binding ligands.

5.3.5 Ligand binding titrations by UV-visible spectroscopy

Ligand binding titrations to determine the binding affinity (K_d values) of ligands were carried out on a Cary 50 UV-visible scanning spectrophotometer (Agilent Technologies, UK) using a 1 cm pathlength quartz cuvette and at 25 ºC. The concentration of CYP143A1 was from 2-10 µM in either 100 mM potassium phosphate, pH 7.0, or in 10 mM Tris-HCl, 150 mM KCl, pH 7.5. Ligands were prepared as stock solutions (10 - 100 mM) in DMSO and titrated as 0.1-0.2 µl aliquots into the 1 ml cuvette containing the buffered solution of CYP143A1, up until the point that the compound became insoluble or where the absorbance spectrum of CYP143A1 no longer changed. Ligands used were azole antifungal drugs and small-molecule fragment compounds. Spectra were recorded between 250 to 800 nm after each addition of ligand. Difference spectra were generated by subtraction of the initial ligand-free CYP143A1 spectrum from each successive ligand-bound spectrum. Binding constants (K_d values) were determined by plotting the maximum change in absorbance for each difference spectrum (ΔA_{peak} - ΔA_{trough}) against the concentration of the ligand used, using the same wavelength pair throughout for each particular titration. Data were fitted using a standard hyperbolic (Michaelis-Menten) equation (Equation 1), the Morrison equation for tight-binding ligands (Equation 2) or the sigmoidal Hill equation (Equation 3), depending on the affinity of the ligand and type of binding curve observed. All data fitting was performed using Origin software (OriginLab, MA). All titrations were repeated in triplicate, and data from representative experiments are reported.

Equation 1: $A_{obs} = (A_{max} \times L\ )/(K_D + L)$

Equation 2: $A_{obs} = (A_{max}/2E) \times \left(\left((L + E + K_D) - ((L + E + K_D)^2) - (4 \times L \times E)\right)^{0.5}\right)$

Equation 3: $A_{obs} = A_{max}(L^n/(K_D + L^n))$

A_{obs} is the observed change in absorbance, A_{max} is the maximum absorbance change at apparent ligand saturation, E is the total enzyme concentration, L is the concentration of ligand and n is number of apparent binding sites.

5.3.6 Nuclear magnetic resonance (NMR) spectroscopy

Fragment screening of an 80 compound rule-of-three compliant library was conducted as cocktails of three fragments by ligand-observed NMR, according to a previously reported method (9). Samples containing fragments alone (200 µl, with fragments at 200 µM), or with fragments plus CYP143A1 protein (20 µM), were prepared in 10 mM Tris-HCl buffer pH 7.5, containing 150 mM KCl. 3-(trimethylsilyl) propionic acid-d_4 (TSP, 20 µM) in D_2O (10% v/v) was added as an internal standard and all samples contained a final DMSO-d_6 concentration of 0.6% v/v. Samples were loaded into 3 mm diameter capillaries for analysis inside 524-PP-8 NMR tubes (Winmad-LabGlass, New Jersey USA) and analysed by ^{1}H NMR employing a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence. Spectra were processed using TopSpin 3.5 software (Bruker, Coventry UK) and scaled relative to the TSP-d_4 peak intensity. A decrease in the intensity of fragment ^{1}H signals of >80% relative to protein-free samples was used to identify fragment hits for the calculation of ligand binding profiles.

5.3.7 Electron paramagnetic resonance (EPR) spectroscopy

EPR spectra of both ligand-free CYP143A1 (200 µM) and CYP143A1 bound to ligands (10-100 mM) were recorded using a Bruker ER-300D series electromagnet and microwave source interfaced with a Bruker EMX control unit. The instrument was fitted with an ESR-9 liquid helium flow cryostat (Oxford Instruments) and dual-mode microwave cavity from Bruker (ER4116DM). Spectra were recorded at a temperature of 10 K, a microwave power of 2.08 mW, and amplitude of 1 mT. Samples were prepared in 10 mM Tris-HCl, 150

mM KCl, pH 7.5 from stock solutions of ligands in DMSO. The Bruker software packages supplied with the EPR instrument were used to calculate the *g*-values for all samples. EPR spectra were collected by Dr. Kirsty McLean and Dr. Hazel Girvan (University of Manchester). Data analyses were performed using Origin software.

5.3.8 Resonance Raman (RR) spectroscopy

CYP143A1 (80 µM) samples were transferred into NMR tubes for measurement using a Renishaw Ramascope 2000 spectrophotometer. RR spectra were measured using 406 nm excitation provided by a Coherent Innova Krypton ion laser. The power from the laser was maintained at 200 mW and controlled through the spectrometer to give 6 mW of power at the sample. Static scans were used to collect data over a spectral range from 200 to 2000 cm^{-1}, with a total accumulation time of 30 min in the high frequency region and 40 min in the low frequency region. Data processing and band assignment of the spectral data were carried out with Grams/Wire software, which accompanies the Renishaw 200 series microscope systems. RR spectra were collected and analysed by Prof. Jim Kincaid and Dr. Yilin Liu (Marquette University, USA).

5.3.9 Crystallization of ligand-free and ligand-bound CYP143A1

All crystallization of CYP143A1 was performed using the sitting drop vapour diffusion method. Initial screening of crystallization conditions was performed on pure CYP143A1 against five commercially available 96-well sparse matrix screens: JCSG-*plus*™ HT-96, PACT *premier*™ HT-96, SG1™ Screen HT-96 and Morpheus HT-96 (all from Molecular Dimensions Ltd., UK). The crystallization drops were comprised of 200 nl of 20 mg/ml CYP143A1 and 200 nl of the reservoir solution that were dispensed by a robotic platform (mosquito® Crystal, TTP Labtech, UK). The crystal trays were sealed and incubated at 4 °C. Clustered and single plate-like rectangular red crystals grew after 3 days to about 50-100 µm in their longest dimensions in most of the crystallization plates.

Attempts to obtain ligand-bound forms of CYP143A1 were performed by soaking and co-crystallization methods. For soaking, each of the ligand-free CYP143A1 crystals obtained in JCSG and PACT crystal screening plates were soaked into a new drop containing the reservoir solution with either a few solid grains or a 2 mM solution of the relevant ligands. The crystal trays were sealed and incubated at 4 ^{o}C for 1 week. All of the soaked crystals maintained their morphology and did not incur significant damage throughout the incubation time. In addition, co-crystallization studies of CYP143A1 with small-molecule fragments were performed by using PACT *premier*™ HT-96 (Molecular Dimensions, UK) crystallization screens. The crystallization drops comprised 250 nl of 20 mg/ml CYP143A1 containing 2-10 mM ligands, and 250 nl of the reservoir solution was dispensed using a robotic platform (mosquito® Crystal, TTP Labtech, UK). The crystal trays were sealed and incubated at 4 ^{o}C. Small cube-like red crystals grew after 5 days to about 50 µm in their longest dimensions in half of the crystallization plates.

5.3.10 Data collection, structure determination and refinement

A single reddish crystal of CYP143A1 was mounted on a 0.05-0.1 mm cryoloop, dipped briefly in paratone oil as cryoprotectant and frozen in liquid nitrogen. X-ray diffraction data were collected at Diamond synchrotron beamline I04 (Harwell UK). The data were indexed and processed with the XDS program package. (10) The initial structure of ligand-free CYP143A1 was solved by a molecular replacement method using PhaserMR (11) from the PHENIX package (12) and CYP142A1, a *M. tuberculosis* cholesterol oxidase P450 (PDB: 2XKR), as the search model. Successive structures of ligand-free or ligand-bound forms of CYP143A1 were solved using the same method, but with the initial ligand-free CYP143A1 model PDB coordinates as the search model. Manual model rebuilding was performed with COOT. (13) Structure refinement and validation were performed with phenix.refine (14) and MolProbity (15) from the PHENIX package, (12) respectively. All molecular graphics were prepared using PyMOL (16) or CCP4mg (17).

5.4 Results

5.4.1 Expression and purification of CYP143A1

CYP143A1 was expressed and purified as a 6xHis–tagged construct, as described in the Materials and Methods. Extraction of CYP143A1 from *E. coli* by sonication and centrifugation resulted in a reddish supernatant and a small amount of red cell pellet, indicating that the enzyme was mostly soluble. Pure CYP143A1 samples were obtained using a three step purification strategy involving nickel affinity, anion exchange and size exclusion column chromatography. Using the Reinheitzahl (Rz) value that is the ratio of absorbance at the heme Soret peak to the protein absorbance at 280 nm ($A_{\sim 420nm}/A_{280nm}$) as the assessment of purity, it was observed that the purity of CYP143A1 increased after each purification step (Figure 5.2A). In addition, the heme Soret peak for CYP143A1 after affinity chromatography was located at 422 nm, while the heme Soret peak for CYP143A1 after anion exchange and size exclusion chromatography was located at 417 nm, resulting from the displacement of heme iron-bound imidazole following the final two purification steps (Figure 5.2A). Typically, purified CYP143A1 had Rz values ranging from ~1.6 to 1.8. Furthermore, the purity of CYP143A1 was also demonstrated by SDS-PAGE analysis with the appearance of a single major band on the gel (Figure 5.2B). The apparent molecular mass of this band, based on comparisons with the standard molecular markers, is 46 kDa. Consequently, the apparent mass of CYP143A1 was confirmed by mass spectrometry to be 46109 Da (Figure 5.2C). This value is consistent with the expected mass from the amino acid sequence of the 6xHis-tagged CYP143A1, minus the initial methionine residue.

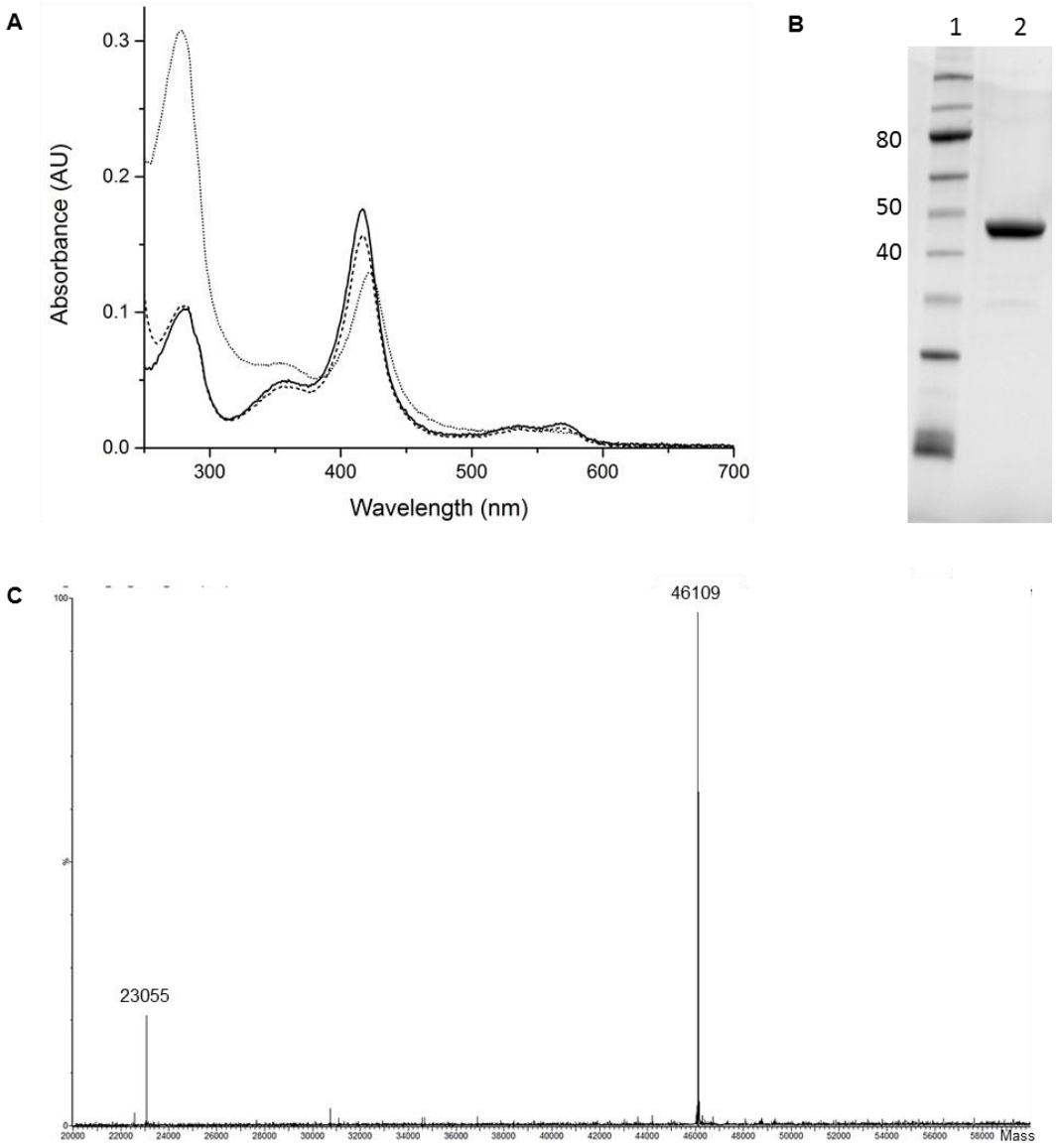


Figure 5.2 Purification of CYP143A1.
(A) UV-visible absorption spectra of CYP143A1 (1-2 µM) after successive chromatographic purification steps: nickel affinity (dotted line), anion exchange (dashed line) and size exclusion (solid line). The relative absorbance of the heme Soret peak (at ~420 nm) increases and the absorbance of total protein (at 280 nm) decreases after each purification stage, indicating a progressive improvement in CYP143A1 purity. (B) SDS-PAGE analysis of the purified CYP143A1, showing standard molecular mass markers (lane 1, NEB Unstained Protein Ladder, broad range (10-250 kDa), bands labelled in kDa) and pure CYP143A1 as a single band (lane 2) with apparent mass of 46 kDa when compared to the molecular mass marker. (C) Accurate mass spectra of CYP143A1 (40 µM) obtained by mass spectrometry under denaturing conditions indicates the mass of the 6xHis-tagged CYP143A1, minus the initial methionine residue, is 46109 Da.

5.4.2 Spectroscopic characterisation of ligand-free CYP143A1

UV-visible spectra for the oxidised and reduced forms of the ligand-free CYP143A1, and for reduced CYP143A1 bound to carbon monoxide (CO), are shown in Figure 5.3. The

UV-visible spectrum of oxidised, ligand-free CYP143A1 has a major heme absorbance peak at 417 nm (Soret) and minor absorbance peaks for the α and β bands are located at 574 and 537 nm, respectively. This indicates that the CYP143A1 ferric heme iron is predominantly low spin with water as its sixth axial ligand. Reduction of CYP143A1 with sodium dithionite results in a Soret band shifts to 414 nm. The binding of CO to the reduced CYP143A1 results in the formation of two ferrous-CO complex species, with Soret features at 423 nm and 448 nm, and minor peaks for the Q-band features at 548 and 574 nm. The spectral features of the ferrous-CO complex of CYP143A1 are consistent with the proximal axial ligation of the P450 heme iron by cysteine in both its thiolate (P450) and thiol (P420) forms (18).

X-band EPR spectroscopy was performed to analyse the CYP143A1 heme iron coordination state and the homogeneity thereof. EPR spectra for ligand-free CYP143A1 (Figure 5.3) displayed a single major low-spin EPR signal with *g*-values of 2.42 (g_z), 2.26 (g_y), and 1.92 (g_x). This indicates that CYP143A1 is in a low-spin, ferric state, and that its heme iron is coordinated by a cysteine thiolate ligand. The homogeneous nature of the EPR spectrum suggests that there is a single, dominant coordination state. There is no evidence for a high spin species in this sample, consistent with the UV-visible spectrum. However, the low temperatures used for EPR spectroscopy of heme proteins (10 K in this case) also favour P450s forming the low spin state.

In addition, resonance Raman (RR) spectroscopy was used to further probe the heme iron oxidation state, spin-state and aspects of heme iron coordination and heme prosthetic group conformation. The spectrum of ligand-free CYP143A1 in the high frequency region (Figure 5.3) shows the main peak at 1373 cm^{-1}, assigned as the oxidation state marker band, v_4, indicating that the heme iron is in ferric state. The spin-state marker bands, v_3, v_2 and v_{10} were observed at 1502, 1585 and 1637 cm^{-1}, respectively, showing the heme iron is in a 6-coordinate, low spin state, likely reflecting cysteine thiolate and water axial ligands. The CYP143A1 $v_{c=c}$ band at 1623 cm^{-1} is assigned to in-plane stretching of the heme 2- and 4-vinyl groups, and likely indicates that these substituent groups are largely co-planar with the heme itself.

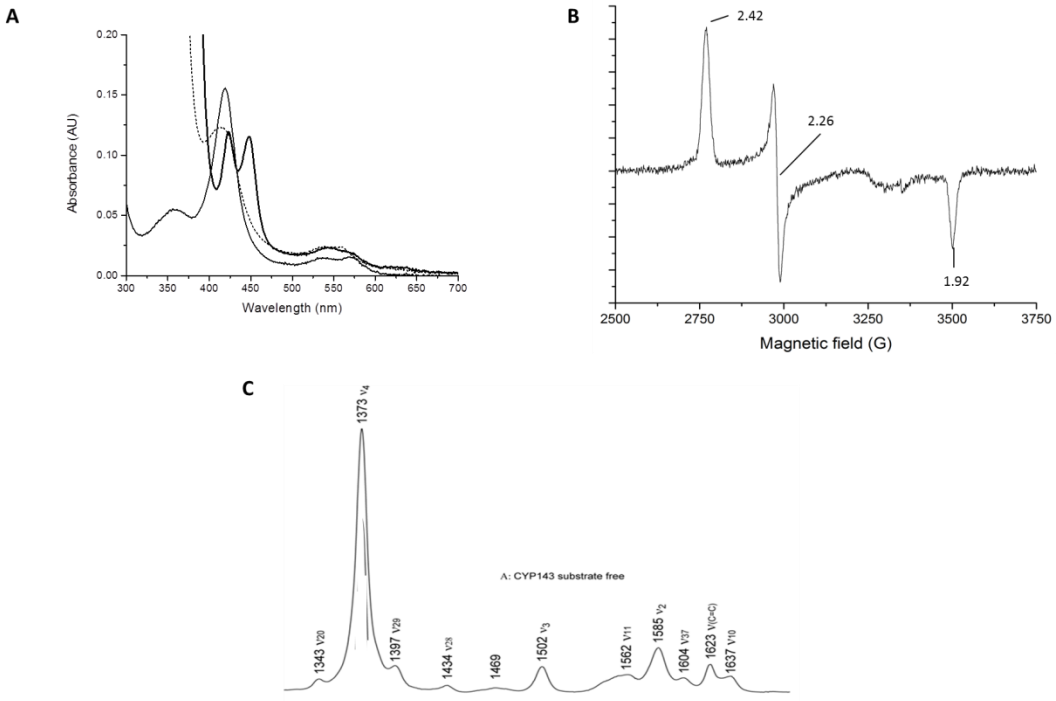


Figure 5.3 Spectroscopic characterisation of ligand-free CYP143A1.
(A) UV-visible absorption spectra of oxidised (ferric) CYP143A1 (1.5 µM, thin line), sodium dithionite-reduced (ferrous) CYP143A1 (dashed line), and reduced (ferrous) CO-bound CYP143A1 (thick line). The Soret peak wavelengths are at 419, 414, and at 423/448 nm, respectively. The ferrous-CO complex has P420 and P450 features consistent with the proximal ligation of the heme iron by both cysteine thiol and cysteine thiolate in equilibrium. (B) X-band EPR spectra for CYP143A1 (200 µM) with *g*-values of 2.42/2.26/1.92. (C) Resonance Raman spectrum for CYP143A1 (80 µM) with laser excitation at 406 nm. The major ν_4 feature at 1373 cm^{-1} indicates a ferric oxidation state of the heme iron.

5.4.3 Spectroscopic characterisation of azole-bound CYP143A1

Like most Mtb P450 enzymes, CYP143A1 binds azole antifungal drugs, which occupy the active site cavity and coordinate to the heme iron (usually through nitrogen atoms from imidazole or triazole functional groups) to produce characteristic Type II spectral shifts with a peak between 420-422 nm and trough at 385-390 nm, as shown in Figure 5.4. The dissociation constant values for CYP143A1 with azoles were obtained from spectral titration curves (Figure 5.5) and are summarised in Table 5.1. Among the azole drugs tested, the highest affinity was observed for bifonazole, miconazole, econazole and clotrimazole (K_d values of 2.5 ± 0.3, 2.8 ± 0.4, 4.5 ± 0.4 and 5.7 ± 0.5 µM, respectively). No significant binding to CYP143A1 was observed for the more polar and bulky ketaconazole, itraconazole and voriconazole drugs. Binding of various phenylimidazoles (1-PIM, 2-PIM and 4-PIM) to

CYP143A1 resulted in relatively weak binding of 1-PIM (K_d = 25.7 ± 1.4 μM) and no significant spectral changes induced by the binding of 2-PIM and 4-PIM. The binding of 1-PIM to CYP143A1 induces a type II inhibitor-like spectral change, with a decrease in the intensity of the Soret band in the 1-PIM complex and a shift of the Soret maximum to 420 nm. The titration curves for bifonazole, clotrimazole and 1-PIM were fitted using a standard hyperbolic function (Equation 1), while the sigmoidal titration plots obtained for econazole and miconazole were best fitted using the Hill equation (Equation 3), with Hill coefficients of approximately 1.8 obtained in both cases.

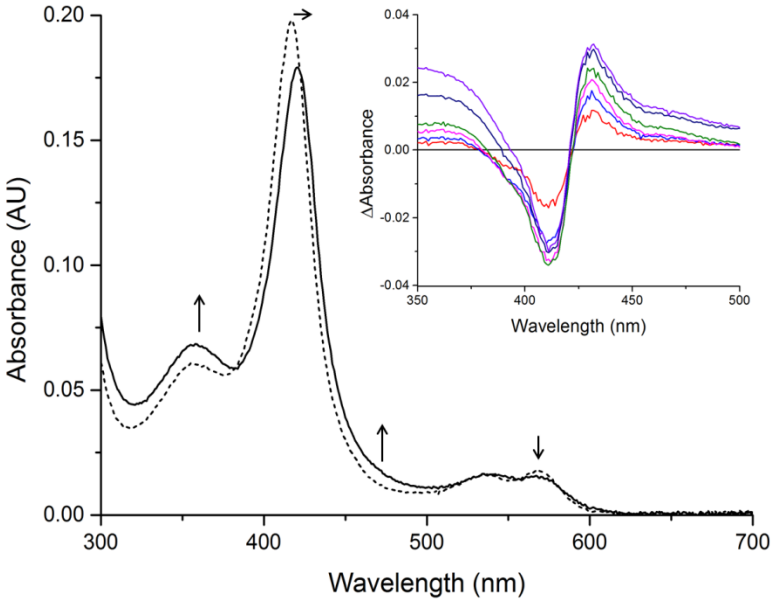


Figure 5.4 UV-visible absorption spectra of CYP143A1 before and after addition of bifonazole.
UV-visible absorption spectra of ligand-free CYP143A1 (2 μM, dashed line) and for bifonazole-bound CYP143A1 (solid line). The major Soret peak shifts from 417 nm to 421 nm upon binding of the azole. The inset shows difference spectra generated by subtracting the spectrum for ligand-free CYP143A1 from selected spectra for bifonazole-bound forms of CYP143A1 collected during a bifonazole titration. Difference absorption maxima and minima are at 432 nm and 411 nm, respectively.

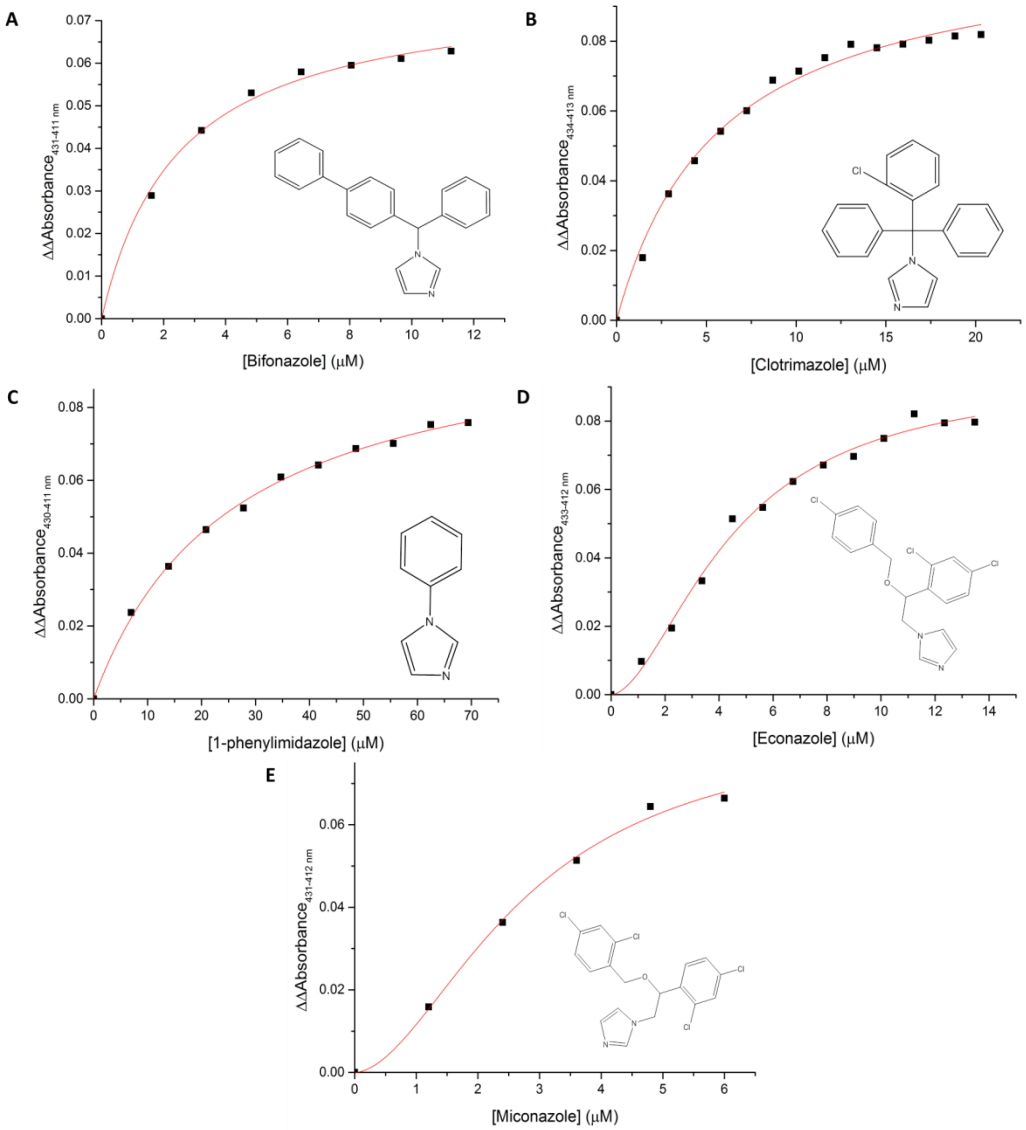


Figure 5.5 Binding of azole antifungal drugs to CYP143A1.
Plots of azole-induced absorption differences were obtained for each spectral data set, and ΔA_{max} minus ΔA_{min} values were calculated, using the same wavelength pair throughout in individual titrations. These data were plotted against the relevant concentration of each azole ligand, and data were fitted using either a standard (Michaelis-Menten) hyperbolic equation (bifonazole, clotrimazole and 1-phenylimidazole in panels A-C) or to the Hill function where a sigmoidal dependence was observed (econazole and miconazole in panels D and E). CYP143A1 was at 2.5 μM in each titration. The structures of the azole drugs are shown as insets in each plot. The K_d values are presented in Table 5.1.

Table 5.1 Dissociation constants for azole antifungals drugs binding to CYP143A1. The K_d values were determined by spectral titration. Data were fitted using either a hyberbolic equation (Equation 1), or using the Hill equation (Equation 3) in the cases of econazole and miconazole, where a sigmoidal dependence of induced absorbance change with azole drug concentration occurred. The n values from both Hill plots were ~1.8. The azoles specified as NB (no binding) gave no detectable spectral change upon their addition to CYP143A1.

Azole drugs	Soret wavelength maximum obtained	K_d (μM)
Bifonazole	421 nm	2.5 ± 0.2
Clotrimazole	421 nm	5.7 ± 0.5
Econazole	420 nm	4.5 ± 0.4
Fluconazole	NB	NB
Itraconazole	NB	NB
Ketaconazole	NB	NB
Miconazole	420 nm	2.8 ± 0.4
Voriconazole	NB	NB
1-PIM	420 nm	25.7 ± 1.4
2-PIM	NB	NB
4-PIM	NB	NB

EPR spectra were also collected for azole-bound CYP143A1 complexes to examine the effects of azole drug nitrogen-coordination to the heme iron. In all cases, the EPR spectra obtained exhibit a characteristic P450 rhombic spectrum indicative of predominantly low-spin, ferric heme iron (Figure 5.6) with the *g*-values of 2.42 (g_z), 2.26/2.25 (g_y) and 1.92/1.91 (g_x). For azole-bound CYP143A1 complexes, except for miconazole, the EPR spectra revealed a small proportion of the enzyme that exhibited additional species with *g*-values consistent with changes to the sixth ligand environment of the heme iron. These minor species have sets of *g*-values at 2.52/2.25/1.87 for bifonazole-bound CYP143A1; 2.49/2.25/1.88 for clotrimazole-bound CYP143A1; and 2.47/2.25/1.89 for 1-phenylimidazole-bound CYP143A1. Distal coordination of the CYP143A1 heme iron by imidazole nitrogen likely occurs in these minor species.

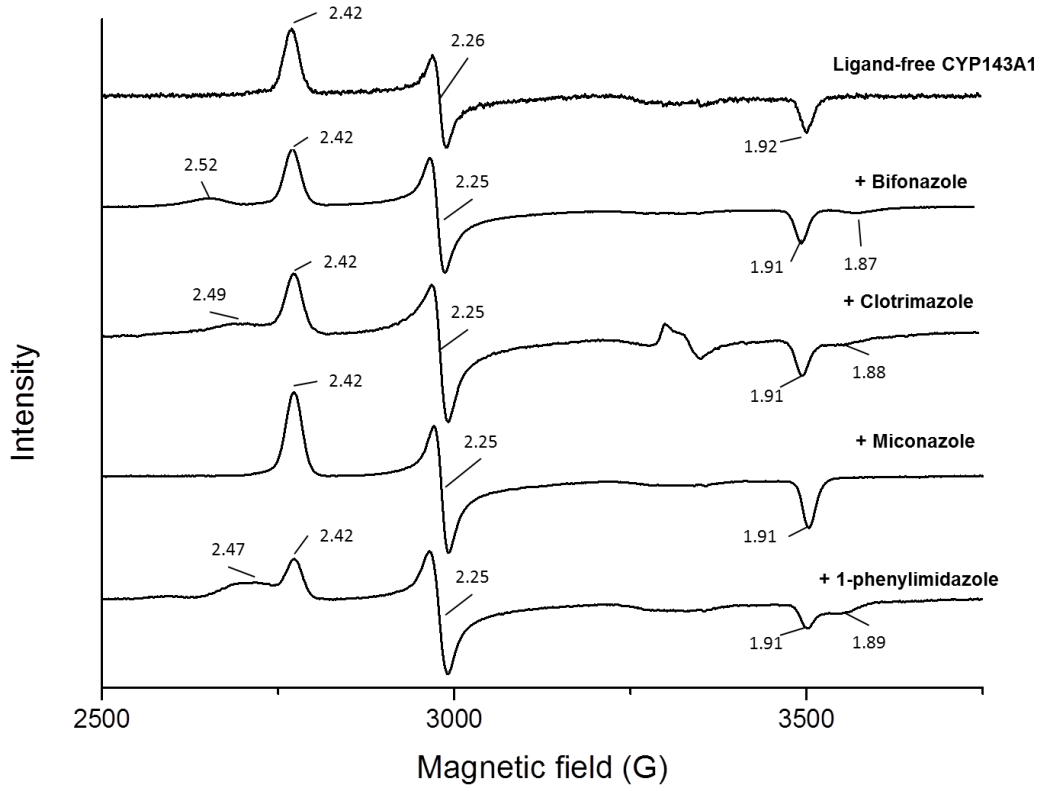


Figure 5.6 EPR spectra for CYP143A1 in complex with 1-phenylimidazole and selected azole antifungal drugs.
X-band EPR spectra were collected for CYP143A1 in its ligand-free form and for its complexes with bifonazole, clotrimazole, miconazole and 1-phenylimidazole. The low spin *g*-values are indicated in each spectrum.

5.4.4 Spectroscopic characterisation of fragment-bound CYP143A1

Initial fragment screening and validation was performed using UV-visible heme absorbance assays and Carr-Purcell-Meiboom-Gill (CPMG) ligand-observed NMR spectroscopy, respectively, in order to identify fragment hits for CYP143A1. Out of 160 fragments in the heme-focused and unbiased NMR library tested against CYP143A1, 10 fragments induced a heme Soret shift and were observed to bind to CYP143A1 (Figure 5.7). An additional 7 fragments did not perturb the heme Soret band, but do bind to CYP143A1 as validated by NMR spectroscopy (Figure 5.8). These fragment hits are as listed in Table 5.2.

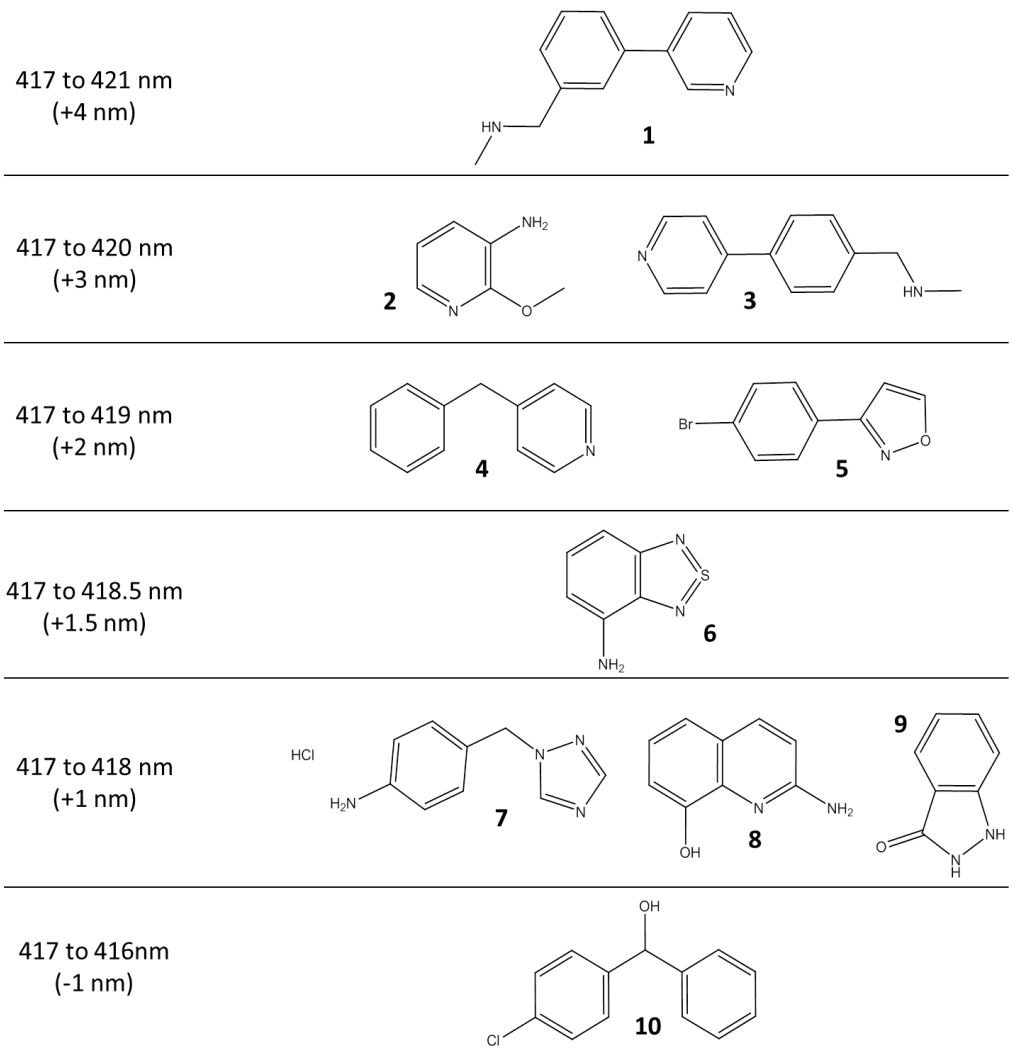


Figure 5.7 Fragment hits for CYP143A1.
Fragment hits from initial fragment screening studies using (i) a UV-visible spectrophotometric heme absorbance shift assay and (ii) CPMG ligand-observed NMR spectroscopy. Fragments shown in this image are those that perturb the Soret peak position of CYP143A1 and are grouped based on the Soret shift induced.

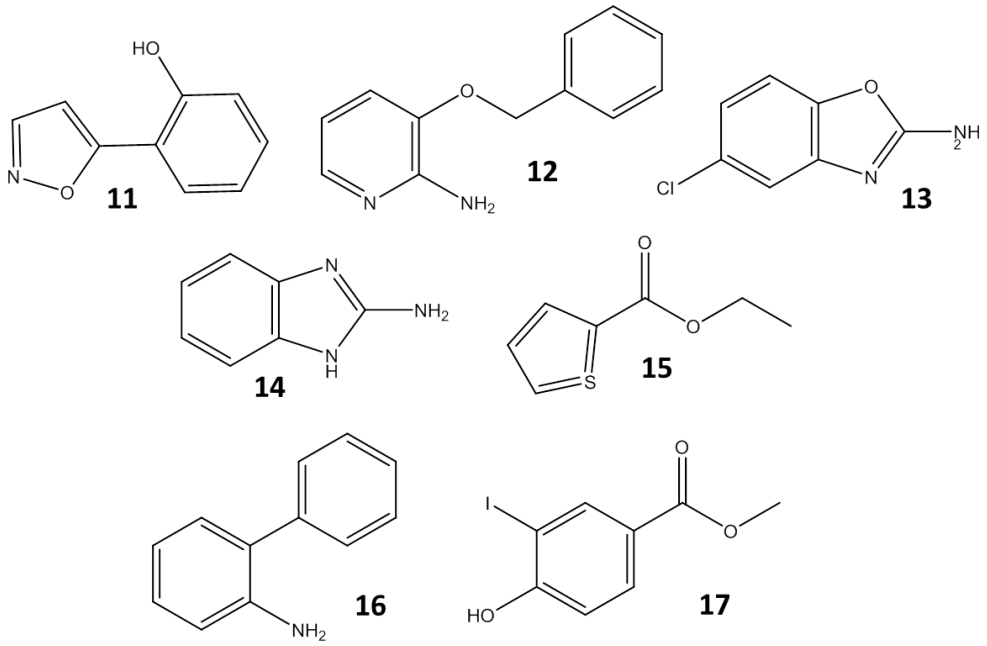


Figure 5.8 Fragments hits for CYP143A1.
Fragment hits are shown from initial fragment screening studies using (i) a UV-visible spectrophotometric heme absorbance shift assay and (ii) CPMG ligand-observed NMR spectroscopy. The fragments displayed did not cause any significant change in the CYP143A1 Soret peak position, but were shown to bind CYP143A1 by NMR analysis.

Table 5.2 List of fragment hits for CYP143A1.
Fragments **1-10** induced CYP143A1 heme Soret shifts in the UV-vis absorbance assay and were also confirmed to bind to CYP143A1 by NMR spectroscopy. Fragments **11-17** were confirmed to bind to CYP143A1 by NMR spectroscopy, but showed negligible effects on the CYP143A1 absorbance spectrum.

Fragment	Identifier	Name
1	NMR195	N-methyl-(3-pyridine-3-benzyl) amine
2	NMR410	2-methoxypyridin-3-amine
3	NMR415	N-methyl-(3-pyridine-3-benzyl) amine
4	NMR170	4-benzylpyridine
5	NMR012	3-(4-bromophenyl)isoxazole
6	NMR054	4-amino-2,1,3-benzothiadiazole
7	NMR015	(1,2,4-triazol-1-ylmethyl) aniline hydrochloride
8	NMR034	2-amino-8-quinolinol
9	NMR070	Bis(4-hydroxyphenyl) methane
10	NMR056	4-chlorobenzhydrol
11	NMR026	2-(5-isoxazolyl) phenol
12	NMR028	2-amino-3-benzyloxypyridine
13	NMR032	2-amino-5-chlorobenzoxazole
14	NMR035	2-aminobenzimidazole
15	NMR036	2-aminobiphenyl
16	NMR072	Ethyl-2-thiophenecarboxylate
17	NMR078	Methyl 4-hydroxy-3-iodobenzoate

Fragment hits were characterised using UV-visible absorption titrations to confirm their binding affinity for CYP143A1. As expected for the fragment hits that did not induce any heme Soret shift in the earlier heme absorbance assay, there were no spectral changes observed for these fragments (fragments **11**-**17**) when titrated into CYP143A1. However, out of the 10 fragment hits that induced heme Soret shifts (referred to as 'heme-binding fragments') in the initial study, only 4 fragments gave sufficient heme absorbance changes to enable the determination of K_d values using a UV-visible absorbance binding assay. These are heme-binding fragments **1**, **2**, **3** and **4** and all bind to CYP143A1 by interacting with the heme iron through their amine or pyridine nitrogens to produce characteristic Type II spectra with Soret peaks between 419-421 nm and troughs at ~380 nm observed in the UV-

visible absorption spectra. The dissociation constant values for CYP143A1 with these heme-binding fragments were obtained from the fitting of their spectral titration curves (Figure 5.9). Heme-binding fragment **2** binds tightest with a K_d value of 14.4 ± 0.5 µM, followed by fragments **1**, **3** and **4** with K_d values of 56.3 ± 1.9, 224.4 ± 19.1 and 269.4 ± 27.6 µM, respectively. Heme-binding fragment **4** data were fitted most accurately using the Hill equation (Equation 3), with a coefficient (n) value of 1.6.

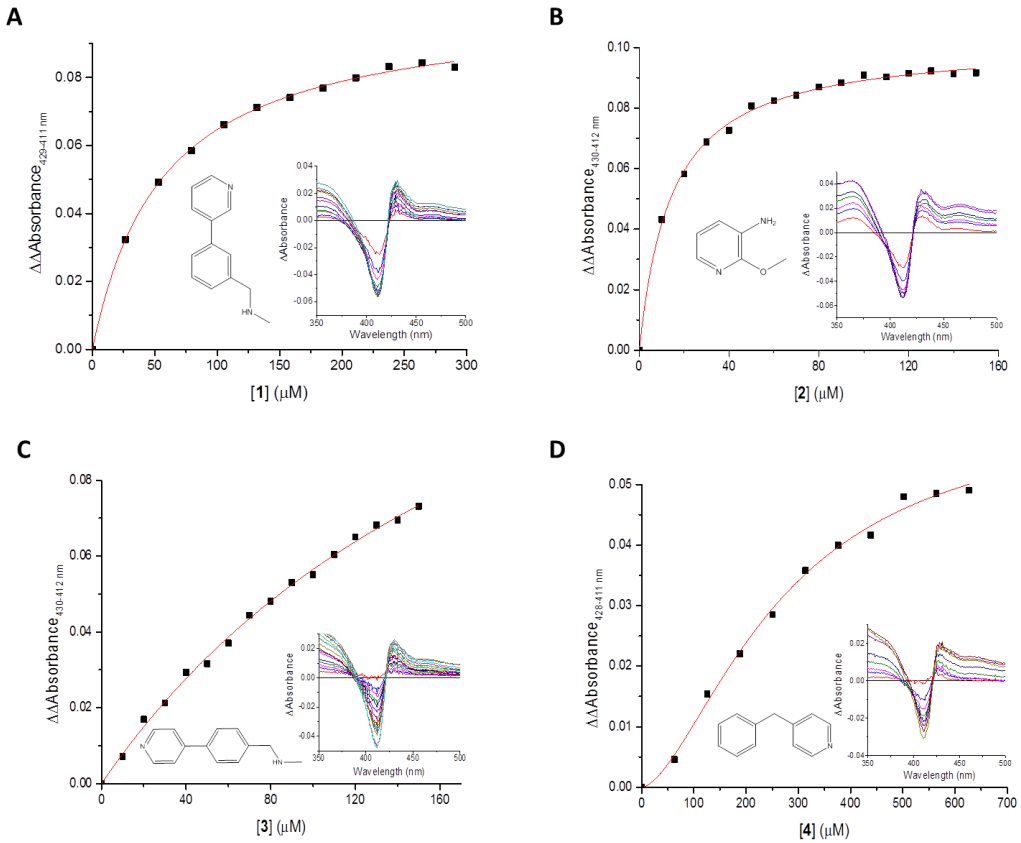


Figure 5.9 Binding of heme-binding fragment hits to CYP143A1.
Panels A-D show fitted data from spectral titrations of CYP143A1 (4 µM) with heme-binding fragments **1-4**, respectively. The insets show difference absorption spectral sets for the relevant titrations and the structures of the fragments used in each case. Data were fitted using either a standard hyperbolic (Michaelis-Menten) function (Equation 1) or the Hill function (Equation 3) in the case of fragment 4. The K_d values determined are (1) 56.3 ± 1.9 µM; (2) 14.4 ± 0.5 µM; (3) 224.4 ± 19.1 µM; and (4) 269.4 ± 27.6 µM.

Subsequently, X-band EPR spectroscopy was performed on the complexes of CYP143A1 with these heme-binding fragments in order to analyse their mode of ligation to CYP143A1. In all cases, the EPR spectra obtained exhibit characteristic P450 rhombic features that indicate predominantly low-spin, ferric heme iron (Figure 5.10), with EPR

spectra in most cases dominated by species with *g*-values of ~ 2.42/2.25(2.26)/1.91. However, heme-binding fragment **2** produced a significant change to the CYP143A1 EPR spectrum and the major signal observed was for a new LS species with *g*-values of 2.45/2.25/1.90. A minor g_z signal at 2.42 likely reflects a small proportion of the resting (fragment-free) form of CYP143A1. In this case, there is clear evidence for the distal coordination of the CYP143A1 heme iron by pyridine or amine nitrogen of fragment **2**. Also, fragment **6** displayed a minor species with a similar set of *g*-values to the major species in the fragment **2** spectrum (2.45/2.26/1.91), suggesting that there is also partial coordination of the CYP143A1 heme iron, probably through the amine nitrogen of this molecule. For the other fragments, it may be the case that binding of fragments is limited by their solubility, leading to poor occupancy of the P450 active site.

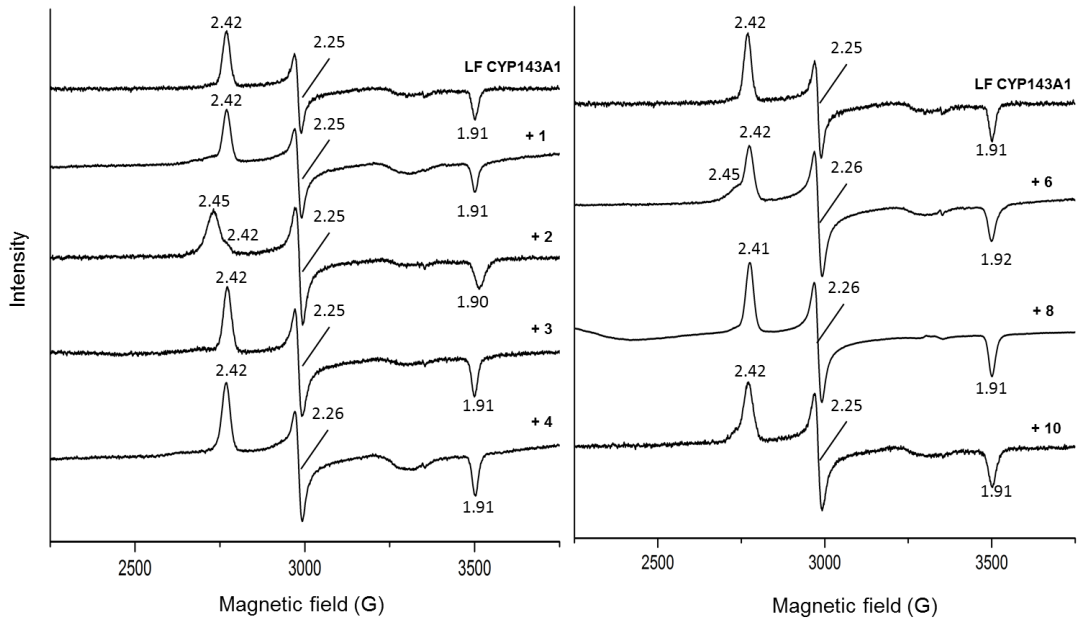


Figure 5.10 EPR spectra for CYP143A1 in complex with fragment hits.
EPR spectra are shown for CYP143A1 (200 μM) in its ligand-free (LF) form, and for complexes with fragment hits. The *g*-values for the low-spin forms of the enzyme are indicated in each spectrum.

5.4.5 Structural characterisation of ligand-free CYP143A1

Ligand-free CYP143A1 crystallized as a monomer, consistent with its behaviour in solution, and diffracted to 1.56 Å. The structure of ligand-free CYP143A1 was solved by molecular replacement using the structure of the CYP142A1 (PDB code 2XKR), a

cholesterol 27-oxidase P450 in Mtb. (19) The ligand-free CYP143A1 structure exhibits the characteristic P450 fold, which resembles a triagonal prism. There are two distinct structural domains with the heme sandwiched between a smaller β sheet-rich domain and a larger α helix-rich domain, which is typical for P450s (Figure 5.11A). Superimposing the ligand-free CYP143A1 and CYP142A1 structures resulted in a root mean square deviation of 1.99 Å over 348 C_α atoms. Almost all secondary structural elements are in similar conformations, except for differences occurring in the conformation of the flexible B'C and FG loops, as well as in the orientation of the F- and G-helices (Figure 5.11B). The B'C and FG loops of ligand-free CYP143A1 have a more "open" conformation than the B'C and FG loops of CYP142A1. The distance between the CYP143A1 B'C and FG loops was measured at 15.2 Å, while the distance between CYP142A1's B'C and FG loops was measured at 7.8 Å. Hence, the CYP143A1 structure is in a relatively "open" conformation with respect to substrate/ligand entry. This is further illustrated by the surface representations of the ligand-free CYP143A1 and CYP142A1 proteins. The route for CYP143A1 substrate entry/product exit to and from the active site is more accessible and unhindered compared to that for CYP142A1. Nonetheless, both the CYP143A1 and CYP142A1 active site access channels are lined by predominantly hydrophobic residues (Figure 5.11C).

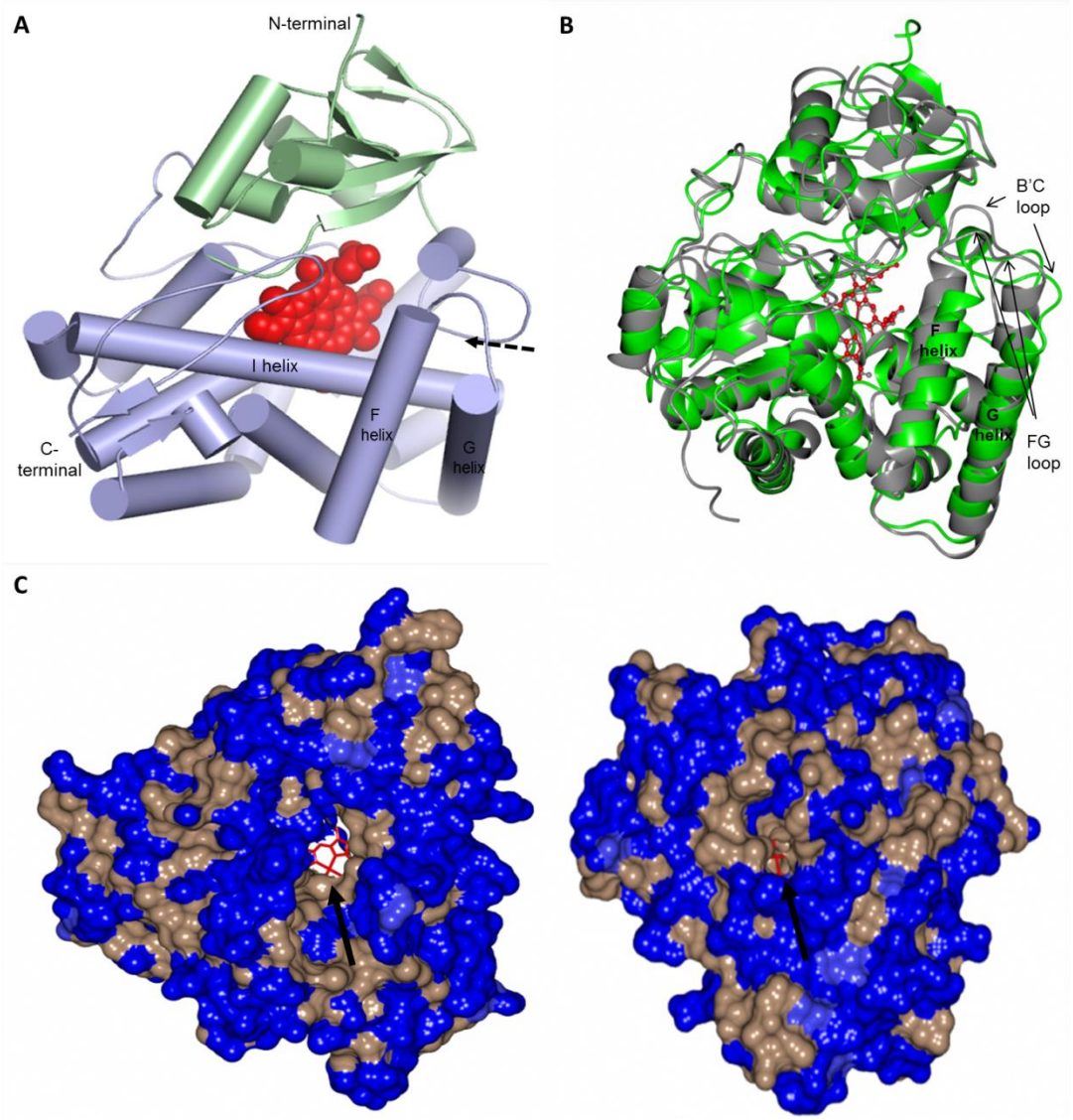


Figure 5.11 Structural features of CYP143A1.
(A) Overall fold of Mtb CYP143A1. The heme (red, spacefill) is sandwiched between the two major domains of the structure; an α helix-rich domain (ash blue) and the smaller β sheet-rich domain (green). An arrow indicates the possible site of substrate entry between the B'C and FG loops. (B) The overlay of Mtb CYP143A1 (green) with CYP142A1 (grey). The heme of CYP143A1 is shown in red ball and stick form. (C) The surface representation of CYP143A1 (left) and CYP142A1 (right) hydrophobicity properties. Hydrophilic and hydrophobic residues are shown in blue and brown, respectively. The arrow indicates the access site entry for CYP143A1 and CYP142A1 ligands. The heme of CYP143A1 (red) is readily accessible as the entry site opening is much more open as compared to that of CYP142A1 (red), where the heme is not clearly visible in the same orientation.

Despite the differences in loop conformations, the CYP143A1 region surrounding the heme cofactor is highly similar to that of CYP142A1. As for other cytochrome P450 enzymes, the proximal and distal axial ligands of the CYP143A1 heme iron are a cysteine (Cys366) and a well-defined water molecule, respectively. The CYP143A1 Fe-S distance is 2.54 Å and the Fe-O distance is 2.25 Å. Both these values are similar to those in other P450

structures obtained. The planar heme cofactor observed with a distal water molecule indicates that the heme iron is in the low-spin state. In addition, many residues in the CYP143A1 active site environment are conserved in identity and relative position with respect to those in CYP142A1. Conserved residues in the I-helix are Leu252, Thr260 and Ala372. In the structural overlay of ligand-free CYP143A1 and CYP142A1 (Figure 5.12), residues Leu108, Ala256 and Ala303 of CYP143A1 are replaced by Ile76, Gly230 and Val277 in CYP142A1. Moreover, Met280 and Phe380 in CYP142A1 are not observed in the proximity of the CYP143A1 distal heme pocket and are likely to have roles specific to CYP142A1 catalysis. The residues Arg119/Arg87, His115/His83, His364/His338 and Arg308/Arg282 are conserved in CYP143A1/CYP142A1, forming ionic/hydrogen bonding interactions with the propionate groups of the heme prosthetic group.

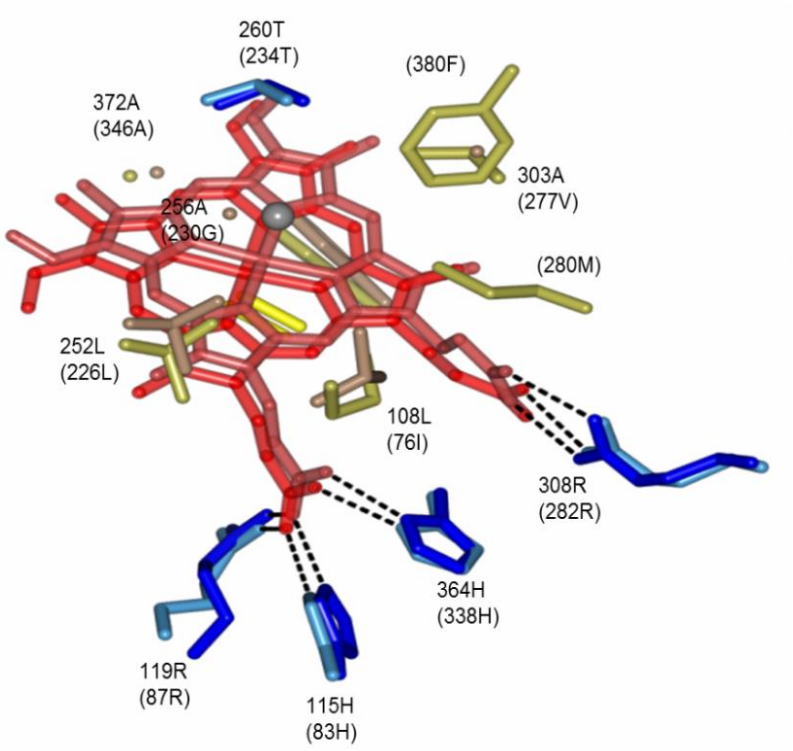


Figure 5.12 Active site structure of ligand-free CYP143A1.
The image shows an overlay of selected residues surrounding the heme cofactor in the ligand-free CYP143A1 active site with those residues in the CYP142A1 active site. The selected residues surrounding the active site are all within a 4 Å radius of the heme iron. CYP143A1 residues are shown in brown and blue for hydrophobic and hydrophilic residues, respectively; the heme cofactor is in dark red; the cysteine-thiolate proximal ligand is in yellow; and a grey sphere represents the sixth distal water ligand. CYP142A1 residues are shown in gold and light blue for hydrophobic and hydrophilic residues, respectively; the heme cofactor is in red; and the cysteine-thiolate ligand is in light yellow. The CYP143A1 residue single letter names and numbers are shown, with their CYP142A1 counterpart residues shown in brackets. Interactions between conserved residues and the heme propionates are shown as dashed lines.

5.4.6 Structural characterisation of ligand-bound CYP143A1

Even though the ligand-free CYP143A1 crystal structure is in an “open” conformation and the active site access channel is exposed, obtaining crystal structures of CYP143A1 in complex with azole drugs or fragment hit compounds was not successful, either by soaking native crystals or by performing co-crystallization experiments. In most cases, the presence of some electron density was observed near the heme iron, but this could not be fitted accurately with the structures of the soaked or co-crystallized ligands, or even with solvent atoms. Instead, electron density in the distal pocket was found to be due to a partially ordered polyethylene glycol (PEG) molecule derived from the crystallization condition, with

the oxygen of its terminal alcohol group binding either directly to the heme iron, or via the sixth water ligand. The coordination of PEG to the heme iron was also observed for CYP142A1. In the structural overlay of PEG-bound CYP143A1 and CYP142A1 (Figure 5.13), both PEGs are located within the distal heme pocket but exhibit different coordination and binding modes. In CYP143A1, PEG coordinates to the heme iron via the sixth water ligand and occupies a region that is sterically impossible in the CYP142A1 structure due to the presence of the Met280 and Phe380 residues, which are absent in the CYP143A1 structure. In addition, binding of PEG does not cause any significant conformational change to the CYP143A1 structure and the active site access channel remains wide and accessible (Figure 5.14).

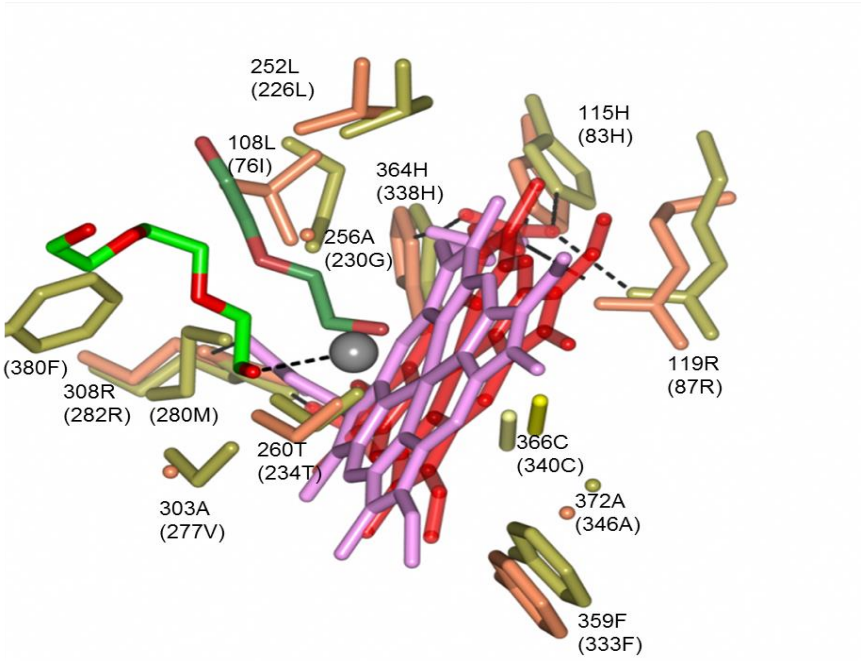


Figure 5.13 Residues surrounding the PEG-bound CYP143A1 active site.
The image shows an overlay of residues surrounding the polyethylene glycol (PEG)-bound CYP143A1 active site with those residues observed in the PEG-bound CYP142A1 active site. The selected residues surround the active site within a 4 Å radius of the heme iron. CYP143A1 residues are shown in orange, with the heme prosthetic group in pink; the cysteine-thiolate ligand in yellow; the distal water ligand as a grey sphere; and the PEG ligand in bright green/red. The CYP142A1 residues are shown in gold, with the heme in pink; the cysteine thiolate in light yellow; and the PEG in dark green/red. The residues and their single letter names and numbers are shown, with the corresponding CYP142A1 residues shown in brackets. Hydrogen/ionic bonds are shown as dashed lines.

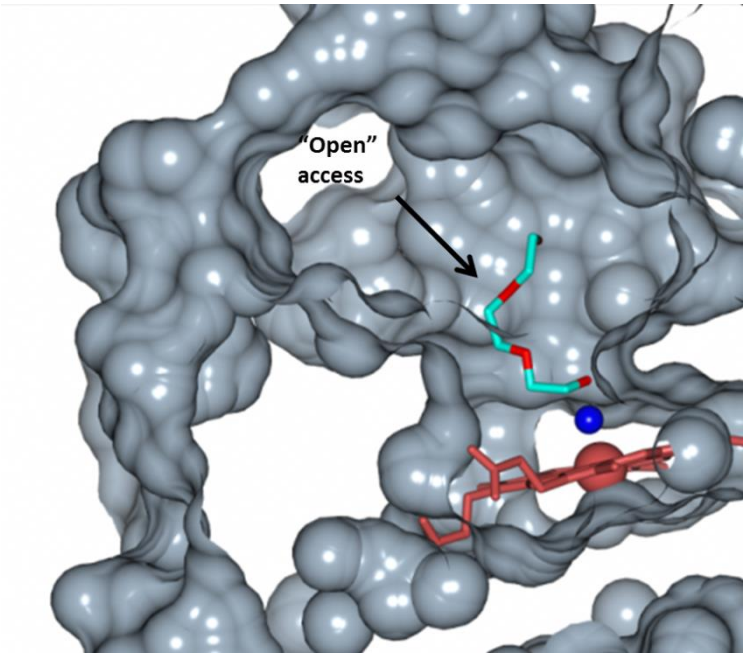


Figure 5.14 Active site access channel of PEG-bound CYP143A1.
A cross-section view of the solvent accessible surface for PEG-bound CYP143A1 shows a wide open active site access channel towards the heme (red). PEG and the sixth water ligand are shown in cyan/red and dark blue, respectively.

However, in an attempt to obtain the crystal structure of CYP143A1 in complex with heme-binding fragment **2** (which is the tightest binding fragment hit for CYP143A1), a different electron density feature was observed near the heme iron. This density does not fit well with the "expected" PEG molecule and was fitted with the structure of fragment **2**. After many rounds of refinement, the electron density for this fragment remains incomplete and it displays only partial occupancy of the active site in its major binding mode (Figure 5.15). Therefore, it is suggested that fragment **2** either does not fully occupy the heme distal pocket, or else that there are multiple binding modes for fragment **2**.

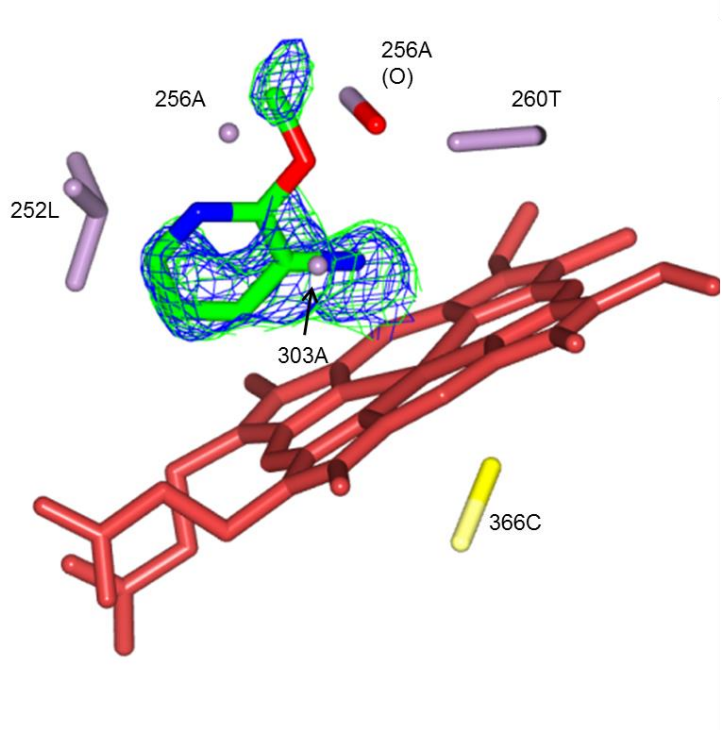


Figure 5.15 Residues surrounding fragment 2-bound CYP143A1 active site.
Electron density for fragment **2** in its CYP143A1 complex structure was obtained by co-crystallization of CYP143A1 with the ligand. The incomplete density for fragment **2** suggests that this molecule does not fully occupy the heme distal pocket, or else that there are multiple different binding modes. Sigma-averaged 2Fo-Fc and difference Fo-Fc electron density maps are both contoured at 0.3 e/Å^3 using CCP4mg.(17) Hydrogen bonds are shown as dashed lines.

5.5 Discussion

M. tuberculosis CYP143A1 was expressed in *E. coli* as a soluble enzyme, indicating that CYP143A1 is likely cytosolic in the Mtb bacterium. All characterized Mtb P450s to date have been expressed as soluble enzymes, (19–24) except for CYP128A1 which is likely to be membrane associated, as its dihydromenaquinone substrate is hydrophobic and located in the Mtb cell membrane. (25) This assumption was also supported by the fact that CYP128A1 is difficult to express in a soluble form and since there is no biochemical characterization for this enzyme available to date. The molecular mass of CYP143A1 is 46 kDa, making it an average sized P450 among the 20 Mtb P450s. The masses of the Mtb P450s ranges between 43 kDa for the structurally characterized cyclodipeptide synthase CYP121A1 (21) through to 56 kDa for the as yet uncharacterized CYP136A1.

During purification of CYP143A1, it was observed that its heme Soret peak position after the later anion exchange and size exclusion chromatography purification steps was different from the Soret position recorded immediately after the initial nickel affinity chromatography step. The Soret peak for the latter is slightly red shifted at 422 nm, due to the presence of imidazole in the elution buffer that readily binds to the CYP143A1 ferric heme iron. Extensive dialysis prior to anion exchange chromatography and successive purification steps led to the complete removal of the imidazole from the CYP143A1 heme iron and a resultant shift of the Soret peak to 417 nm, its native position. The CYP143A1 UV-visible spectrum with Soret maximum at 417 nm is characteristic of a ligand-free, ferric enzyme with cysteine thiolate and water axial ligands, and shows similar spectral properties to other characterized Mtb P450s, such as CYP121A1 with its Soret maximum at 416.5 nm (21) and CYP130A1 at 418 nm (22). An outlier is Mtb CYP144A1 that has its resting Soret band at 420.5 nm (23).

Upon heme iron reduction and binding with CO, the CYP143A1 ferrous-CO complex forms a mixture of the P450 (thiolate-ligated) and P420 (thiol-ligated) states. This is not atypical for P450 enzymes, some of which undergo a transition from the Fe^{II}-CO P450 state to the P420 state due to the instability of the former species to cysteine thiolate protonation. A good example is the Mtb CYP51B1 sterol demethylase enzyme, which undergoes a thiolate-to-thiol (P450-to-P420) transition of the proximal cysteine ligand, although the rate of this process can be decreased on binding the substrate analogue estriol. (20) In addition, the P450 species can sometimes be regenerated from the P420 species by binding substrate, as shown for the P450epoK enzyme from *Sorangium cellulosum*, or by manipulating the pH of the buffer, as demonstrated for Mtb CYP121A1. (26) It may also be the case for CYP143A1 that substrate binding could stabilize the cysteine thiolate-coordinated heme state, though this experiment awaits the identification of its substrate. In its resting state, EPR and resonance Raman spectroscopy demonstrate that the CYP143A1 ferric heme iron is in the low-spin, thiolate-coordinated form, similar to spectra previously reported for the heme domain of the well-studied flavocytochrome P450 BM3 enzyme from *Bacillus megaterium.* (27,28)

Spectral changes (type II Soret shifts to 420-422 nm) observed upon binding of azole drugs to CYP143A1 were consistent with the coordination of nitrogen on the aromatic heterocycles of the azole compounds to the CYP143A1 ferric heme iron. The binding preference of CYP143A1 was higher towards imidazoles including bifonazole, clotrimazole, econazole and miconazole, but with poorer affinities to those observed for the Mtb P450s CYP151B1 and CYP121A1. (21,29) However, Soret shifts were not detected in titrations of CYP143A1 with fluconazole, itraconazole, ketoconazole or voriconazole, possibly due to inability of these compounds to access the CYP143A1 active site, in turn perhaps due to the nature of the active site or to constriction of the heme distal pocket (discussed in detail later). In addition, apparent cooperativity in the binding of azoles to CYP143A1 was observed for econazole and miconazole ligands (Figure 5.5). Similar phenomena were previously reported for Mtb CYP130A1 (22) and CYP144A1 (23), where the authors suggest that this could be due to the effect of enzyme dimerization induced on binding of azoles. However, no further data are available for CYP143A1 to suggest a similar effect. Another potential explanation for apparent cooperativity of CYP143A1 ligand binding is multiple site cooperativity in which two or more ligands bind simultaneously to the same protein molecule, as demonstrated by the *Saccharopolyspora erythraea* (EryF, CYP107A1) P450 enzyme. (30,31)

EPR spectroscopy can provide information relating to the interaction of ligands with the low-spin P450 heme iron, and may also be useful in identifying high spin heme iron. When bound to clotrimazole and miconazole, the CYP143A1 *g*-values were not significantly altered from those of the ligand-free CYP143A1. However, it may be that these azoles do not bind efficiently to CYP143A1 at the high concentrations required. Another possibility is that they may coordinate indirectly to the heme iron via the axial water ligand that remains as the distal ligand. Structural data for the Mtb CYP121A1 enzyme provided proof of indirect ligation of fluconazole to the CYP121A1 ferric heme iron via the water sixth ligand, as reported by Seward et al (32). In addition, they also reported that a direct binding conformation of fluconazole to the CYP121A1 heme iron was also feasible. EPR analysis of CYP143A1 bound to bifonazole and 1-phenylimidazole identified some heterogeneity in their ligation modes. In the complexes of these inhibitors, CYP143A1 exhibits low spin EPR

spectral features with *g*-values of 2.52/2.25/1.87 (bifonazole) and 2.47/2.25/1.89 (1-phenylimidazole) that are consistent with the direct coordination of the heme iron in a proportion of the molecules. EPR studies also showed that there is negligible high spin heme iron content in any of the CYP143A1 samples analysed, although this is also commonly observed in P450 studies conducted at 10 K.

The binding of selected azoles to CYP143A1 may identify this P450 as a potential therapeutic target in *M. tuberculosis*, as azole drugs have been proven to be effective inhibitors of the growth of *M. tuberculosis*, *M. smegmatis* and *Streptomyces* strains. (29,33) Consequently, antimycobacterial potency of the azole drugs has motivated research into the development of novel inhibitors that have the potential to coordinate to the Mtb P450 heme iron. We have adopted a fragment-based approach to provide a platform for rational design and synthesis of more potent and selective *M. tuberculosis* P450 inhibitors that should avoid undesirable interactions with human P450s and target Mtb P450 specifically.

The fragment screening hit rates for CYP143A1 are low (1-11%) compared to previous studies with other Mtb P450s such as CYP142A1 (64%) and CYP125A1 (50%). (9) Most of the CYP143A1 fragment hits reported here were from a heme-focused library in which each fragment contained a known heme binding functional group, such as heteroaromatic nitrogen, aniline, aliphatic amine, nitrile or carboxylic acid. Further studies using UV-visible spectroscopy showed that CYP143A1 has a binding preference for fragments with imidazole as the heme binding scaffold, and the highest binding affinity of CYP143A1 is for fragment **2**. In addition, EPR spectroscopy showed that CYP143A1 in complex with fragment **2** generated a new low spin species with *g*-values of 2.45/2.25/1.90 that are consistent with the direct ligation of the fragment's amine or pyridine nitrogen to the CYP143A1 heme iron. This ligation mode will be studied in more detail using an X-ray crystallography approach.

Although the sequence similarity is usually less than 35% within Mtb P450s, their secondary and tertiary structures are relatively well conserved. The three-dimensional structures of CYP143A1 and CYP142A1 show highly similar folds, but there are differences in the positioning of their various secondary structural elements. These structural features

are commonly involved in heme binding, substrate recognition and the formation of the active site cavity, and are readily identified and comparable with other Mtb P450s for which structures are known. (19,22,24,34–37) In general, the structures are more conserved and rigid the closer they are to the heme, especially for the I and L helices, which directly contact the heme. Key residues are conserved between P450s, including the cysteine residue that provides the heme proximal axial ligand (Cys366 for CYP143A1), and a conserved threonine residue (Thr260 for CYP143A1) that is involved in delivery of protons to iron-bound oxygen species in the catalytic cycle, which are required for the conversion of the ferric-peroxo intermediate to the ferric-hydroperoxo species, and for its subsequent protonation that results in the cleavage of the O-O bond and which leads to the formation of the highly reactive compound 1 intermediate. (38,39) It was suggested that ordered solvent molecules in the P450 active site also participate in the proton delivery process. (40) On the other hand, the structures that differ the most are those flexible regions that control substrate specificity, such as the B', F and G helices. Conformational mobility of the B'C and FG loops was previously observed in other structurally defined P450s, and these regions serve as the 'gates' for substrate entry into and product release from the active site. The structure of the palmitoleic acid-bound P450 BM3 heme domain demonstrated that conformational changes occur upon substrate binding by the movement of the F and G helices "sliding" over the surface of the I helix to close off the entry channel, indicating that substrates enter near the FG loop region. (41) The conformational changes were also found to be regulated by the repositioning of the F, G and B' helices and the FG loop that covers the ligand binding site and reshapes the protein surface. (42,43) In the case of Mtb CYP130A1 and a 6-coordinated cytochrome P450 from *Thermus thermophilus* HB8 (PDB ID: 1WIY), the BC and FG loop regions were also involved in the dimerization of the enzymes. (22) The active site cavity of CYP143A1 is defined by the F and G helices and the long I helix at its N-terminal region, with the majority of residues lining the active site being hydrophobic. The hydrophobicity of the substrate binding site was also observed in other Mtb P450s, although their substrate preferences differ (e.g. the cyclodipeptide cYY for CYP121A1, and cholesterol for CYP125A1); but there are no data to suggest CYP143A1 binds to these substrates, despite cholesterol also being a substrate for the structurally similar CYP142A1 enzyme. The sizes

and shapes of the various substrates of the Mtb P450s are diverse and the structural basis for determining substrate specificity remains unclear for a number of the Mtb (and other) P450s.

Fragment-based screening using X-ray crystallography to identify fragment binding modes was employed to further explore the structural properties of the CYP143A1 active site. However, efforts to extend the characterization of the CYP143A1 active site were hampered by the propensity of PEG (from the crystallization condition) to occupy the active site. Nonetheless, PEG binding suggests that the candidates for CYP143A1 substrate might be unusual fatty acids, long chain lipids or other hydrophobic molecules that mimic aspects of the PEG structure and can bind close to the heme iron. As seen in many P450 substrate-bound structures, the substrates are located such that the atom to be hydroxylated is within ~4-5 Å of the heme iron in the catalytically competent pose. In the case of the CYP143A1 fragment **2**-bound crystal structure, PEG was not observed in the active site, which could be due to tighter affinity of CYP143A1 for fragment **2** (K_d = 14.4 ± 0.5 µM) compared to PEG. Binding titration experiments using available fatty acids, lipids and PEG did not cause significant CYP143A1 spectral changes. However, as seen in the crystal structure of PEG-bound CYP143A1 (Figure 5.13), the PEG is bound indirectly to the heme iron via the sixth water ligand. Therefore, the lack of an apparent spectral change may be due to this binding mode producing a spectrum similar to that of the low-spin, water-ligated form.

5.6 Summary

This manuscript presents the first report of the expression, purification, spectroscopic analysis and structural characterization of the *M. tuberculosis* P450 enzyme CYP143A1. This P450 shows structural resemblance to the Mtb CYP142A1 enzyme, though does not bind cholesterol/cholestenone – which are known substrates for CYP142A1. Ongoing studies are aimed at identifying the substrate(s) for this novel P450s, and at the progression of a fragment screening and elaboration process to enable the production of specific inhibitors of CYP143A1.

5.7 References

1. World Health Organization. Global tuberculosis report 2016. Geneva: WHO Press. www.who.int/tb/publications/global_report/en/ Date last accessed: July 7, 2017. Date last updated: 2016.

2. Mahajan R. Bedaquiline: First FDA-approved tuberculosis drug in 40 years. Int J Appl basic Med Res. 2013 Jan;3(1):1–2.

3. Zumla A, Schito M, Chakaya J, Marais B, Mwaba P, Migliori GB, Hoelscher M, Maeurer M, Wallis RS. World TB Day 2016: reflections on the global TB emergency. Lancet Respir Med. 2016;4(4):249–51.

4. Wong EB, Cohen KA, Bishai WR. Rising to the challenge: new therapies for tuberculosis. Trends Microbiol. 2013;21(9):493–501.

5. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov. 2007 Jan 8;6(1):29–40.

6. Koul A, Arnoult E, Lounis N, Guillemont J, Andries K. The challenge of new drug discovery for tuberculosis. Nature. 2011;469(7331):483–90.

7. Marchetti C, Chan DSH, Coyne AG, Abell C. Fragment-based approaches to TB drugs. Parasitology. 2016;1–12.

8. Mendes V, Blundell TL. Targeting tuberculosis using structure-guided fragment-based drug design. Drug Discov Today. 2017;22(3):546–54.

9. Kavanagh ME, Chenge J, Zoufir A, McLean KJ, Coyne AG, Bender A, Munro AW, Abell C. A fragment profiling approach to inhibitors of the orphan *M. tuberculosis* P450 CYP144A1. Biochemistry. 2017;56(11):1559–72.

10. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr. 2010 Feb;66(Pt 2):133–44.

11. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, IUCr, C. AS, A. P, M. N, J. RR, M. RL, T. S, L. WG. *Phaser* crystallographic software. J Appl Crystallogr. 2007 Aug 1;40(4):658–74.

12. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010 Feb 1;66(2):213–21.

13. Emsley P, Cowtan K. *Coot* : model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr. 2004 Dec 1;60(12):2126–32.

14. Afonine P V., Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallogr Sect D Biol Crystallogr. 2012 Apr 1;68(4):352–67.

15. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. *MolProbity* : all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr. 2010 Jan

1;66(1):12–21.

16. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.8. 2015.

17. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: The CCP4mg molecular-graphics software. Acta Crystallogr Sect D Biol Crystallogr. 2011 Apr 1;67(4):386–94.

18. Omura T, Sato R. Liver microsomes spermidine in the extraction of the deoxyribosyl-synthesizing system from *Escherichia*. J Biol Chem. 1962;237(April):1375–6.

19. Driscoll MD, McLean KJ, Levy C, Mast N, Pikuleva IA, Lafite P, Rigby SEJ, Leys D, Munro AW. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: Evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. J Biol Chem. 2010;285(49):38270–82.

20. McLean KJ, Warman AJ, Seward HE, Marshall KR, Girvan HM, Cheesman MR, Waterman MR, Munro AW. Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. Biochemistry. 2006;45(27):8427–43.

21. McLean KJ, Cheesman MR, Rivers SL, Richmond A, Leys D, Chapman SK, Reid GA, Price NC, Kelly SM, Clarkson J, Smith WE, Munro AW. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. J Inorg Biochem. 2002;91:527–41.

22. Ouellet H, Podust LM, Ortiz De Montellano PR. *Mycobacterium tuberculosis* CYP130: Crystal structure, biophysical characterization, and interactions with antifungal azole drugs. J Biol Chem. 2008;283(8):5069–80.

23. Driscoll MD, McLean KJ, Cheesman MR, Jowitt TA, Howard M, Carroll P, Parish T, Munro AW. Expression and characterization of *Mycobacterium tuberculosis* CYP144: Common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. Biochim Biophys Acta - Proteins Proteomics. 2011;1814(1):76–87.

24. Chenge JT, Le D Van, Swami S, McLean KJ, Kavanagh ME, Coyne AG, Rigby SEJ, Cheesman MR, Girvan HM, Levy CW, Rupp B, von Kries JP, Abell C, Leys D, Munro AW. Structural characterization and ligand/inhibitor identification provide functional insights into the *Mycobacterium tuberculosis* cytochrome P450 CYP126A1. J Biol Chem. 2016;292(1):1310–29.

25. Holsclaw CM, Sogi KM, Gilmore SA, Schelle MW, Leavell MD, Bertozzi CR, Leary JA. Structural characterization of a novel sulfated menaquinone produced by stf3 from *Mycobacterium tuberculosis*. ACS Chem Biol. 2008;3(10):619–24.

26. Dunford AJ, McLean KJ, Sabri M, Seward HE, Heyes DJ, Scrutton NS, Munro AW. Rapid P450 heme iron reduction by laser photoexcitation of *Mycobacterium tuberculosis* CYP121 and CYP51B1. Analysis of CO complexation reactions and reversibility of the P450/P420 equilibrium. J Biol Chem. 2007 Aug 24;282(34):24816–24.

27. Girvan HM, Seward HE, Toogood HS, Cheesman MR, Leys D, Munro AW. Structural and spectroscopic characterization of P450 BM3 mutants with unprecedented P450 heme iron ligand sets: New heme ligation states influence conformational equilibria in P450 BM3. J Biol Chem. 2007 Jan 5;282(1):564–72.

28. Smith SJ, Munro AW, Smith WE. Resonance Raman scattering of cytochrome P450 BM3 and effect of imidazole inhibitors. Biopolymers. 2003 Dec;70(4):620–7.

29. McLean KJ, Marshall KR, Richmond A, Hunter IS, Fowler K, Kieser T, Gurcha SS,

Besra GS, Munro AW. Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes. Microbiology. 2002;148(10):2937–49.

30. Xiang H, Tschirret-Guth RA, Ortiz De Montellano PR. An A245T mutation conveys on cytochrome P450eryF the ability to oxidize alternative substrates. J Biol Chem. 2000 Nov 17;275(46):35999–6006.

31. Cupp-Vickery J, Anderson R, Hatziris Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. Proc Natl Acad Sci. 2000 Mar 28;97(7):3050–5.

32. Seward HE, Roujeinikova A, McLean KJ, Munro AW, Leys D. Crystal structure of the *Mycobacterium tuberculosis* P450 CYP121-fluconazole complex reveals new azole drug-P450 binding mode. J Biol Chem. 2006;281(51):39437–43.

33. McLean KJ, Carroll P, Lewis DG, Dunford AJ, Seward HE, Neeli R, Cheesman MR, Marsollier L, Douglas P, Smith WE, Rosenkrands I, Cole ST, Leys D, Parish T, Munro AW. Characterization of active site structure in CYP121: A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem. 2008;283(48):33406–16.

34. Leys D, Mowat CG, McLean KJ, Richmond A, Chapman SK, Walkinshaw MD, Munro AW. Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450. J Biol Chem. 2003;278(7):5141–7.

35. Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR, Montellano PRO De. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(49):20687–92.

36. McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D, Munro AW. The structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection. J Biol Chem. 2009;284(51):35524–33.

37. Chenge J, Kavanagh ME, Driscoll MD, McLean KJ, Young DB, Cortes T, Levy CW, Rigby SEJ, Leys D, Abell C, Munro AW. Structural characterization of CYP144A1 – a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*. Sci Rep. 2016;6(26628):1–12.

38. Yeom H, Sligar SG, Li H, Poulos TL, Fulco AJ. The role of Thr268 in oxygen activation of cytochrome P450(BM-3). Biochemistry. 1995;34(45):14733–40.

39. Clark JP, Miles CS, Mowat CG, Walkinshaw MD, Reid GA, Daff SN, Chapman SK. The role of Thr268 and Phe393 in cytochrome P450 BM3. J Inorg Biochem. 2006;100(5):1075–90.

40. Denisov IG, Makris TM, Sligar SG, Schlichting I. Structure and chemistry of cytochrome P450. Chem Rev. 2005;105(6):2253–77.

41. Li H, Poulos TL. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat Struct Biol. 1997 Feb;4(2):140–6.

42. Lee YT, Wilson RF, Rupniewski I, Goodin DB. P450cam visits an open conformation in the absence of substrate. Biochemistry. 2010 Apr 27;49(16):3412–9.

43. Nagano S, Li H, Shimizu H, Nishida C, Ogura H. Crystal structures of Epothilone D-bound, Epothilone B-bound, and substrate-free forms of cytochrome P450epoK. J

Biol Chem. 2003;278(45):44886–93.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Conclusions

Tuberculosis (TB) is an airborne disease caused mainly by the aerobic bacterium *Mycobacterium tuberculosis.* (1) The infection is spread person-to-person via the inhalation of infectious particles. Most people can overcome the initial *M. tuberculosis* infection and develop asymptomatic latent TB. However, approximately 10% of the world's population, particularly individuals with immunocompromised conditions such as HIV infection, develop the active TB disease. In this case, the *M. tuberculosis* bacilli multiply rapidly beyond the control of the host immune system, which can lead to death if not treated effectively. Current treatment of active TB relies on a combination of anti-bacterial drugs which are isoniazid, rifampicin, pyrazinamide and ethambutol that have been in use for more than 40 years. (2) This front-line treatment regimen is 95% effective in curing active TB patients, but these drugs must be taken for a period of 6 months. The emergence of multidrug-resistant (MDR-TB) and extensively drug-resistant (XDR-TB) strains of *M. tuberculosis* requires replacing and adding to the front-line regimen with second- and third-line bactericides for successful treatment. (3) These MDR-TB and XDR-TB treatment drugs are more expensive, have greater side effects and can result in poor patient compliance. (4) Moreover, the lengths of a successful treatment regimen and patient death rate have both increased significantly with the spread of MDR-TB and XDR-TB strains. Therefore, there is a critical need for the development of new anti-TB drugs with a novel mechanism of action that can speed up treatment duration and help avoid resistance.

A major breakthrough for TB drug discovery was the complete genome sequencing of the pathogenic Mtb H37Rv strain that revealed several unusual or unique features of the pathogen, and highlighted several new drug target candidates. (5) A particularly interesting finding for our research group was the presence of 20 genes encoding cytochrome P450 enzymes in the 4.4 Mb sized genome. P450s are a large family of monooxygenases that

typically catalyse the activation of molecular oxygen and the insertion of an oxygen atom into a wide variety of substrates. The relatively large number of P450 enzymes suggests that they have important roles within Mtb. Over the past two decades, the characterisation of selected Mtb P450s has highlighted their diverse and unexpected roles in the metabolism of cholesterol and lipids, the production of secondary metabolites (specifically the oxidative crosslinking of the aryl groups of the cyclic dipeptide cyclo-*L*-Tyr-*L*-Tyr by CYP121A1 to form the metabolite mycocyclosin) and the oxidative modification of respiratory menaquinone. (6) The analysis of the structural and functional properties of Mtb P450 enzymes remains highly significant in understanding the biology of Mtb, and knowledge of the properties of these cytochrome P450 enzymes may be exploited for drug discovery and could offer new compounds that are active against Mtb.

With respect to the further understanding of the structure and function of these Mtb P450s, high priority was given to the CYP141A1 enzyme. The CYP141A1-encoding gene (*Rv3121*) is located in one of the *Mycobacterium bovis* BCG RD regions and is thus absent from the genome of the vaccine strain. Consequently, CYP141A1 is potentially a determinant of Mtb virulence or otherwise important for host pathogenicity. In view of CYP141A1's potential importance in Mtb physiology/pathogenicity and as a potential anti-Mtb drug target, characterisation of this enzyme gained our interest. Therefore, the main objectives of this study were to determine the CYP141A1 crystal structure and to establish CYP141A1's spectroscopic and structural characteristics in order to understand its active site properties and its mode of binding to substrate-like and inhibitor-like ligands. By having this information, the study of CYP141A1 can be moved forward to structure-based drug discovery, particularly through the use of a fragment-based screening approach.

In this thesis, data presented in Chapter 2 and Chapter 3 addressed the primary goal of obtaining a CYP141A1 crystal structure. The first strategy employed was N-terminal truncation, which is described in Chapter 2. This strategy was mainly chosen due to the successful characterisation of a truncated version of CYP144A1, the design of which was based on the outcome of a transcriptomic study. (7,8) Although the outcome of N-terminal truncation trials on CYP141A1 did not result in the production of good quality protein

crystals, this strategy revealed some surprising findings. Failure in overexpression of the Del30CYP141A1 construct, but successful overexpression of the Del54CYP141A1 construct in *E. coli* did not meet the expectation that removal of a shorter amino acid sequence would cause fewer complications during expression, purification and enzyme characterisation. Moreover, the ability of Del54CYP141A1 to partially incorporate heme in its active site and still retain the full-length CYP141A1's unique spectroscopic properties was also unexpected. The possibility that the Del54CYP141A1 species is a physiologically relevant form of the Mtb CYP141A1 protein cannot be confirmed as yet, but remains a possibility. In addition, the failure to produce Del54CYP141A1 crystals was proposed to be due to internal heterogeneity of the Del54CYP141A1 protein, and provides some new insights into cytochrome P450 crystallization. Even though this CYP141A1 N-terminal truncation strategy can be optimised, another strategy was considered in order to obtain CYP141A1 crystals.

As modification (truncation) of the CYP141A1 amino acid sequence appeared unfavourable, the second strategy focused on improving the quality of the CYP141A1 protein sample prior to crystallization which is described in Chapter 3. Initially, it was assumed that the process of screening for optimum CYP141A1 buffer conditions could be performed in a high throughput manner. A quick and easy method to identify a buffer composition which increases the stability and possibly the conformational homogeneity of a protein sample is the fluorescence-based thermal shift assay known as Thermofluor. (9) However, attempts to use this method on CYP141A1 to identify improved conditions were not successful (data not shown). Therefore, a lower throughput method was employed for buffer screening. Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) was the method of choice, and is a particularly sensitive method for the detection of protein oligomerization and polydispersity. In one study, it was estimated that 70% of proteins that show a monodisperse size distribution in a light-scattering experiment should give crystals in crystallization trials. (10) However, this method required larger sample volumes, which in turn affected the extent of buffer screening that could be done. Instead of screening a wide range of different buffer composition, the buffer screening process was performed in a more focused but rational way. The screening began by selecting the best type of buffer such that the desired protein pH falls within the effective

buffering capacity of that buffer. In the case of CYP141A1, the desired protein pH range is between 6.5 and 9.0, and hence phosphate, HEPES and Tris buffers are reasonable buffers to consider. Following that, salts such as sodium chloride and reducing agents such as DTT and TCEP were added and screened for the optimum concentration which will keep the protein in solution and avoid protein aggregation, respectively. Finally, the ionic strength and pH of the buffer were optimized for protein solubility, stability and homogeneity. As expected, the optimization of CYP141A1 buffer conditions resulted in the formation of reproducible CYP141A1 crystals and is considered to be a crucial outcome for CYP141A1 X-ray crystallography studies. Improvement of crystallization condition and modification of data collection strategy were required in order to obtain a good resolution CYP141A1 crystal structure. At the time of preparation of this thesis, the highest resolution crystal structure of CYP141A1 was solved at 2.7 Å, by experimental phasing method.

The second objective of this study was addressed in Chapter 4 of this thesis. Various biophysical techniques were applied to establish CYP141A1's spectroscopic and structural characteristics. The active site of CYP141A1 consists of a mixed spin-state ferric heme iron coordinated by a cysteinate residue in the proximal position and a water molecule in the sixth (distal) position, which is typical for a functional cytochrome P450 enzyme. However, CYP141A1 does not display the typical Soret peak at ~450 nm when reduced and in complex with carbon monoxide. Ferrous CO-bound CYP141A1 displayed an unusual Soret peak feature at 440 nm, indicating important structural/electronic differences in the CYP141A1 heme compared to most other P450 enzymes. In other spectroscopic studies, spectral changes were observed (mainly by UV-visible and EPR spectroscopy) upon binding of ligands such as azole compounds, and these provided important information on the mode of binding of such inhibitors. Preliminary fragment-based screening against CYP141A1 also identified some fragment hits that can be the starting point for the development of more potent, isoform-specific inhibitors of this P450. Fragment-based drug discovery projects relies heavily on X-ray crystallography as the P450 structure can provide clear evidence for where and how a fragment molecule binds to the target enzyme. Prior to CYP141A1 crystallization trials, circular dichroism, differential scanning calorimetry and light scattering analysis of the purified CYP141A1 was done to provide insights on CYP141A1 secondary

structure composition, its protein fold and aggregation state. These biophysical techniques also indicated that CYP141A1 has a propensity to crystallize, although the crystallization and structure determination process for CYP141A1 was not as straightforward as expected. Nonetheless, initial structural determination of the ligand-free CYP141A1 by X-ray crystallography revealed key features of its active site cavity and identified critical residues that may be important for ligand interactions and for the catalytic function of the P450. Structural information on CYP141A1 can be useful for fragment-based drug discovery (FBDD) approaches in the future since most FBDD projects reported to date relied on using targets that have known structures. (11)

Since fragment hit characterisation was not performed on CYP141A1 due to slow progress in the CYP141A1 crystallization and structure determination processes, another Mtb P450 enzyme was adopted for this purpose. The selected Mtb P450 enzyme was CYP143A1, an orphan cytochrome P450 encoded by the *M. tuberculosis* H37Rv *Rv1785c* gene. This P450 gene is located alongside a ferredoxin encoding gene (*Rv1786*) in the Mtb genome which is likely required for this P450's catalytic reaction. Therefore, by having a preferred redox partner, CYP143A1 may be inferred to have an important function in Mtb. In terms of the application of fragment-based drug discovery, the ease of obtaining a high-resolution crystal structure of CYP143A1 was an advantage and it was considered that this P450 should be amenable to fragment binding and provide a route to producing tight-binding inhibitors for CYP143A1 through a fragment screening and elaboration process. Similar to the data described in Chapter 4, spectroscopic and structural characterisation was performed on the ligand-free CYP143A1, as well as on CYP143A1 in complex with azole drugs and fragment hits, as detailed in Chapter 5. As is typical of other Mtb P450 enzymes, the ligand-free CYP143A1 heme was predominantly in a low-spin ferric state, and cysteinate-coordinated in the proximal position with a distal water molecule as the sixth (distal) ligand to the heme iron. Spectroscopic changes were observed (by UV-visible and EPR spectroscopy) upon binding of azole compounds and fragments, providing insights into the mode of binding of these inhibitors. EPR spectroscopy analyses of CYP143A1 in complex with nitrogen-containing ligands were consistent with the coordination of the P450 heme iron by nitrogen atoms originating either from an aromatic heterocyclic ring of an azole

drug, or from an amine group of a fragment ligand. Unfortunately, the unexpected presence of a polyethylene glycol molecule in the active site of CYP143A1 hindered the validation of this binding mode by X-ray crystallography. Nevertheless, this study provides insights into CYP143A1 ligand preferences and binding interactions made within its active site cavity, which will also be beneficial for future structure-based drug design studies.

Overall, the data presented in this thesis has addressed many of the objectives set out at the beginning of the research programme. Although there were some unexpected complications, mainly with regards to X-ray crystallography, this research remains significant in providing new knowledge on the structural and spectroscopic properties of the *M. tuberculosis* P450s CYP141A1 and CYP143A1.

6.2 Future recommendations

The major issue with the present CYP141A1 crystal structure is that the resolution at 2.7 Å is not sufficient to obtain detailed information on its active site architecture, thus making the current model structure inappropriate for fragment-based drug discovery. A resolution of around 2.5 Å or better is desired to confidently build ligands into the density observed in the active site and to describe accurately their binding interactions. To date, Mtb P450 enzymes characterised structurally, either in the absence or presence of ligands, were solved at resolutions ranging from 1.06 – 2.2 Å. (7,12–18) Moreover, the first successful application of a fragment-based approach to an Mtb P450 reported structures of CYP121A1 bound to fragments and elaborated fragments at resolutions of between 1.2 – 2.5 Å. (19,20)

In order to obtain a higher resolution crystal structure of CYP141A1, protein engineering should be performed. This is based on the hypothesis that, rather than the external factors such as protein buffer and crystallization conditions, it is the intrinsic characteristics of the enzyme that are causing a problem with high resolution structural determination in this P450. In contrast to the strategy described in Chapter 2, the recommended protein engineering approach for CYP141A1 could be through rational

surface point mutations which are guided by the currently available crystal structure. It can be argued that the CYP141A1 enzyme has surface properties that do not favour good crystal contacts and well-ordered crystal packing. This could be approached by modelling and the recommended mutations could be made by mutating surface exposed lysine and glutamate residues (the least favoured residues in crystal contacts) to arginine and glutamine residues (the most favoured residues in crystal contacts), respectively, and by mutating any surface exposed cysteine residue to serine residues. The latter should be considered first since the addition of reducing agents such as DTT and TCEP did affect the oligomerisation of CYP141A1. There are a large number of examples in the literature in which rational protein surface engineering was shown to be an effective strategy to improve the quality of existing protein crystals. In many of these examples, only a limited number of mutations were required to achieve an improvement in the quality of the crystals. (21–23) As a consequence of protein engineering, the physical appearance of CYP141A1 crystals might differ and belong to a different space group. Proteins crystallizing in lower symmetry systems have lower solvent content to those crystallizing in higher symmetry (24) and protein crystals with less solvent tend to diffract better. (25)

In the case of CYP143A1, the major issue lies in the unsuccessful determination of ligand-bound CYP143A1 crystal structures. In order to obtain crystals of CYP143A1 in complex with azoles and fragments, both soaking and co-crystallization methods were employed. However, co-crystallization seems to be the better method for this P450. As co-crystallization can be affected by temperature, protein concentration, ligand concentration, the use of additives to improve ligand binding, ligand exchange prior to co-crystallization and cross-seeding, (26) these variables should be tested in the future in efforts to optimise efficient ligand binding. In addition, the issue with the presence of PEG derived from the crystallization conditions used could be overcome by screening for new PEG-free crystallization condition using seeding techniques and by the displacement of PEG using tighter binding ligands.

In order to move forward with fragment-based approaches on CYP141A1 and CYP143A1, the binding modes of fragments hits may need to be determined by methods

other than X-ray crystallography. Computational studies and biophysical techniques such as NMR spectroscopy could be considered. *In silico* methods are widely available and are frequently used in drug discovery and ligand docking experiments, and these approaches could provide structural data on fragment hit binding modes. Furthermore, molecular modelling could be used to develop fragments even in the absence of experimental structures. (26) In addition, NMR spectroscopy could also be used to obtain structural information on protein-fragment complexes for the purpose of lead optimisation. Various methods such as Nuclear Overhauser Effects (NOEs) have been used previously and were shown to be suitable for drug design. (27) Once the binding modes of the fragment hits are confirmed, fragment optimisation could begin by using fragment growing, merging or linking strategies. Optimised compounds typically have much tighter binding affinities (i.e. lower K_d values) than the fragments from which they are derived, and this should be advantageous for X-ray crystallography studies and for detailed structure-activity relationship studies prior to compound testing for anti-Mtb activity.

6.3 References

1. Koch R. Die Aetiologie der Tuberculose (Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März gehaltenem Vortrage). Berliner klin Wochenschr. 1882;19(15):221–30.

2. Horsburgh CR, Barry CE, Lange C. Treatment of tuberculosis. N Engl J Med. 2015;373(22):2149–60.

3. Parsons LM, Driscoll JR, Taber HW, Salfinger M. Drug resistance in tuberculosis. Infect Dis Clin North Am. 1997 Dec;11(4):905–28.

4. World Health Organization (WHO). Global tuberculosis report 2016. Geneva: WHO Press. www.who.int/tb/publications/global_report/en/ Date last accessed: July 7, 2017. Date last updated: 2016..

5. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537–44.

6. McLean KJ, Belcher J, Driscoll MD, Fernandez CC, Le Van D, Bui S, Golovanova M, Munro AW. The *Mycobacterium tuberculosis* cytochromes P450: physiology, biochemistry & molecular intervention. Future Med Chem. 2010;2(8):1339–53.

7. Chenge J, Kavanagh ME, Driscoll MD, McLean KJ, Young DB, Cortes T, Levy CW, Rigby SEJ, Leys D, Abell C, Munro AW. Structural characterization of CYP144A1 – a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*. Sci Rep. 2016;6(26628):1–12.

8. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. Cell Rep. 2013;5(4):1121–31.

9. Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS. Optimization of protein buffer cocktails using Thermofluor. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2013;69(2):209–14.

10. Zulauf M, D'Arcy A. Light scattering of proteins as a criterion for crystallization. J Cryst Growth. 1992;122(1–4):102–6.

11. Erlanson DA, McDowell RS, O'Brien T. Fragment-based drug discovery. J Med Chem. 2004;47(14):3463–82.

12. Leys D, Mowat CG, McLean KJ, Richmond A, Chapman SK, Walkinshaw MD, Munro AW. Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450. J Biol Chem. 2003;278(7):5141–7.

13. Ouellet H, Podust LM, Ortiz De Montellano PR. *Mycobacterium tuberculosis* CYP130: Crystal structure, biophysical characterization, and interactions with antifungal azole drugs. J Biol Chem. 2008;283(8):5069–80.

14. McLean KJ, Warman AJ, Seward HE, Marshall KR, Girvan HM, Cheesman MR, Waterman MR, Munro AW. Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. Biochemistry. 2006;45(27):8427–43.

15. McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D, Munro AW. The structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection. J Biol Chem. 2009;284(51):35524–33.

16. Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR, Montellano PRO De. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2009;106(49):20687–92.

17. Driscoll MD, McLean KJ, Levy C, Mast N, Pikuleva IA, Lafite P, Rigby SEJ, Leys D, Munro AW. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: Evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. J Biol Chem. 2010;285(49):38270–82.

18. Chenge JT, Le D Van, Swami S, McLean KJ, Kavanagh ME, Coyne AG, Rigby SEJ, Cheesman MR, Girvan HM, Levy CW, Rupp B, von Kries JP, Abell C, Leys D, Munro AW. Structural characterization and ligand/inhibitor identification provide functional insights into the *Mycobacterium tuberculosis* cytochrome P450 CYP126A1. J Biol Chem. 2016;292(1):1310–29.

19. Hudson SA, McLean KJ, Surade S, Yang YQ, Leys D, Ciulli A, Munro AW, Abell C. Application of fragment screening and merging to the discovery of inhibitors of the *Mycobacterium tuberculosis* cytochrome P450 CYP121. Angew Chemie - Int Ed. 2012;51(37):9311–6.

20. Kavanagh ME, Coyne AG, McLean KJ, James GG, Levy CW, Marino LB, De Carvalho LPS, Chan DSH, Hudson SA, Surade S, Leys D, Munro AW, Abell C. Fragment-Based Approaches to the Development of *Mycobacterium tuberculosis* CYP121 Inhibitors. J Med Chem. 2016;59(7):3272–302.

21. Dale GE, Oefner C, D'Arcy A. The protein as a variable in protein crystallization. J Struct Biol. 2003 Apr;142(1):88–97.

22. Derewenda ZS. Rational protein crystallization by mutational surface engineering. structure. 2004 Apr;12(4):529–35.

23. Derewenda ZS, Vekilov PG. Entropy and surface engineering in protein crystallization. Acta Cryst. 2006;62:116–24.

24. Chruszcz M, Potrzebowski W, Zimmerman MD, Grabowski M, Zheng H, Lasota P, Minor W. Analysis of solvent content and oligomeric states in protein crystals--does symmetry matter? Protein Sci. 2008 Apr;17(4):623–32.

25. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 2003 Sep;12(9):1865–71.

26. Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng SJJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT, Rocque WJ, Wang L, Weaver KL, Williams SP, Wisely GB, Xu R, Shewchuk LM. Crystallization of protein-ligand complexes. Acta Crystallogr Sect D Biol Crystallogr. 2006 Jan;63(1):72–9.

27. Harner MJ, Frank AO, Fesik SW. Fragment-based drug discovery using NMR spectroscopy. J Biomol NMR. 2013;56(2):65–75.