Investigating Orphan Cytochrome P450s in *Mycobacterium tuberculosis*: Insights into Enzyme Structure, Function and Inhibitor Design

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Life Sciences

2016

Jude Chenge
Table of Contents

Abbreviations ........................................................................................................................................9
Declaration ............................................................................................................................................11
Copyright Statement ..........................................................................................................................12
Acknowledgements .............................................................................................................................13
Preface to the Alternative Thesis ........................................................................................................14
Abstract ............................................................................................................................................16

1 Chapter One ........................................................................................................................................18

1.1 Introduction .....................................................................................................................................18
1.2 The *Mycobacterium* Genus ........................................................................................................18
1.3 The *Mycobacterium tuberculosis* Complex ................................................................................20
1.4 Tuberculosis (TB) – History and Epidemiology ...........................................................................20
1.5 The pathogen *Mycobacterium tuberculosis* ................................................................................23
1.6 Immune responses to *M. tuberculosis* .......................................................................................24
1.7 Importance of Cholesterol to *Mtb* Pathogenicity and Persistence .............................................27
   1.7.1 *M. tuberculosis* and cholesterol: a do-or-die affair ..........................................................27
   1.7.2 *M. tuberculosis* catabolizes cholesterol ............................................................................28
1.8 Treatment of Tuberculosis – The journey so far ............................................................................29
1.9 The Cytochrome P450s (CYPs) .....................................................................................................33
1.10 The Structure of Cytochrome P450 .............................................................................................39
1.11 Cytochrome P450 redox partners ...............................................................................................41
1.12 The activities of bacterial and eukaryotic cytochrome P450s .....................................................46
1.13 Azoles and *Mtb* CYP inhibition ...............................................................................................46
1.14 CYPs of the *M. tuberculosis* Genome ........................................................................................48
   1.14.1 CYP51B1 ............................................................................................................................50
   1.14.2 CYP121A1 ..........................................................................................................................51
   1.14.3 CYP128A1 ..........................................................................................................................53
   1.14.4 CYP130A1 and CYP141A1 .................................................................................................53
   1.14.5 CYP124A1, 125A1 and 142A1 ............................................................................................54
1.15 Conclusion and study objectives .................................................................................................56
   1.15.1 *M. tuberculosis* CYP144A1 and CYP126A1 ....................................................................56
   1.15.2 Research Objectives ..........................................................................................................58

References ............................................................................................................................................60
2 Chapter Two .......................................................................................... 69
2.1 Abstract .................................................................................................. 70
2.2 Introduction ............................................................................................. 71
2.3 Results and Discussion ........................................................................... 73
  2.3.1 Bioinformatics and evolutionary studies of CYP144A1 ...................... 73
  2.3.2 Identification of a truncated form of the CYP144A1 P450 from transcriptome analysis .......................................................... 76
  2.3.3 Expression and purification of CYP144A1 in its truncated and full-length forms 78
  2.3.4 UV-Visible spectroscopic properties of CYP144A1-FLV and CYP144A1-TRV 80
  2.3.5 EPR spectroscopic analysis of the CYP144A1-FLV and CYP144A1-TRV proteins 81
  2.3.6 Differential Scanning Calorimetry (DSC) studies on the CYP144A1-FLV and CYP144A1-TRV enzyme forms ......................................................... 83
  2.3.7 Crystal structure determination of CYP144A1-TRV .............................. 84
  2.4 Conclusions ........................................................................................... 87
  2.5 Methods .................................................................................................. 90
    2.5.1 Bioinformatics studies ........................................................................ 90
    2.5.2 Identification of CYP144A1 transcriptional start sites ...................... 90
    2.5.3 Cloning of the CYP144A1-FLV and TRV genes ................................. 90
    2.5.4 Expression and purification of the CYP144A1-FLV and CYP144A1-TRV forms of CYP144A1 .............................................................. 91
    2.5.5 Mass spectrometry .............................................................................. 93
    2.5.6 Determination of molecular weight and aggregation state of CYP144A1-FLV and CYP144A1-TRV by multi-angle laser light scattering (MALLS) 94
    2.5.7 EPR spectroscopic analysis of CYP144A1 ........................................... 94
    2.5.8 Determination of CYP144A1 thermal stability by Differential Scanning Calorimetry (DSC) ................................................................. 95
    2.5.9 CYP144A1 crystallization and structure determination ..................... 95
References ..................................................................................................... 97

3 Chapter Three ............................................................................................ 100
3.1 Abstract .................................................................................................. 101
3.2 Introduction ............................................................................................. 102
3.3 Results ...................................................................................................... 106
  3.3.1 Bioinformatic and structural analysis .................................................. 106
3.3.2 Ligand screening using UV-Vis spectrophotometry ........................................ 110
3.3.3 Fragment screening using ligand-observed NMR ........................................ 117
3.3.4 Characterization of fragment binding mode using NMR spectroscopy .......... 118
3.3.5 Characterization of ligand binding mode by competition UV-Vis spectroscopy. 120
3.3.6 Characterization of ligand binding mode using EPR spectroscopy ............... 122
3.3.7 Characterization of ligand binding using nanoESI native mass spectrometry.... 125
3.4 Discussion ........................................................................................................ 128
3.5 Supporting Information ....................................................................................... 139
3.6 Experimental Methods ...................................................................................... 146
3.7 General ............................................................................................................... 146
3.7.1 Bioinformatics, modelling and ligand docking studies .................................. 146
3.7.2 Expression and purification of CYP144A1 .................................................. 148
3.7.3 Optical titrations ............................................................................................ 148
3.7.4 UV-Visible spectroscopy, ligand screening and competition assays ............ 149
3.7.5 Isothermal titration calorimetry ...................................................................... 151
3.7.6 Ligand-observed NMR .................................................................................. 151
3.7.7 EPR spectroscopy .......................................................................................... 153
3.7.8 Differential scanning calorimetry .................................................................. 153
3.7.9 Nano-ESI native mass spectrometry .............................................................. 153
Supplementary Methods S1. Biochemical turnover analysis of compound 1.......... 154
Supplementary Methods S2. Potential substrates screened against CYP144A1 using UV-Vis spectroscopy.................................................................................................................. 155

References ............................................................................................................... 156

4 Chapter Four ....................................................................................................... 161
4.1 Abstract ............................................................................................................. 162
4.2 Introduction ....................................................................................................... 163
4.3 Results ............................................................................................................... 166
4.3.1 Expression and Purification of CYP126A1 .................................................. 166
4.3.2 UV-Visible Spectroscopic Properties of CYP126A1 and Identification of Novel CYP126A1 Ligands and their Binding Properties .................................................. 168
4.3.3 EPR Spectroscopy of CYP126A1 and Ligand-Bound forms ....................... 173
4.3.4 MCD spectroscopy of CYP126A1 ............................................................... 176
4.3.5 Determination of the heme iron reduction potentials of substrate-free and ligand-bound forms of CYP126A1 ................................................................. 178
4.3.6  Analysis of CYP126A1-mediated oxidation of HTS molecules .............................. 181
4.3.7  Crystal structure of CYP126A1 in the imidazole-bound form .............................. 182
4.3.8  Crystal structure of CYP126A1 in complex with inhibitor and substrate-like molecules from compound screens .............................................................. 185

4.4  Discussion and Conclusions .................................................................................. 188

4.5  Experimental Procedures ..................................................................................... 193
4.5.1  Cloning and Production of a CYP126A1 Expression Plasmid Construct .......... 193
4.5.2  Expression and purification of CYP126A1 ......................................................... 194
4.5.3  Native Mass and NanoESI Mass Spectrometry .................................................... 196
4.5.4  UV-Visible Absorbance Titrations .................................................................. 197
4.5.5  EPR and MCD Spectroscopy ............................................................................ 199
4.5.6  Redox Potentiometry ....................................................................................... 199
4.5.7  Analysis of CYP126A1-dependent oxidation of substrate-like compounds from compound screening ................................................................. 201
4.5.8  Crystallography and CYP126A1 Structure Determination ............................... 201
4.5.9  Materials ........................................................................................................ 203

References ..................................................................................................................... 204

5  Chapter Five – Summary, Conclusions & Future Prospects ................................. 208

References ..................................................................................................................... 217

Word Count: 58,669

List of Figures

Figure 1.1. The *Mycobacterium* genus. ................................................................. 19
Figure 1.2. The *Mycobacterium* Cell wall ............................................................. 24
Figure 1.3. Immune response and host-directed therapy against mycobacterial infection. ..................................................................................................................... 26
Figure 1.4. Mechanisms of action of tuberculosis drugs ........................................ 32
Figure 1.5. The structures of the major forms of heme ........................................... 35
Figure 1.6. Absorption spectra of cytochrome P450 .............................................. 35
Figure 1.7. Cytochrome P450 reaction mechanism .................................................. 38
Figure 1.8. A schematic representation of cysteinate protonation ......................... 41
Figure 1.9. Different electron transfer systems in P450 catalyzed reactions .......... 45
Figure 1.10. Chemical structures of selected azole antifungals ............................. 48
Figure 1.12. Phylogenetic analysis of the amino acid sequences of the 20 Mtb P450 enzymes

Figure 1.13. Scheme for various reactions catalysed by Mtb CYPs

Figure 2.1. Phylogenetic tree illustrating the evolutionary relationship of CYP144A1 to other mycobacterial CYP144 family P450s

Figure 2.2. Alternative transcripts of the CYP144A1 gene produced by Mtb H37Rv

Figure 2.3. Expression and purification of CYP144A1-FLV and CYP144A1-TRV

Figure 2.4. Spectroscopic properties of CYP144A1-TRV

Figure 2.5. Differential Scanning Calorimetry (DSC) analysis of CYP144A1-TRV

Figure 2.6. Protein disorder prediction and hydrodynamic features of CYP144A1

Figure 2.7. The crystal structure of CYP144A1-TRV

Figure 3.1. Structural overlay of CYP144A1 with Mtb CYP121A1, CYP125A1 and CYP142A1, and comparative structural data

Figure 3.2. Ligand profiling and characterization strategy for the orphan Mtb enzyme CYP144A1

Figure 3.3. Structures of CYP144A1 ligands and fragments

Figure 3.4. Binding affinity titrations of CYP144A1 with amino-pyrazole ligands and fragments

Figure 3.5. Structural composition and hit rates from fragment screening by ligand-observed NMR spectroscopy

Figure 3.6. CPMG ligand-observed NMR spectra of CYP144A1 with competitive and cooperatively binding fragments

Figure 3.7. CYP144A1 UV-Vis absorption spectra from competition binding assays

Figure 3.8. X-band EPR spectra of CYP144A1 with a selection of heme binding and non-heme binding ligands

Figure 3.9. NanoESI native mass spectra of CYP144A1-ligand binding interactions

Figure 3.10. Rationalization of experimentally characterized CYP144A1-ligand binding interactions using ligand docking

Figure S3.11. Aminopyrazole compounds identified using UV-Vis spectroscopy to bind to CYP144A1
Figure S3.12. ITC binding isotherm and integrated enthalpy plot for the titration of CYP144A1 (50 μM) with compound 3 (1 mM). .................................................................139
Figure S3.13. Ligand observed NMR screening and binding mode characterization. .................................................................................................................................................140
Figure S3.14. Twenty-six fragment hits identified from cocktails screened by ligand-observed NMR CPMG experiment.................................................................140
Figure S3.15. Twenty fragment hits identified by singleton ligand observed NMR screening using STD, CPMG and WaterLOGSY experiments. .........................141
Figure S3.16. Fragments showing competitive or cooperative binding interactions with clotrimazole in CPMG NMR experiments. ........................................141
Figure S3.17. Fragments (0.5-1 mM) which displaced miconazole (12.5 μM) in competition UV-Vis spectroscopy experiments. ..................................................141
Figure S3.18. X-band EPR spectra of CYP144A1 bound to azole antifungal drugs.142
Figure S3.19. Fragments identified to bind CYP144A1 by EPR spectroscopy..............143
Figure S3.20. NanoESI mass spectra of ligand-free CYP144A1 demonstrating the change in charge state with d₆-DMSO.................................................................144
Figure S3.21. DSC traces demonstrating the effect of ligand binding on the thermal stability of CYP144A1......................................................................................144
Figure 4.1. Purification and mass spectrometry of CYP126A1.................................168
Figure 4.2. UV-visible spectral properties of CYP126A1......................................169
Figure 4.3. Binding of the azole inhibitor ketoconazole and HTS compound 1 to CYP126A1.................................................................................................................171
Figure 4.4. Inhibitors and substrate-like ligands for CYP126A1. .........................172
Figure 4.5. Electron Paramagnetic Resonance Spectra of CYP126A1. .............176
Figure 4.6. Magnetic Circular Dichroism analysis of CYP126A1. .....................178
Figure 4.7. Determination of the heme reduction potentials of ligand-free and compound 1-bound CYP126A1...........................................................................180
Figure 4.8. CYP126A1-dependent oxidation of Compound 4. .........................182
Figure 4.9. The crystal structure of CYP126A1.....................................................184
Figure 4.10. Structures of CYP126A1-ligand complexes..................................187
Figure 4.11. Active site structures of CYP126A1-ligand complexes..................187
List of Tables

Table 1.1. Various drugs and progress made towards repurposing them for TB treatment .................................................................31
Table 2.1. Data reduction and final structural refinement statistics for *Mycobacterium tuberculosis* CYP144A1 .................................................................96
Table 3.1. Binding mode and affinity of small molecule ligands and fragments for CYP144A1 and *Mtb* P450s CYP121A1, CYP125A1 and CYP142A1 ...........116
Table S3.2. The g-values from X-band EPR spectra of CYP144A1 with selection of heme binding and non-heme binding ligands ........................................142
Table S3.3. The g-values generated by fragments and azole antifungal drugs binding to CYP144A1 in X-band EPR spectroscopy experiments .........................143
Table S3.4. Melting transitions measured for CYP144A1-ligand complexes by DSC .................................................................145
Table 4.1. Binding constants for azole drugs with CYP126A1 and other *M. tuberculosis* cytochrome P450 enzymes ......................................................173
Table 4.2. Binding constants for compounds identified to bind CYP126A1 through high-throughput screening studies ........................................173
Table 4.3. Data reduction and final structural refinement statistics for the *M. tuberculosis* H37Rv CYP126A1 and drug complexes .........................................203
Abbreviations

°C Degrees celcius
A Absorbance
Å Angstroms (10^-10 m)
AIDS Acquired immunodeficiency syndrome
Amp Ampicillin
BCG Bacillus Calmette-Guérin
bp Basepair
CO Carbon monoxide
CPB Cytoplasmic phospholipid bilayer
CMPG Car-Purcell-Meiboom-Gill
CT Charge transfer
CYP or P450 Cytochrome P450
cYY Cyclo-L-Tyr-L-Tyr
DNA Deoxyribonucleic acid
EMB Ethambutol
EPR Electron paramagnetic resonance
FAD Flavin adenine dinucleotide
FDR Ferrodoxin reductase
FDX Ferrodoxin
FMN Flavin mononucleotide
FLV Full length version of CYP144A1
HIV Human immunodeficiency virus
HS High-spin
HT-SPOTi High Throughput Spot culture growth inhibition
IPTG Isopropyl-beta-D-1-thiogalactopyranoside
Kd Dissociation constant
KO Knockout
KPi Potassium phosphate
L Litre
LB Luria-Bertani (Lysogenic broth)
LC Liquid chromatography
LS Low-spin
m meter
m Milli (10^-3 m)
M Molar
MCD Magnetic Circular Dichroism
mce Mammalian cell entry
MDR-TB Multi-drug resistant tuberculosis
MIC Minimum inhibitory concentration
MOM Mycobacterial outer membrane
Mtbo Mycobacterium tuberculosis
n Nano (10^-9 m)
NanoESI MS NanoElectrospray ionization mass spectrometry
NAD Nicotinamide adenine dinucleotide
NHE Normal hydrogen electrode
NMR Nuclear magnetic resonance
NO Nitric oxide
OD Optical density
P450 BM3  CYP102A1 from Bacillus megaterium
P450cam  CYP101A1 from Pseudomonas putida
PCW     Periplasmic cell wall
PDIM    Phenolphthiocerol-dimycocerate
Pdr     Putidaredoxin reductase
Pdx     Putidaredoxin
PMSF    Phenylmethanesulfonyl fluoride
RIF     Rifampicin
rpm     Revolutions per minute
sp      Spinach
TB      Tuberculosis
TRV     Truncated version of CYP144A1
UV      Ultraviolet
WHO     World Health Organisation
Water LOGSY Water-Ligand Observed via Gradient Spectroscopy
XDR-TB  Extensively or extremely drug-resistant
X-rays  Electromagnetic radiation with a wavelength in the
        range of 0.01 to 10 nm
ε       Extinction coefficient
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 alternative format, which allows the inclusion of work already published.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialization of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectualproperty.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Acknowledgements

I first of all thank my PhD supervisor, Professor Andrew Munro. He has given me the opportunity to study for a Doctorate in his Lab, and taught me a great deal. He showed great patience, faith and trust in me all through the PhD and without such attributes I may not be where I am now. Not only did he mentor me during the PhD, he also initiated my training as a Postdoc in his lab, which further shows his confidence in my professional and scientific methodologies. Another unquantifiable debt of gratitude goes to Dr. Kirsty McLean. She has taught me all I know about scientific research to date and has been a great advisor on scientific, personal and professional matters. Thanks for your help and always having the right answer.

I want to say a huge thank you to my parents. All my life I have had so much love and support.

Gratitude must go to all the members of the Munro, Scrutton and Leys groups. They have been wonderful friends both in and out of the lab.

Finally, the greatest of all thanks to the love of my life Kate Chenge. We got there in the end! I owe you a great deal. You were there for me when my crystals did not diffract well, when things went bad at work, when I needed to talk and you put up with me during this last months. We have lots of catching up to do, and am looking forward to it.
Preface to the Alternative Thesis

This thesis is presented in the style of the University of Manchester’s alternative PhD format. This format is different to a conventional thesis because it allows the author to include data in a format suitable for publication, although the data presented in each must be unique. The structure of each results chapter follows that specified by the journal to which it is associated. Therefore, this thesis does not contain a separate methods chapter, as all the materials and methods have been described in the individual results chapters. Due to this, the main body of the thesis consists of the following chapters; an abstract, introduction, three results chapters and a summary & conclusion chapter. Due to the nature of the self-contained chapters (papers), each chapter has been given a reference list specific to that section. For continuity, the formatting of the text and page layout has been maintained throughout the thesis and numbering of pages, figures and tables is continuous, although still relatable to each respective chapter.

The alternative format thesis has been introduced to be more relevant to the processes involved in scientific research, with particular importance to the reporting of such research and publication of papers. The nature of scientific publications is collaborative; in order to better achieve a scientific goal. For this reason, as part of the alternative format, the contributions of each co-author to the papers must be established.

Papers included as results chapters

Chapter Two


Chapter Three


Chapter Four


Contributions from authors
As a supervisor for the PhD, Andrew W. Munro contributed to both data analysis and manuscript preparation for all papers.

**Chapter Two**

Chenge J expressed, purified and crystalized the protein and performed all spectroscopic, hydrodynamic and DSC experiments.

Kavanagh ME, Abell C and Matak-Vincovic D performed all mass spectrometry analysis of CYP144A1.

McLean KJ and Driscoll MD aided in the cloning and expression of the full length CYP144A1.

Young DB and Cortes T performed transcriptomic analysis of the *CYP144A1* gene.

Leys D and Levy C collected X-ray diffraction data and aided in solving the crystal structure of CYP144A1 by molecular replacement.

Rigby SE performed EPR experiments and data analysis

**Chapter Three**

Chenge J expressed and purified the protein and performed UV-Vis, EPR, and DSC experiments and contributed to manuscript preparation.

Kavanagh ME performed NMR experiments to identify novel CYP144A1 ligands.

McLean KJ and Driscoll MD aided in the cloning of the full length CYP144A1.

Munro AW and Abell C performed data analysis and contributed to manuscript preparation.

**Chapter Four**

Chenge J expressed, purified and crystalized the *apo* protein and performed UV-Visible titrations and contributed to manuscript preparation.

Le Van, D and Swami S aided in cloning and expression and contributed to co-crystallization of CYP126A1

McLean KJ aided in cloning and expression of the *CYP126A1* gene, turnover experiments and data analysis.

Kavanagh ME performed all mass spectrometry analysis of CYP126A1 and assisted in data analysis and manuscript preparation.

Coyne A performed NMR fragment screening experiments to identify novel small molecule scaffolds for CYP126A1.
Cheeseman MR and von Kries JP performed MCD and HTS experiments, respectively.

Rigby SE performed EPR experiments and data analysis.

Levy C and Leys D collected X-ray diffraction data and aided in resolving crystal structures in this paper.

Abell C and Munro analyzed data and contributed to manuscript preparations.
Abstract

The World Health Organization regards tuberculosis as a world pandemic disease. There is increased demand for new drugs to tackle this threat. This threat has been further elevated with the emergence of drug resistant strains of the causative pathogen, *Mycobacterium tuberculosis* (*Mtb*), thereby increasing the urgency for development of novel anti-tubercular drugs. Success in whole genome sequence determination of *Mtb* revealed a large cohort of cytochrome P450 (CYP) enzymes. Research on these *Mtb* P450s has shown that several of them are critical to the survival of the pathogen. CYP121A1 and CYP128A1 have been demonstrated to be essential using knockout experiments. CYP125A1 and CYP142A1 have been shown to play crucial roles in bacterial catabolism of host steroids, with CYP125A1 also shown to be located within a gene cluster highly important for bacterial virulence and infectivity. CYP144A1 was shown to be one of the genes whose expression is elevated when *Mtb* was exposed to macrophage-like conditions, and gene knockout studies using the H37Rv virulent strain of *Mtb* indicated the ΔCYP144A1 mutant to be more sensitive to the clotrimazole antifungal. CYP126A1 was shown to be located within a cluster of genes highly important for the *de novo* synthesis of purines in *Mtb*. These and other data suggested these enzymes to be important to the growth process of *Mtb* and thus potential drug targets for developing novel therapeutics.

Findings in this PhD have revealed that many characteristics of CYP144A1 and CYP126A1 are comparable to previous *Mtb* P450s reported to date. CYP144A1 is highly conserved within the *Mycobacterium* genus and specifically within pathogenic species. Transcriptomic analysis has revealed an alternative truncated transcript leading to the production of two physiologically relevant versions of CYP144A1. Our comparative biophysical characterization of both versions (CYP144A1-FLV and -TRV) show both enzymes to be similar in their binding tightly to azole antifungals. EPR and DSC studies show that the 30 amino acid truncation (to form CYP144A1-TRV) does not affect the heme electronic environment and the overall thermal stability of the enzymes. X-ray crystallography was utilized to determine the first crystal structure of a *Mtb* CYP144 family enzyme. The structure reveals that CYP144A1 possesses a large hydrophobic active site primed for accommodating large hydrophobic substrates. Further chemoproteomic profiling identified novel compounds, which bind in both inhibitor-like and substrate-like modes to CYP144A1, resulting in the development of novel CYP144A1 compounds for use as chemical probes for this P450. Fragment and compound screening identified several ligands with varying binding affinities for CYP126A1, suggesting that this P450 is capable of binding and catalyzing reactions with a wide range of substrates. Turnover experiments proved catalytic activity of CYP126A1 on one of these compounds (Compound 4). Crystallization of CYP126A1 with various compound “hits” (compounds 1 and 7, the azole drug ketoconazole) revealed involvement of several important residues within the active site of CYP126A1 in interactions with these molecules, thus providing important information for designing inhibitors for this enzyme.

Both CYP144A1 and CYP126A1 display important characteristics that contribute to our general understanding of cytochromes P450 as a whole, and of *Mtb* P450s in particular. This PhD project has established the first instance of leaderless transcripts in *Mtb* P450s and has presented the first crystal structures of both CYP144A1 and CYP126A1, as well as identifying novel, useful chemicals that can be used as mechanistic probes for these enzymes as well as providing the basis for *Mtb* P450 isoform-specific inhibitors.
1 Chapter One

1.1 Introduction

1.2 The *Mycobacterium* Genus

The appearance of mycobacteria on earth dates back to about 150 million years ago, but it was around 2400 BC that tuberculosis began presenting itself amongst communities in ancient Egypt\(^1\). Various species of the *Mycobacterium* genus have since continued to thrive within human and animal communities, and in most cases they are linked to, or associated with, a disease\(^2\). The development of sensitive techniques involving sequence-based approaches for identifying bacterial variants has led to studies focused on the non-tuberculosis mycobacteria (NTM) species, which presently has enabled the identification of ~150 species in this genus\(^2,3\). This species number is anticipated to expand due to two reasons; the first is the discovery of organisms in obscure habitats, which were previously not classified as members of the *Mycobacterium* genus. The second reason is cases of redefining organisms that were previously considered to be variants of other species\(^2\). The *Mycobacterium* genus has been classically subdivided into two categories of organisms; those that are rapid-growing (colonies visualized within a week) and those that are slow-growing. Figure 1.1 shows a phylogenetic representation of these species based on the completed genome sequence analysis of these organisms. The rapid-growing category has only one human pathogen (*M. abscessus*), while the slow-growing category is comprised of the *M. tuberculosis* complex, and other species associated with disease in humans, such as *M. avium, M. intracellulare, M. leprae, M. ulcerans* and *M. kansasii*\(^2\).
The evolution of mycobacteria is by both vertical inheritance (gene sharing across species) of core genes, as well as by horizontal gene transfer in the form of chromosomal DNA or plasmids. These two methods of evolution in bacteria have already been reported for *M. marinum*, *M. ulcerans* and the *M. avium* complex. Recently, this method has been applied to describe the genesis of *M. tuberculosis* from a related NTM species, which revealed that, after its common ancestry with a *M. kansasii*-like organism, the *M. tuberculosis* complex acquired about 55 new genes through horizontal gene transfer, further confirming the evidence of horizontal gene transfer in *M. tuberculosis*. Presently, ongoing studies (involving targeted gene disruption, or heterologous putative virulence gene complementation into lesser virulent genus groups, such as *M. kansasii*) are focused on defining the extent to which these genes acquired by the *M. tuberculosis* complex contributed to the unique properties and pathogenicity of *M. tuberculosis*.

**Figure 1.1.** The *Mycobacterium* genus.

The diagram shows the phylogeny of the *Mycobacterium* genus and a schematic representation of the acquisition of foreign genes at different strata of *M. tuberculosis* evolution. The radial tree was generated using MEGA 6. The different mycobacterial lineages, based on slow-growing versus rapid-growing organisms, are also indicated. Figure adapted from Veyrier et al, 2009.
1.3 The Mycobacterium tuberculosis Complex

The mycobacterial organisms grouped within the *Mycobacterium tuberculosis* complex (MTBC) exhibit about 99% identity at the nucleotide level and have an identical 16S rRNA sequence\(^7\). The classical members of the MTBC are comprised of *M. tuberculosis*, *M. africanum*, *M. microti* and *M. bovis*. Other recently categorized organisms include *M. caprea*, *M. pinnipedii* and *M. canettii* \(^7\). *M. tuberculosis*, *M. africanum* and *M. canettii* are the pathogens primarily responsible for tuberculosis disease in humans \(^8\)-\(^{10}\), while *M. microti*, *M. bovis*, *M. caprae* and *M. pinnipedii* are responsible for tuberculosis disease in animals, albeit with a few transmissions also seen in humans \(^11\)-\(^{14}\). Rare potential inclusions are the geographically restricted variants of MTBC that are comprised of *M. orygis* and *M. mungi*, which are responsible for causing tuberculosis (TB) disease in banded mongoose \(^15\)-\(^{17}\). Human and animal infections by MTBC organisms, particularly *M. tuberculosis* and *M. bovis*, have become common and are linked to the increasing global burden of TB\(^7\).

1.4 Tuberculosis (TB) – History and Epidemiology

Tuberculosis, an ancient scourge, has plagued the human race since their prehistory. The disease was well documented in classical Greece, where it was originally called phthisis\(^1\). Hippocrates identified TB and understood its clinical symptoms. He stated in his writings (his so-called “aphorisms” “Phthisis makes its attacks mainly in young adults between the age of eighteen and thirty five”, which means that he could recognize that TB infection is active mainly in young adults \(^1\),\(^{18}\). Clarissimus Galen, a Greek physician, attended to the Roman Emperor Marcus Aurelius in 174 AD. He also documented TB and prescribed milk, fresh air and sea voyages for treatment. In Alexandria, Egypt, however, the disease seemed not to be a problem at the time\(^1\). TB
continued to spread through the years, moving into Europe in the middle ages, even though documentation on infection was sparse in this region. Nonetheless, there is documented archaeological evidence of TB from widespread sites from all over Europe, dating from the fall of Rome in the 5th century to the 15th century. As cases of TB infection spread from Italy to the northern part of Europe, communities started to become aware of the symptoms of the disease. Laennec, who is known today for inventing the stethoscope, was the first to clearly describe TB pathogenicity, consolidating the concepts of TB infection from the classical pulmonary and extrapulmonary descriptions. Indeed, modern understanding of TB can be traced to findings from the work of Laennec, which introduced new terminologies for describing the disease, such as bronchial and vesicular breathing, murmurs and thrills (the latter being vibrations associated with cardiac or vascular murmurs) and pectoriloquy (increased resonance in the lungs, that he described as a sign of tubercular cavities), that are still used to this day.

Presently, TB remains one of the most devastating diseases worldwide, with infection of people of all ages and from across the globe. According to studies by the World Health Organization (WHO) Burden of Disease project, TB is the eighth leading cause of mortality and the tenth cause of DALYs (Disability Adjusted Life Years). Still more disturbing is the fact that its effect is spread among all age groups and is responsible for the deaths of individuals mostly within the age range of 14-44 years.

TB disease and infection is primarily associated with the aerobic bacterium *Mycobacterium tuberculosis* (*Mtb*), which mainly infects the lungs. The spread of the infection is further escalated via the inhalation of aerosolized bacilli from
individuals infected with *Mtb*. In most cases infected persons may overcome the initial *Mtb* infection, resulting in the development of asymptomatic latent TB; where *Mtb* remains dormant but persists within inactivated alveolar macrophages and evades other immune defenses of the body. About 10% of individuals may develop the active disease after infection; where the bacteria activates its virulence factors, undergoes more rapid growth and overcomes the host immune system\(^\text{24}\). As a result of treatment failure, often due to multi-drug (MDR) and extreme drug (XDR) resistant strains, the bacteria propagate and attack the host, leading to death from systemic infection\(^\text{2,7,24}\). HIV (Human immunodeficiency virus) and other diseases that compromise the immune system can increase the likelihood of developing active TB, from either the initial infection or from reactivation of latent TB \(^\text{24}\). About 33% of the world’s population may be infected with latent TB, and there are estimated to be 9 million new incidences of active TB each year, mostly in developing countries. Consequently, the WHO has declared the TB epidemic a global health emergency\(^\text{23}\). Due to this worldwide scourge of TB, both the academic and industrial communities have initiated intensive research with the common aims of understanding details of the biochemistry and physiology of *Mtb*, and to develop new therapeutics, in order to counter the issues of *Mtb* drug- and multi-drug resistance, and the increase in redundancy of classical medications\(^\text{25,26}\). Recently, with the identification of novel target enzyme systems\(^\text{27}\), the development of new drugs has been initiated\(^\text{28}\), with the anticipation that the drift may reverse against the further global spread of MDR-*Mtb* strains resistant to the classical anti-TB drugs. The properties of MDR and XDR strains of *Mtb* are discussed in more detail in section 1.7.
1.5 The pathogen *Mycobacterium tuberculosis*

As discussed earlier, *Mycobacterium tuberculosis* is the causative pathogen for the disease tuberculosis. *Mtb* was first discovered by Robert Koch in 1882, studies for which he was awarded a Nobel prize in 1905\(^29\). In addition to belonging to the *Mycobacterium* genus, *Mtb* also belongs to the sub order of Actinomycetales, a group of Gram-positive bacteria with high GC-content DNA. The *Mycobacterium* genus is quite different from other taxa based on its ability to synthesize mycolic acids as constituents of its cell wall, hence the name *Mycobacterium*. This unique, specialized envelope surrounds these bacteria. It is comprised of an inner, cytoplasmic membrane, an intermediate cell wall with an essential mixture of sugar molecules, and an exterior wall containing mycolic acid based lipids, as well as surface signaling and transmembrane proteins\(^30\). The exact arrangement of the membrane constituents is still being studied. However, a theoretical model has been postulated\(^31\). Advanced microscopic techniques including cryo-electron tomography (CET) have now revealed new features of the unique cell wall of the mycobacteria.

The presence of this unique cell wall in mycobacteria also aids in the resistance of *M. tuberculosis* to external stresses such as macrophage environments and anti-tubercular drugs. A schematic representation of the envelope is shown below. The mycobacterial outer membrane (MOM) is important in giving the mycobacteria their unique pathogenic features. It is comprised of 40% mycolic acids, which are covalently linked to arabinose and galactose subunits called arabinogalactans (or Pentaarabinosyl motifs) that are themselves linked to the peptidoglycan middle layer (the periplasmic cell wall, or PCW) (Figure 1.1).
Some of the major lipids of interest are the LAMs (lipoarabinomannans) of the outer layer. These contribute to the structural integrity and the strength of the layer, making it difficult to design anti-tubercular drugs that compromise its assembly. The porin proteins present in the outer layer are responsible for trafficking molecules across the outer membrane, with roles including the acquisition of nutrients to aid metabolism. These proteins also associate with other surface proteins responsible for signaling.

Figure 1.2. The Mycobacterium Cell wall.

The envelope is subdivided into three layers. The MOM outer layer (mycobacterial outer membrane), which is comprised of mycolic acids and various lipids such as branched and capped portions of LAM (lipoarabinomannans). The middle layer PCW (periplasmic cell wall) consists mainly of peptidoglycans and galactans, as well as some LM (lipomannans) portions of LAM. There are also significant amounts of PIMs (phosphatidylinositol mannosides) that link the PCW to the inner third layer, the CPB (cytoplasmic phospholipid bilayer). This layer consists mainly of basic plasma membrane phospholipids as well as some polyprenyl sugars. Figure adapted from Park S, 2000.

1.6 Immune responses to *M. tuberculosis*

The acute response mediated by macrophages residing in the lungs initiate the early events in the course of infection that result in inflammasome activation and cytokine
production, as well as in the induction of multiple host defence mechanisms, such as initiation of antimicrobial peptide production and generation of reactive oxygen species (ROS)\(^3\) (Figure 1.3). However, in most cases these immune responses fail to prevent \(Mtb\) infection, which must then be contained by adaptive immune responses. In this case, responding T cells produce interferon-\(\gamma\) (IFN\(\gamma\)) and chemokines, which facilitate further T cell recruitment (Figure 1.3). The IFN\(\gamma\) signaling requirement is evident in individuals with mutations in the IFN\(\gamma\) gene, and these individuals are susceptible to disseminated infection by mycobacterial and other intracellular pathogens\(^3\). Although clinical interventions that reduce the risk of mycobacterial infection, such as vaccinations with BCG (\(M. bovis\) Bacillus Calmette-Guerin) or initiation of antiretroviral therapy in HIV co-infected persons, have been consistently shown to inhibit intracellular growth of mycobacteria in \(ex\ liv\) models\(^3\), the results from these effects are limited in terms of true \(in\ liv\) bactericidal activity. The continuous persistence of mycobacteria \(in\ liv\) induces the production of necrotic granulomas, within which mycobacterial growth can be inhibited by a lack of oxygen and nutrients\(^3\). Nevertheless, under these conditions \(Mtb\) switches into a dormancy mode that limits bacterial cell wall synthesis, cell division and the dependence on aerobic respiration\(^3\). The present anti-TB medications (further discussed in subsequent sections) are limited in their ability to penetrate granulomas and thus have reduced effects on dormant bacilli. Also, as a result of sustained inflammation leading to tissue damage, permanent pulmonary disability occurs in about half of patients that survive mycobacterial infection. These patients consistently experience a persistent cough, intermittent breath shortages and a reduced life expectancy, despite a successful tuberculosis cure\(^\_4\)\(^\_4\).
Figure 1.3. Immune response and host-directed therapy against mycobacterial infection.

Inhibition of phagolysosome fusion by mannose-capped lipoarabinomannan (ManLAM) and secretion of ESX-1 (early secreted antigen 6 secretion system 1) facilitates intracellular survival of mycobacteria. This may be overcome by agents that promote autophagy or by the production of cathelicidin and other antimicrobial peptides and polypeptides such as ubiquitin. Adapted from Wallis et al.\textsuperscript{43}
1.7 Importance of Cholesterol to \textit{Mtb} Pathogenicity and Persistence

1.7.1 \textit{M. tuberculosis} and cholesterol: a do-or-die affair

The importance of cholesterol to \textit{Mtb} pathogenicity cannot be over-emphasized. Cholesterol is a major building block of animal cell membranes that is required to maintain membrane permeability and fluidity. Cholesterol is also an important precursor in bile acids synthesis and in steriodogenesis. Plants and fungi can also produce cholesterol, but this is in smaller quantities compared to animals. In contrast, bacterial steriodogenesis has been somewhat controversial. Only the proteobacterium \textit{Methylococcus capsulatus} and the plactomycete \textit{Gemmata obscuriglobus} have been recognized unambiguously to produce enzymes required for steriodogenesis\textsuperscript{44,45}. By contrast, \textit{Mtb} lacks the key enzymes squalene monooxygenase and oxidosqualene cyclase involved in sterol biosynthesis. However, \textit{Mtb} catabolizes cholesterol, using a cluster of genes that include the cytochromes P450 CYP125A1 and CYP142A1 (and possibly CYP124A1) that are required to initiate the degradation of the side chain of cholesterol (and 4-cholesten-3-one) substrates\textsuperscript{46,47}. The \textit{Mtb} H37Rv genome (the most studied pathogenic \textit{Mtb} strain) was also shown to encode the \textit{CYP51B1} gene, a member of a family of cytochrome P450 (\textit{CYP}) genes that produce an enzyme that catalyzes the 14α-demethylation of lanosterol, generating the 8,14-diene derivative in a crucial step in cholesterol biosynthesis\textsuperscript{48}. \textit{CYP51B1} was the first recognized prokaryotic member of the CYP51 family, although its role in \textit{Mtb} remains uncertain.

\textit{Mtb} can persist in the harsh environment of the macrophage and survive attack by the human immune system via mechanisms that are not yet fully clarified\textsuperscript{49,50}. Altered cholesterol levels are speculated to be associated with \textit{Mtb} infection\textsuperscript{51}, with high
cholesterol levels in the diet significantly increasing *Mtb* infection in the lungs as well as reducing immunity to *Mtb*\(^\text{52,53}\). This is because cholesterol is particularly involved in phagocytosis of mycobacteria by macrophages\(^\text{54,55}\). Interestingly, phagocytes engulf mycobacteria via segments of the membrane rich in cholesterol\(^\text{56}\). Also, studies have shown that cholesterol is a major requirement for retaining the host protein coronin 1, also known as TACO (tryptophan aspartate containing coat protein – a host associated cytoskeleton–plasma membrane linking molecule in leukocytes\(^\text{52}\)), on *Mtb* infected phagosomes, which results in the inhibition of phagosome–lysosome fusion\(^\text{57,58}\). This study shows clear evidence suggesting that cholesterol is a key component of *Mtb* virulence and persistence. In fact, the higher the cholesterol levels in host macrophages, the higher the rate of *Mtb* lung infection and impairment of host immunity\(^\text{52}\).

1.7.2 *M. tuberculosis* catabolizes cholesterol

*Mtb* utilization of nutrients during infection is not yet fully understood, but it has been hypothesized that host lipids play an important role in *Mtb* survival in the lungs\(^\text{57}\). Various studies suggest that the mycobacteria utilize fatty acids as the major source of carbon, rather than carbohydrates and other macromolecules\(^\text{59}\). Experiments on respiration of *Mtb* in mouse lungs showed that rate of infection is increased by fatty acids, but is unresponsive to carbohydrates\(^\text{57,59}\). Mce4 (an ABC-like transport system) has been identified in *Mtb*, and is involved in cholesterol import. The deletion of this *mce4* gene resulted in growth defects of the *Mtb* deletion strain exposed to cholesterol as the primary source of carbon\(^\text{60}\). The *mce4* gene is also found in other cholesterol catabolizing actinobacterial species such as *Rhodococcus jostii*\(^\text{60,61}\), *M. smegmatis*\(^\text{62}\), and *M. bovis* BCG\(^\text{61}\). Studies with \(^{14}\text{C}\) isotope radiolabelled cholesterol derivatives have shown that *Mtb* degrades cholesterol\(^\text{63}\). It was shown that the carbon atoms in the
sterol (cholesterol) framework and those in the aliphatic side chain, respectively, ended up being used for energy production and for lipid biosynthesis in the bacterium\textsuperscript{63}. Yang and coworkers further used a lipidomics approach to show that cholesterol metabolism in \textit{Mtb} significantly elevated the lipid virulence factor phenolthiocerol dimycocerosate (PDIM), probably as a result of high influx of cholesterol metabolites derived from the breakdown of cholesterol\textsuperscript{64}. Furthermore, studies with heptadeuterated 25,26,26,27,27,27-D\textsubscript{7}-cholesterol conclusively explained the function of the \textit{Mtb} CYP125A1 P450 in the oxidation and breakdown of cholesterol, and demonstrated the incorporation of the side chain into the virulence factor PDIM, confirming the crucial role of cholesterol in the chronic phase of infection\textsuperscript{65}.

1.8 Treatment of Tuberculosis – The journey so far

While the current medications for treating drug-sensitive TB are highly effective when there is optimum adherence of patients, as well as under trial conditions, the generalized outcomes are relatively far from ideal when administered to TB-infected patients who generally do not adhere strictly to dosage, which is a common problem in a real-life programmatic condition\textsuperscript{66-68}. Other recommended regimens by the WHO for treating drug-sensitive and drug-resistant tuberculosis also usually have several inherent problems\textsuperscript{69}, some of which are discussed below.

The first problem is that the duration for treating drug-sensitive TB is long – usually lasting about 6 months\textsuperscript{69}. Also, the four most effective agents from the first-line panel of oral drugs, specifically rifampicin, isoniazid, pyrazinamide, and ethambutol, must all be administered in a single prescription for the first two months of treatment, and
two of the four must be taken for a subsequent four months in the continuation phase, resulting in patient adherence issues. When administered in sub-optimum conditions, this results in increased rates of mortality and, in most scenarios, creates chronic cases of infectious drug-resistant tuberculosis\textsuperscript{69,70}. Therefore, a major aim is to develop novel TB drugs that will shorten regimens. Other work on the life cycle of \textit{M. tuberculosis} showed that mycobacteria can develop a dormancy phenotype under nutrient depletion and anaerobiosis conditions, and in this state they are tolerant to several anti-TB drugs, which may be why long treatment durations are usually required\textsuperscript{71-75}.

The second problem is the need for new drugs to tackle the growing global situation of multidrug (MDR) and extensively drug-resistant (XDR) tuberculosis\textsuperscript{76}. MDR-TB is caused by the resistance of bacilli to at least rifampicin and isoniazid, and is becoming prevalent around the globe, with at least five hundred thousand infections recorded in 2012\textsuperscript{77}. Around 92 countries have reported incidences of XDR-TB, which is caused by bacilli being resistant to both rifampicin and isoniazid, as well as being resistant to at least one of the three injectable, second-line anti-TB drugs (kanamycin, amikacin or capreomycin) and to any of the fluoroquinolone drug series\textsuperscript{77}.

The third problem is the issue of interaction between anti-TB drugs and antiretroviral medications, which has been a major problem in Africa, where TB infections are in most cases co-infections with HIV. The fourth problem is the need for novel drugs for treating latent TB infected patients before latent TB switches to the active form of the disease\textsuperscript{78}. Consequently, efforts have been made in the past five years to repurpose and re-use old drugs such as thalidomide and clofazimine (both originally used in
treatment of leprosy) to discover new compounds. Progress in the latter field has led to important developments, with a number of newly discovered drugs going into the preclinical stage of testing (Table 1.1)\(^{79,80}\). Researchers shifting from a single-enzyme targets approach to whole bacterial cell phenotypic screening have helped achieve such progress. Among novel compounds identified are diarylquinolines, benzothiazonines and imidazopyridine amide\(^{81-83}\). Their modes of inhibition are described in Figure 1.4 and most of these drugs are now in the preclinical stage of drug development.

**Table 1.1.** Various drugs and progress made towards repurposing them for TB treatment

<table>
<thead>
<tr>
<th>Name</th>
<th>Class</th>
<th>Current use</th>
<th>Stage of repurposing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivermectin</td>
<td>Avermectin</td>
<td>Anti-helminthic</td>
<td>Anti-TB property detected by MTT(^1) assay</td>
</tr>
<tr>
<td>Carprofen</td>
<td>2-arylpropanoid acid NSAID(^2)</td>
<td>Analgesic</td>
<td>Anti-TB property detected in vitro by HT-SPOTi(^3)</td>
</tr>
<tr>
<td>Chlorpromazine</td>
<td>Phenothiazine</td>
<td>Anti-psychotic</td>
<td>Mouse model studies using MDR-TB strains</td>
</tr>
<tr>
<td>Disulfiram</td>
<td>Thiocarbamate</td>
<td>Alcohol withdrawal drug</td>
<td>Anti-TB property detected by broth dilution tests</td>
</tr>
<tr>
<td>Gatifloxacin</td>
<td>Fluoroquinolone</td>
<td>Respiratory infections</td>
<td>Phase III: enrolment complete</td>
</tr>
<tr>
<td>Metronidazole</td>
<td>Nitroimidazole</td>
<td>Broad spectrum antibiotic</td>
<td>Phase II completed</td>
</tr>
<tr>
<td>Linezolid</td>
<td>Oxazolidinone</td>
<td>Gram-positive bacteria</td>
<td>Phase II completed</td>
</tr>
<tr>
<td>Linezolid</td>
<td>Oxazolidinone</td>
<td>Gram-positive bacteria</td>
<td>Phase II completed</td>
</tr>
<tr>
<td>Moxifloxacin</td>
<td>Fluoroquinolone</td>
<td>Acute bacterial sinusitis</td>
<td>Phase III: in enrolment</td>
</tr>
<tr>
<td>Nitazoxanide</td>
<td>Nitrothiazole</td>
<td>Antiprotozoal</td>
<td>In vitro activity detected</td>
</tr>
<tr>
<td>Oxyphenbutazone</td>
<td>Pyrazolidinedione</td>
<td>Analgesic</td>
<td>In vitro activity detected</td>
</tr>
</tbody>
</table>

\(^1\)MTT is a tetrazolium dye (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) used to assay cellular metabolic activity. \(^2\)NSAID is a class of drugs referred as Nonsteroidal anti-inflammatory drug. \(^3\)HT-SPOTi is a culture assay technique called high throughput spot culture growth inhibition.
Figure 1.4. Mechanisms of action of tuberculosis drugs.

Figure adapted from Zumla et al (2014).
In addition to phenotypic screening, genomic studies have also proved important in characterizing *Mtb* and identifying potential new drug targets. These studies were led by the unravelling of the genome sequence of the virulent *Mtb* H37Rv strain by Cole et al., revealing numerous genes with unknown functions and other unexpected findings, such as annotation of genes coding for kinases, phosphatases and for an adrenodoxin reductase homologue, all of which are typically thought of as eukaryotic proteins. Such genes are probably derived from earlier host organisms and have been assimilated by the bacterium during its evolution. Also, among these unanticipated genes of the *Mtb* H37Rv genome sequence were 20 different *CYP* genes coding for cytochrome P450 enzymes. This large number of *CYP* genes was uncommon for a bacterium, likely unveiling novel and pivotal roles of the P450s in *Mtb*. As a result, studies over the past ten years have focused on characterizing these *Mtb* P450s and have shown new, crucial roles played by these enzymes in the survival of *Mtb*, leading to their potential exploitation as novel drug targets. In the next sections of this chapter, a review of the chemistry of cytochromes P450, including their structure and redox properties is presented. Furthermore, azole anti-fungals and their anticipated anti-TB properties are outlined. Finally, a review of the *Mycobacterium tuberculosis* cytochrome P450s, their importance as novel anti-TB drug targets, as well as their involvement in the metabolism and persistence of the bacterium are discussed.

### 1.9 The Cytochrome P450s (CYPs)

The P450s (CYPs) are heme *b*-containing enzymes (see Figure 1.5 that illustrates the differences between the three major heme groups), which catalyze the reductive scission of molecular oxygen at the heme iron in the active site of the P450, which
results in mono-oxygenation of a substrate in close proximity, and in the generation of a water molecule from the second oxygen atom\textsuperscript{87}, according to the scheme below:

\[
\text{RH} + \text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{ROH} + \text{H}_2\text{O}
\]

P450s are so called because of the unique absorbance characteristics of their ferrous-carbon monoxide bound complex, positioned at \~450 nm for the heme thiolate complex, hence the name pigment 450 or P450. This main absorption feature is referred to as the Soret band of the P450 heme protein (Figure 1.6).

The protein thiolate ligand \textit{trans} to the CO comes from a conserved cysteine residue and is crucial for P450 activity as a consequence of its electron donating character. The Fe(II)-CO absorbance shifts to a lower wavelength of \~420 nm when the cysteine thiolate is protonated to thiol (the P420 form)\textsuperscript{88}. This protonation of cysteine thiolate (further discussed in the next section) is associated with P450 inactivation, although deprotonation of the cysteine thiol can occur in certain P450s to reactivate the enzyme\textsuperscript{47}.
Figure 1.5. The structures of the major forms of heme.

Hemes are cofactors consisting of an Fe$^{2+}$ (ferrous) iron positioned in the centre of a heterocyclic ring referred to as a protoporphyrin. The most common are hemes $a$, $b$ and $c$. The differences in the structure of heme $a$ from heme $b$ is that the methyl side chain at ring position 8 is oxidized to a formyl group and that a hydroxyethylfarnesyl group, an isoprenoid chain, is attached to the vinyl side chain at ring position 2. The difference between heme $c$ and heme $b$ is that the two vinyl side chains of heme $b$ are replaced by covalent, thioether linkages to the apoprotein.

Figure 1.6. Absorption spectra of cytochrome P450.

A typical absorption spectrum for a P450 enzyme (CYP51B1 from *M. tuberculosis*). The oxidized (thin line) ferric state is shown with absorbance at ~420 nm. The dithionite-reduced, ferrous CO-bound complex has an absorbance of ~450 nm (thick line). The major absorption (Soret) band shift to ~450 nm in the CO complex is a hallmark of the cysteinate-coordinated heme iron of P450 enzymes. Figure taken from Munro et al. 2007.

The schematic representation below shows various processes involved in the P450’s canonical catalytic pathway (Figure 7). P450s are mono-oxygenases and are best known for their involvement in human liver drug metabolism and detoxification, and for their roles in steroidogenesis and the synthesis and inter-conversion of steroids in the adrenal glands$^{90,91}$. Evolutionary and phylogenetic studies on P450 have shown that these enzymes may have evolved in archaea about 3000 million years ago$^{92}$. However, the present human xenobiotic
metabolizing P450s may date back to during the Devonian and Silurian periods about 400 million years ago, at the time of the co-evolutionary “warfare” between plants and animals due to increasing atmospheric oxygenation and to detoxification of plant toxins (in part produced using P450s) by evolution of animal P450s\textsuperscript{92}. As a result, P450 isoforms span from archaea through prokaryotes (including bacteria) and on to eukaryotes (plants and animals). The P450s in eukaryotes are integral membrane proteins, associated with the membrane by an N-terminal membrane-spanning helical segment of the protein, and interact with their redox partner enzymes (the FAD- and FMN-binding NADPH-cytochrome P450 reductase)\textsuperscript{93} in the case of microsomal P450s, and with the flavoprotein adrenodoxin reductase and the iron-sulfur protein adrenodoxin in the case of the adrenal mitochondrial P450s). The P450s catalyze the reductive scission of oxygen (O\textsubscript{2}) bound to their heme iron in the active site by the timely provision of two electrons to enable oxygen activation and the generation of a highly reactive oxidizing species (compound I, a ferryl-oxo porphyrin radical cation) and the insertion of a single oxygen atom into the substrate\textsuperscript{94}. On the other hand, the prokaryotes and archaea have soluble, cytosolic P450s that require interaction with other soluble partner proteins (typically ferredoxin reductases and ferredoxins) in order to catalyze a similar reaction process to those seen in the eukaryotes (albeit it often with quite different substrates) using distinct redox apparatus. Even though they are located in different cellular environments in prokaryotes/archaea and eukaryotes, virtually all the P450s catalyze the reductive scission of molecular oxygen in a similar manner in these different types of organisms\textsuperscript{87}. The typical reaction process involves the donation of two electrons to the heme iron from NAD(P)H by one or more redox partners, along with the donation of two protons, usually from bulk solvent with proton transfer to heme iron-oxo complexes mediated by specific active site amino
acid side chains\textsuperscript{95}. In the case of the well-studied \textit{Pseudomonas putida} camphor hydroxylase P450cam (CYP101A1), a conserved acid/alcohol amino acid pair (Asp251 and Thr252) are crucial in proton delivery to iron-oxo species in the P450cam catalytic cycle. A similar mechanism is retained in most other P450 monooxygenase enzymes\textsuperscript{96}.
Figure 1.7. Cytochrome P450 reaction mechanism.

The various intermediates involved in the catalysis of oxidation of a substrate (R-H) to product (R-OH) by a P450 are shown (a-g). The oxidation state of the heme iron for each stage of the cycle is depicted on the heme structure. The cysteine thiolate ligand is depicted as an S-atom positioned at the proximal side of the heme, and the distal water molecule (changing to dioxygen and other oxy-intermediates as the reaction proceeds) is also shown. The first step involves the displacement of the water ligand by the approaching substrate (R-H). This results in a more positive potential of the ferric heme iron as it favours the shift from low spin to high spin ferric heme iron, enabling the transfer of an electron (e\(^{-}\)) from the redox partner. The resulting ferrous heme iron binds oxygen to form the ferrous-oxy intermediate (which is in equilibrium with the ferric-superoxy form). A second electron from the redox partner reduces the heme iron to the ferric-peroxy state. The ferric-hydroperoxo form (compound 0) is then produced via protonation. A second protonation results in scission of the bound dioxygen, leading to the release of a water molecule and the formation of the reactive ferryl-oxo porphyrin radical cation species called compound I. The compound I is considered to be the major species responsible for oxidizing the substrate in most P450s. With the substrate already in close proximity, compound I abstracts a hydrogen atom to form the ferryl-hydroxo form (compound II), and this species (through a process referred to as “radical rebound”) reacts with the substrate radical formed to hydroxylate this species. The hydroxylated product is then released and a water molecule re-ligates to the distal face of the ferric heme iron, completing the catalytic cycle. Non-productive pathways leading to collapse of oxy intermediates are also depicted. Adapted from Munro et al 2007\(^{89}\).
1.10 The Structure of Cytochrome P450

All known P450 enzymes have a similar structural fold and conformation. Although they have widely different levels of amino acid sequence similarity, they are all still globular proteins that are assembled into what is commonly referred to as the “P450 fold”. This P450 tertiary structure is mainly a formation of alpha-helices with a small domain with a substantial component of beta-sheets. While each secondary structural element contributes key amino acids need for folding, flexibility/stability and catalysis for the various roles of the different P450 enzymes, their overall amino acid sequences can be very different. However, they all contain approximately 80 percent alpha-helices, 15 percent beta-sheet, and 5 percent loops and turns\(^{97,98}\). The alpha helices of all P450s are assigned alphabetically (A-L) from the N-terminal. The construction of the helices varies considerably between the enzymes, especially for the shorter ones and those at the periphery of the P450. Key examples are the helices A, B and K that show variations in length between P450s and are sometimes assigned with a gap, with two parts of the helix joined by a short linking peptide. A hypothesis to explain this is that these differences may arise from difficulties in the assignment of secondary structure due to flexibility and movement shown by peripheral elements, resulting in disordered regions as defined from electron density in structures elucidated by X-ray crystallography. Alternatively, unstructured regions could be essential for helical flexibility and to enable the binding of larger substrates. The Figure 1.8 below shows a representation of a typical P450 secondary structure.
A common feature retained in all P450s is the heme-binding motif. A sequence of amino acids present in the loop region just before the start of the L-helix is responsible for positioning the heme group, and is essential for P450 catalytic activity. The heme iron is coordinated to a conserved cysteine residue that is critical to the binding motif. The cysteine’s side chain contains a sulfur atom that acts as the heme iron’s proximal 5th axial ligand in its octahedral geometry. This cysteine ligand, in its thiolate form, is responsible for the characteristic P450 Soret absorbance spectrum, as discussed earlier. The characteristic absorbance feature of the P450 Fe$^{2+}$-CO species is also what provides the name for this enzyme superfamily. The ferrous-CO complex may undergo a change in proximal coordination through thiolate protonation to the thiol, converting it to the P420 species, as discussed earlier. This can be visualized by a shift in the UV-visible Soret spectrum from $\sim$450 nm to $\sim$420 nm$^{99}$. This P420 species is considered to be an inactive form of the enzyme$^{100}$. However, more recent studies suggest that the formation of these P420 species may in fact not be associated with permanent inactivation and that there is an equilibrium between the two species that may be perturbed by e.g. adjusting the pH or by addition of a P450 substrate$^{101,102}$.
The cysteine side-chain sulfur that coordinates the heme iron may thus exist either in the unprotonated thiolate state or as a protonated thiol. The protonation process is the basis of the P450-to-P420 transition chemistry\(^8\). The figure below depicts a typical P450/P420 transition mechanism (Figure 1.9).

![Diagram of cysteine protonation](image)

**Figure 1.9.** A schematic representation of cysteinate protonation

### 1.11 Cytochrome P450 redox partners

As can be observed in the P450 catalytic cycle above, the reaction process catalyzed by P450s requires a source of electrons and protons. The electrons, in most cases, are supplied by NADH and NADPH, while the protons are abundantly available from the bulk solvent and delivered to the heme iron using a conserved acid-alcohol amino acid pair in the P450 active site (first recognized as Asp251 and Thr252 in P450cam)\(^103,104\). The protons are thus shuttled onto the oxygen atoms of the ferric peroxo and ferric-hydroperoxo species in the catalytic cycle using a conserved amino acid motif in many P450s. A notable example is observed in the *Bacillus megaterium* fatty acid hydroxylase P450 BM3, where mutations of key active site residues (T268 and F393) reduced the ability of the enzyme to proceed through the classical P450 catalytic cycle. The T268A mutation is postulated to hinder the donation of protons from the T268 side chain hydroxyl group to ferric-peroxy and ferric-hydroperoxy intermediates, while the F393 residue modulates the reduction potential of the heme
iron through its influence on the cysteinate heme ligation. The F393H and F393W mutations perturb the heme iron potential, decrease electron transfer rate to the heme iron, alter the stability of the ferrous-oxy complex and diminish the mutant’s ability to activate oxygen\textsuperscript{105}. In addition, the conserved T252 and T303 residues in the active sites of P450cam and CYP2E1 (purified from rabbit) have also been shown to be involved in the donation of protons to the iron-bound dioxygen species and in stabilization of these species, consistent with a retained role of the acid-alcohol amino acid pair in P450 catalysis from bacteria through to mammals\textsuperscript{104,106}.

An initial hypothesis was that P450 enzymes interacted with redox partners that can be classified into two major categories (class I being bacterial-like ferredoxins and ferredoxin reductases; and class II being eukaryotic CPRs). However, this general classification is gradually being unraveled. Studies continue to unveil new examples of systems specifically evolved to employ redox machinery to perform electron transfer from NAD(P)H to specific P450 enzymes via novel routes, while certain other P450s do not even employ such redox partners for electron transfer\textsuperscript{94,107}. Nevertheless, most known P450s utilize redox partner systems that fall into the major categories of the class I and class II model systems.

Initial 1970’s studies of CYP101A1 (P450\textsubscript{cam}) from \textit{Pseudomonas putida}\textsuperscript{108-110} showed that electrons were transferred through a redox system containing three soluble enzymes. In this case, electrons derived as a hydride ion (two electrons and one proton) from NADH are transferred to an FAD-containing reductase called putidaredoxin reductase (Pdr), and then, one at a time, the electrons pass onto the P450 through electron transfer mediated by the Pdr FAD onto the iron-sulfur protein
putidaredoxin (Pd). This electron transfer mechanism from NADH to a ferredoxin reductase, to a ferredoxin and then onto the P450 was classified as the bacterial or prokaryotic system, and in the class I category. The steroid-metabolizing P450 systems in eukaryotic mitochondria utilizes a similar type of redox system, with the class I P450 redox partners here being the FAD-binding adrenodoxin reductase and the 2Fe-2S cluster-binding adrenodoxin. Thus, the class I system is clearly also used in higher eukaryotes as well as in prokaryotes/archaea. The alternative class II mechanism, initially thought to be exclusive to eukaryotes, was described from studies on liver P450 systems. A notable example is in the CYP3A subfamily (which comprises about 30% of all the liver P450s) and the sterol demethylase CYP51, which are respectively responsible for the metabolism of a large number of drugs, and for the catalysis of the rate-limiting step in the cholesterol biosynthetic pathway, which involves the removal of the 14α-methyl group of lanosterol in fungal systems. The key FAD- and FMN-containing enzyme, called NADPH-cytochrome P450 reductase (CPR), was discovered to be the single redox partner enzyme responsible for shuttling electrons directly from NADPH to the P450, and thus was termed a “diflavin reductase”. The electrons are received as a hydride ion from NADPH by the FAD, and are then passed one at a time onto the FMN, and then to the P450 heme iron. In this case, these enzymes (CPR and P450) are both membrane-bound forms, anchored to the microsomal membrane by a N-terminal transmembrane helical segment.

This class I/class II classification system was the accepted model for a few years, until the discovery of P450 BM3 (CYP102A1) from the soil bacterium Bacillus megaterium. This P450 contains a fatty acid hydroxylase P450 attached to a
dilflavin reductase CPR, but with no membrane anchor regions on either of these domains, meaning that CYP102A1 is highly soluble. The discovery of this novel, single component, class II-like system in a bacterium raised speculations that there were likely other alternative P450 redox partner systems in Nature, other than the aforementioned class I/class II models. BM3 was also the first example in which the redox partner is fused to the P450$^{117}$. The scheme below shows various selected classes of P450 redox systems (Figure 1.10).
Different electron transfer systems in P450 catalyzed reactions.

(a) Three-protein systems: The P450 and reductase partners can either be soluble or membrane-bound (class I) – as first described for the D-camphor 5-exo hydroxylating CYP101A1 (P450cam) from *Pseudomonas putida* and initially considered to be the generalized redox system in prokaryotes. (b) Two-protein systems: CPR and P450 are both membrane-bound (class II) – and were considered to be the generalized redox system employed by eukaryotic P450s. Subsequent studies showed that mitochondrial P450s instead use a class I system in which the P450 and reductase (adrenodoxin reductase) components are membrane-associated. (c) One-protein systems: soluble or membrane-bound systems formed by fusions of a CPR-like reductase and P450 (class III) – first demonstrated for the long-chain fatty acid hydroxylating CYP102A1 (P450 BM3) from the soil bacterium *Bacillus megaterium*. Instead of using a ferredoxin, some bacterial P450s employ flavodoxins as the electron carriers. This type of system was shown to be functional for mammalian P450c17 (when expressed in *E. coli*, and using the host flavodoxin and flavodoxin reductase enzyme system), and naturally in the cineole-metabolizing P450cin (CYP176A1 from *Citrobacter braakii*), and for the *Bacillus subtilis* fatty acid oxygenase CYP107H1 (P450 Biol). The system is also used in the more recently identified P450:flavodoxin fusion protein XplA (CYP177A1) from the soil bacterium *Rhodococcus rhodochrous*, which degrades the recalcitrant explosive RDX (hexydro-1,3,5-trinitro-1,3,5-triazine). Other non-redox partner systems utilized by other P450s have also been described. These require none of the systems above and instead employ hydroperoxides or organic peroxides, thereby bypassing the requirement for NAD(P)H, molecular oxygen and redox partners. Examples are the CYP74A family (plant P450 allene oxide synthase) and CYP5A1 (mammalian thromboxane synthase) – both of which are isomerases; CYP152A1 (a *Bacillus subtilis* fatty acid hydroxylating P450) and related enzymes which use hydrogen peroxide to replace NAD(P)H-dependent redox partner systems; and CYP55A1 (a *Fusarium oxysporum* nitric oxide reducing P450) which uses NAD(P)H directly to reduce heme iron-bound nitric oxide. Figure adapted from Urlacher VB et al., 2012.
1.12 The activities of bacterial and eukaryotic cytochrome P450s

As discussed briefly above, a key disparity between bacterial and mammalian CYPs is the oxidoreductases and electron transferases that partner with the P450s to enable their catalytic activity. Bacterial P450s are generally reduced using a NAD(P)H-binding and FAD-containing ferredoxin reductase, which donates electrons to the P450 via an iron-sulfur cluster-containing ferredoxin. Also, most P450s currently recognized in prokaryotes are “orphan” enzymes with no known activity or substrates identified. However, those with known functions catalyze various reactions ranging from catabolism of carbon compounds for the growth of the host bacterium through to the biosynthesis of secondary metabolites with useful bioactivities, such as antifungals and anti-parasitics. Typical examples include oxidative modification of polyketides, such as the hydroxylation of the erythromycin precursor 6-deoxyerythronolide B, and the epoxidation of epothilones C and D to A and B, respectively, which are known anticancer agents. The structural disparities between different P450s (due to active site alterations that enable distinctive substrate recognition) show promise for developing potent inhibitor specificity between different P450 isoforms. Furthermore, it is important to note that, between catabolic and biosynthetic P450s, biosynthetic P450s are more often the target for development of therapeutics due to their high substrate specificity. Therefore, Mtb P450s involved in biosynthesis are more likely to be the focus of anti-TB drug targets, with the anticipation that their biosynthetic pathways are different to those in the human host.

1.13 Azoles and Mtb CYP inhibition

Studies on azoles that inhibit CYP51, and which are also used clinically for the treatment of skin and systemic fungal infections, have presented these compounds as
potential new types of anti-TB drugs. Most of these azole anti-fungals show a wide profile of P450 inhibition in vivo. They also show tight binding to P450s in vitro with high affinity (some with nanomolar $K_d$ values). This inhibition of P450s is usually through a reversible coordination of the azole nitrogen in the sixth distal ligand position with the heme iron in the active site, a mechanism called type-II inhibition. This results in the displacement of the water molecule that typically occupies this position in the P450 resting state, as well as competitively inhibiting the substrate from accessing the active site. Anti-mycobacterial activity has also been noticed in many azole antifungals. Therefore, if we hypothesise that azole antitubercular activity is also due to P450 inhibition, consequently, the results should strongly support the speculations about the essentiality of various Mtb CYP genes. Some of the well-studied azole anti-fungals that exhibit potent inhibition on Mtb P450s include: ketoconazole, clotrimazole, econazole and miconazole. Selected members of the triazole class (e.g. fluconazole and voriconazole) have proven less potent (Figure 1.11). This anti-mycobacterial potency of azoles has shed light on a possible new route for the discovery of novel regimens for anti-tubercular drugs, thereby instigating research in this area. Consequently, this has also led to the development of various novel compounds with imidazole, triazole, pyridine and other groups that have the capability of coordinating with P450 heme iron. However, these ongoing studies remain in their early stages.
Achievements in the development of drugs that inhibit human biosynthetic P450s (such as letrozole and anastrozole, which are potent inhibitors of aromatase) have helped formulate the hypothesis that, through understanding the functions of the unexpectedly high number of CYPs present in the genome of the *Mtb* H37Rv strain, there may be a route to the discovery of novel therapeutics for this persistent human pathogen\(^{86,142-144}\). As shown in the phylogenetic tree below, there are 20 *Mtb* CYP genes encoded in the *Mtb* genome and their localizations on the chromosomes generally provide little information about their physiological roles (Figure 1.12). This, in part, is due to the fact that they do not reside in gene clusters or operons of defined function, with the notable exception of the cholesterol oxidases CYP125A1 and CYP124A1, which reside in a gene cluster dedicated to the catabolism of host cholesterol\(^{46}\). Presently, approximately seven of the 20 *Mtb* P450s have been expressed and characterized\(^{46,47,115,130,145-147}\).

**Figure 1.11.** Chemical structures of selected azole antifungals.

1.14 **CYPs of the *M. tuberculosis* Genome**
Figure 1.12. Phylogenetic analysis of the amino acid sequences of the 20 *Mtb* P450 enzymes.

The seven P450 isoforms that have been characterized with structures solved are depicted by blue circles. Further characterization and P450s structure determination has been achieved for CYP126A1 (in this thesis) and for CYP143A1 (Swami et al, unpublished data). The tree was generated after amino acid sequence alignment between the various enzymes. The clustering of CYPs 124A1, 125A1 and 142A1 may infer similar substrate specificity for sterols and fatty acids.\textsuperscript{46,47,131,146} The presence of CYP101A1 (P450cam) and *Mtb* CYP143A1 on adjacent branches may also hint at both structural and functional similarities in these enzymes. From the previously characterized CYPs, the isoforms that show most promise as anti-TB drug targets are CYPs 121A1 and 125A1, which are considered essential for bacterial viability and infectivity, respectively.\textsuperscript{46,56,129,145} Other isoforms that show promise as potential secondary drug targets are CYPs 124A1 and 142A1, which are likely to play important roles in catabolism of host cholesterol.\textsuperscript{47,146} The figure is adapted from McLean et al., 2006.\textsuperscript{86}

It has been difficult to assign specific functions to many of the *Mtb* P450s due to a significant lack of sequence similarity to other P450s of known function, as well as extensive divergence amongst the individual *Mtb* CYPs. Only *Mtb* CYP135A1 and CYP135B1 are sufficiently closely related to be classified in the same P450 family, although their roles remain obscure. CYP51B1 was shown to be a sterol demethylase, but its role in *Mtb* remains unknown. Nonetheless, the involvement of many of these
P450s in pathogenicity is still a high possibility. Subsequent genome studies with *M. smegmatis* and *Streptomyces coelicolor* revealed 39 and 18 CYPs in these organisms, respectively, pointing to multiple important functions for P450 enzymes in the actinobacteria. Surprisingly, analysis of the *E. coli* genome sequence revealed this bacterium is devoid of any P450s, making this organism a good host for the heterologous expression of other P450 enzymes, but also again highlighting the importance of the P450s in the biochemistry and likely in the virulence of pathogenic mycobacteria, and probably also in other pathogenic bacteria. Consequently, various projects have been initiated to explain the roles of these P450s in *M. tuberculosis*. Below, a review of those P450s studied and characterized is presented, as also highlighted in Figure 1.12 above.

**1.14.1 CYP51B1**

CYP51B1 was the first of the *M. tuberculosis* CYPs to be studied. CYP51B1 was shown to have ~30-40% amino acid sequence identity to CYP51s in eukaryotic organisms. CYP51 enzymes in eukaryotes catalyze a 14α-demethylation reaction during steroidogenesis (Figure 1.13). CYP51B1 in *Mtb* was also confirmed to catalyze the demethylation of lanosterol and of 24,25-dihydrolanosterol following its expression and purification by Bellamine and co-workers. Even though this oxidative demethylation process is always associated with steroidogenesis in eukaryotes, it is not yet clear how this activity is relevant to *Mtb*, since several components of the complete steroidogenesis pathway are not present in *Mtb*. Crystal structures of the ligand-free, substrate-bound, as well as inhibitor-bound forms of CYP51B1 have been presented. From these data, a detailed overview of its structural and conformational dynamics has been extrapolated. For instance,
because of the flexible nature of the BC-loop region and the F- and G-helices, catalysis is likely achieved through enabling “large scale” conformational change\textsuperscript{150}. A fluconazole-bound CYP51B1 structure, in comparison with CYP51s in other eukaryotic organisms, also provides insights into the mode of binding of fluconazole and other azole inhibitors, as well as giving insights into the mechanisms of drug resistance in selected fungal species\textsuperscript{151}.

In addition, CYP51 is the enzyme targeted by azole antifungals in the treatment of fungal infections. Thus, its presence in \textit{M. tuberculosis} (as CYP51B1) supports the hypothesis that the P450s in \textit{Mtb} are suitable drug targets in the pursuit for novel anti-tubercular drugs. Further, the data suggest these azole antifungals may be important new compounds for the treatment of TB and that design of new azoles may provide important anti-tubercular drugs. In view of this information, ligand titration studies with azoles (particularly clotrimazole, econazole and miconazole) revealed their tight binding to CYP51B1, with selected azoles (econazole and clotrimazole) exhibiting binding affinities in the nanomolar range\textsuperscript{129}. These findings were consequently the “torch-bearing” factors that stimulated further investigations into the remaining \textit{Mtb} P450s as potential anti-tubercular targets and azole-binding enzymes.

\subsection{1.14.2 CYP121A1}

CYP121A1 was the second \textit{M. tuberculosis} P450 to be characterized by McLean et al\textsuperscript{56}, and was shown to be essential for the viability of \textit{Mtb} grown \textit{in vitro} using gene disruption studies\textsuperscript{152,153}. Also, CYP121A1, along with CYP128A1 and CYP141A1, are the only \textit{Mtb} P450s that appear to be conserved only within members of the \textit{Mycobacterium tuberculosis} complex. This is because their homologues do not
appear in the genomes of the other members of the wider family of actinobacteria. The structure of the ligand-free CYP121A1 was also the second Mtb P450 structure to be solved (following CYP51B1)\(^{154}\). The ligand-bound structure with fluconazole, complemented by spectroscopic data, revealed that the enzyme coordinates this azole in two different modes. The first mode showed the azole nitrogen directly coordinated as the heme iron’s 6\(^{th}\) axial ligand, which required a significant shifting of the I-helix away from the heme, creating a wider active site cavity for the approaching inhibitor. The second mode of fluconazole binding was possible without this structural modification, and the azole here coordinates instead through the axial water molecule, which is retained as the 6\(^{th}\) heme iron axial ligand (rather than being displaced by the triazole nitrogen, as seen for fluconazole in the first mode)\(^{155}\).

CYP121A1 was also the first M. tuberculosis P450 with an identified physiological substrate. Even though CYP51B1 was demonstrated to exhibit sterol demethylase activity, it is still not certain how this process fits into Mtb metabolic pathways. CYP121A1’s activity was revealed after the gene Rv2275 (immediately upstream of Rv2276, which encodes CYP121A1) was shown to encode an enzyme that catalyses the formation of the cyclodipeptide (CDP) cyclo-L-Tyr-L-Try (cYY)\(^{156}\) from tRNA-bound L-Tyr substrates. CYP121A1 was thereafter also investigated for the binding of this cYY product, and the results indicated that cYY was a genuine substrate, as demonstrated from both structural studies and spectroscopic binding studies. Further studies with NMR and LC-MS showed that CYP121A1 catalyzed the formation of an intramolecular C-C bond between the side chains of the tyrosine molecules in cYY, forming the product mycocyclosin (Figure 1.13).
1.14.3 CYP128A1

CYP128A1 is another P450 that was highlighted by Sassetti et al. from *Mtb* genome-wide gene disruption studies as being essential for *M. tuberculosis* viability. There is still rather little information about this P450, and it has yet to be expressed and purified. Nonetheless, a potential function has been postulated following analysis of the surrounding genes in the cluster that embeds the CYP128A1 gene (Rv2268c). The adjacent Rv2267c gene overlaps the CYP128A1 gene by 3 bp, which raises questions as to the location and characteristics of the Rv2267c promoter sequences. The Rv2267c gene encodes a sulfotransferase enzyme, Stf3, which is involved in the production of the sulfolipid S881. This S881 compound was later confirmed to be a sulfated product of dihydromenaquinone, which is the major electron carrier during *Mtb* respiration. CYP128A1 is postulated to perform a terminal-hydroxylation of this dihydromenaquinone, a reaction that precedes its sulfation by Stf3. A stf3 gene knock-out strain of *Mtb* demonstrated decreased virulence. However, the postulate that the CYP128A1 gene should be essential to the growth process apparently contradicts these findings. However, a process that allows the fluctuation of virulence effects in this bacterium could be viewed as a disguised form of persistence. These data show that a great deal of knowledge is yet to be actualized on how these enzymes function in *M. tuberculosis*.

1.14.4 CYP130A1 and CYP141A1

CYP130A1 was highlighted as a potential *M. tuberculosis* drug target after it was discovered to be absent from the *M. bovis* genome, as well as from its avirulent BCG strain. This case is synonymous to that for CYP141A1, which is located in close proximity to CYP130A1 in the *M. tuberculosis* genome. These findings suggested that
these P450 enzymes are important to the pathogenic process in the human host, while not being essential towards the viability of the bacterium. Consequently, studies by Ortiz de Montellano et al. attempted to assign a physiological function to CYP130A1, as well as to link CYP130A1 to azoles and related inhibitors as a potential antitubercular drug target. While a physiological role was not identified for CYP130A1, ligand-free and econazole-bound crystal structures were solved, which showed similar flexibilities in the BC-loop and FG-helices, and their structural reorganization upon ligand binding, as described earlier for CYP51B1. Further compound screening techniques identified polycyclic arylamines as novel inhibitors for CYP130A1, with the structural studies revealing that mutation of a key residue (G243A mutant) led to a total conversion of some of the arylamines from inhibitor-like (type II) to substrate-like (type I) binding modes. This suggests a role for the conserved A(G)243-G244 I-helix residues in modulating both the binding affinity of the axial water ligand and in substrate selectivity of cytochrome P450 enzymes.\textsuperscript{158}

1.14.5 CYP124A1, 125A1 and 142A1

Further studies on the \textit{Mtb} CYPs CYP124A1, CYP125A1 and CYP142A1 have shown that these enzymes co-operate in a pathway that leads to metabolism of host derived steroids (cholesterol and cholestenone) by the bacterium (\textbf{Figure 1.13})\textsuperscript{46,47,146}. CYPs 125A1 and 142A1 are considered the major enzymes involved in these sterol oxidation reactions\textsuperscript{46,47}, with CYP124A1 likely mainly involved in long, branched-chain fatty acid hydroxylation, albeit also being capable of sterol oxidation activity\textsuperscript{146}. The bacteria achieve this sterol catabolism by oxidizing the aliphatic tail of the cholesterol at C-27 to the alcohol, aldehyde and then to the carboxylic acid\textsuperscript{47}. This process is speculated to be the initial and rate-limiting step in the degradation of sterol side chains to be utilized as metabolic fuel for the bacterium\textsuperscript{47,159}. 

54
Figure 1.13. Scheme for various reactions catalysed by *Mtb* CYPs

The various points of modification and the functional groups added by each enzyme are highlighted in red. CYP51B1 inserts a double bond at position 15 of the lanosterol substrate. CYP125A1 and CYP142A1 perform sequential terminal oxidations on the aliphatic chain of cholesterol, forming the carboxylic acid product. CYP124A1 hydroxylates the terminal methyl group of branched-chain fatty acids, though it can also perform similar reactions to CYP125A1 and CYP142A1. CYP121A1 catalyzes the intramolecular carbon-carbon (C-C) coupling of cyclodityrosine (cYY) to form the product mycocyclosin.
1.15 Conclusion and study objectives

1.15.1 *M. tuberculosis* CYP144A1 and CYP126A1

CYPs have been considered as bacterial enzymes that function primarily in catabolism of unusual carbon sources\(^{127}\). This would seem to indicate that bacterial (including *Mt*) CYPs could not be considered as potential therapeutic targets. However, the recent discovery of 20 *Mt* CYPs from the H37Rv genome sequence by Cole et al.\(^{27}\) launched research into the physiological and pharmacological importance of these enzymes to *Mt*, and in order to establish if they could be targets for novel therapeutics\(^{27}\). To date, there have been reports of expression and purification of 7 of these *Mt* CYPs (CYP51B1, CYP121A1, CYP125A1, CYP130A1, CYP124A1, CYP142A1 and CYP144A1)\(^{46-48,56,130,131,145-147}\). Six of these CYPs (CYPs 51B1, 121A1, 124A1, 125A1, 130A1 and 142A1) have already had their crystal structures reported\(^{47,48,145-147,151,154}\), and in view of the crucial roles (especially CYP121A1 and CYP125A1) demonstrated for these enzymes in *Mt*, further studies on the remaining orphan CYPs are imperative. In the context of this study, I have focused on two orphan *Mt* P450s - CYP144A1 and CYP126A1. Their biophysical and structural characterizations will be extensively elaborated on in the subsequent chapters 2, 3 and 4 of this thesis. Firstly, I will justify the reasons for choosing them as priority subjects.

CYP144A1 is expressed as both a full length and a N-terminal truncated version (from alternative transcripts), and both these forms may be physiologically relevant. CYP144A1 is encoded by the *Mt* gene *Rv1777* (*CYP144A1*). CYP144A1 exhibits tight binding to azole anti-fungal drugs\(^{130}\). Studies by Driscoll et al. showed that (in addition to tight azole binding to clotrimazole, econazole and miconazole) deletion of
the *CYP144A1* gene, even though not being essential for *Mtb* viability, compromised the growth rate of the deletion strain with respect to the wild-type *Mtb* H37Rv strain on exposure to clotrimazole and econazole\(^{130}\). These data hint towards an important role for CYP144A1 in the physiology of *Mtb*. Also, CYP144A1 was first noticed for being amongst a group of *Mtb* genes identified as drug candidates due to their involvement in dormancy in the *Mtb* bacterium, which is one of the tactics employed by this bacterium to evade their destruction when engulfed by macrophages\(^{160}\). This finding was further supported by another study simulating macrophage-like conditions using a dendritic cell model, where *CYP144A1* was found to be upregulated upon subjection to oxidative stress and an acidic environment\(^{161}\).

Regarding its functionality, CYP144A1 (like most other CYPs) does not share close structural similarities with other CYPs of known function, from which clues to CYP144A1’s role might be deduced. In addition, the gene cluster containing CYP144A1 is poorly characterized, with most of the genes within the cluster termed “hypothetical proteins”\(^{162,163}\). Consequently, substrate specificity of CYP144A1 has proved difficult to predict.

CYP126A1 is another *Mtb* P450 that has emerged as a promising drug target for anti-TB drug development. CYP126A1 is encoded by the gene *Rv0778*\(^{162,163}\), which is located near essential *Mtb* genes encoding enzymes involved in the de novo anabolism of purines\(^{163}\). CYP126A1 is also positioned within an operon that encodes PurB, an adenylosuccinate lyase enzyme, raising anticipations that it may also play an important role in nucleotide metabolism in *Mtb*\(^{163}\). Also, bioinformatics show that CYP126A1 shares extensive sequence homology (35% identity) to the *Mtb* cholesterol hydroxylases CYP125A1 and CYP124A1, as well as being highly
conserved across strains of both pathogenic and non-pathogenic actinobacteria, further supporting the model that in has an important and widespread role in these bacteria\textsuperscript{143,164}. Previous studies on predicting the biological function of CYP126A1 involved exploiting a virtual TB metabolome, to analyse the binding of potential substrate-like type-I “biofragments” to this P450. The data collected indicated that this enzyme exhibited a binding preference for aromatics, including those containing chlorophenol scaffolds that bind near the CYP126A1 heme within the active site, suggesting its potential involvement in substrate oxidative dehalogenation\textsuperscript{164-166}.

1.15.2 Research Objectives

Even though CYP144A1 and CYP126A1 have already been expressed and biophysically characterized\textsuperscript{130,165}, most of the data in these studies were collected with full length versions of the proteins. CYP144A1 is expressed as both full length (FLV) and truncated (TRV) versions. The truncated version is 30 amino acids shorter (at the N-terminal region), as defined by the generation of a construct that expresses the TRV from the second methionine residue in the protein (further discussed in chapter two). It was found that the FLV CYP144A1 was routinely proteolytically cleaved at or near this position during \textit{E. coli} expression. Given that both forms may be physiologically relevant in the bacterium, and that both can be expressed successfully, it is of interest to establish the properties of both forms. In order to gain full understanding of the structural features and functional roles of the CYP144A1 forms, there is need to investigate if there are any major differences between the properties of the FLV and TRV forms of CYP144A1. In the course of these studies, I also performed crystallography experiments to structurally characterize both versions of this enzyme. Based on the results from structural studies, I then further investigated novel ligands
in a search for potential substrates, as well as inhibitors that could later be developed into novel anti-TB drugs or used as mechanistic probes for further studies on this enzyme. The results are presented in Chapters 2 and 3 of this thesis.

In the case of CYP126A1, its gene cluster had already hinted at the potential biological activity of this orphan enzyme. Therefore I proceeded to investigate substrate/ligand interactions and to do work towards the development of novel inhibitors of this \textit{Mtb} P450 enzyme. I cloned, expressed and purified intact CYP126A1 for biophysical characterizations. Mass spectrometry and Nano-electrospray ionization Mass Spec (NanoESI) were utilized to determine the molecular weight of the protein and to analyze its various oligomerization states in its ligand-free and ligand-bound forms. I performed compound- and fragment-screening to identify novel ligands for this P450 and then undertook further biophysical characterizations to probe their mode of binding and affinity for CYP126A1. Turnover studies were also performed on the substrate-like compounds to establish the catalytic activity of CYP126A1. I also performed structural experiments using X-ray crystallography to characterize the CYP126A1 structure and to determine the binding modes of different ligands and to identify important residues participating in ligand selectivity of this P450. These results are presented in chapter 4 of this thesis.
References

8. Lerner, B.W. & Lerner, L.K. World of Microbiology and Immunology. 1 326-328 (Gale, Detroit, 2003).


Chapter Two

Structural characterization of CYP144A1 – a cytochrome P450 enzyme expressed from alternative transcripts in \textit{Mycobacterium tuberculosis}

Jude Chenge$^{1\dagger}$, Madeline E. Kavanagh$^{2\dagger}$, Max D. Driscoll$^1$, Kirsty J. McLean$^1$, Douglas B. Young$^3$, Teresa Cortes$^4$, Dijana Matak-Vinkovic$^2$, Colin W. Levy$^1$,

Stephen E. J. Rigby$^1$, David Leys$^1$, Chris Abell$^2$, Andrew W. Munro$^{1\ast}$

$^1$Manchester Institute of Biotechnology, Centre for Synthetic Biology of Fine and Specialty Chemicals (SYNBIOCHEM), Faculty of Life Sciences, The University of Manchester, Manchester M1 7DN, United Kingdom. $^2$Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom. $^3$Centre for Molecular Microbiology and Infection, Imperial College London, London, United Kingdom. $^4$Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.

$\dagger$Jude Chenge and Madeline E. Kavanagh are joint first authors on this article
2.1 Abstract

*Mycobacterium tuberculosis* (*Mtb*) causes the disease tuberculosis (TB). The virulent *Mtb* H37Rv strain encodes 20 cytochrome P450 (CYP) enzymes, many of which are implicated in *Mtb* survival and pathogenicity in the human host. Bioinformatics analysis revealed that CYP144A1 is retained exclusively within the *Mycobacterium* genus, particularly in species causing human and animal disease. Transcriptomic annotation revealed two possible CYP144A1 start codons, leading to expression of (i) a “full-length” 434 amino acid version (CYP144A1-FLV) and (ii) a “truncated” 404 amino acid version (CYP144A1-TRV).

Computational analysis predicted that the extended N-terminal region of CYP144A1-FLV is largely unstructured. CYP144A1 FLV and TRV forms were purified in heme-bound states. Mass spectrometry confirmed production of intact, His$_6$-tagged forms of CYP144A1-FLV and -TRV, with EPR demonstrating cysteine thiolate coordination of heme iron in both cases. Hydrodynamic analysis indicated that both CYP144A1 forms are monomeric. CYP144A1-TRV was crystallized and the first structure of a CYP144 family P450 protein determined. CYP144A1-TRV has an open structure primed for substrate binding, with a large active site cavity. Our data provide the first evidence that *Mtb* produces two different forms of CYP144A1 from alternative transcripts, with CYP144A1-TRV generated from a leaderless transcript lacking a 5’-untranslated region and Shine-Dalgarno ribosome binding site.
2.2 Introduction

*Mycobacterium tuberculosis* (Mtb) is the causative agent for tuberculosis (TB), a chronic, infectious human disease that is responsible for the death of more than 1.5 million people annually. Major efforts have been made in recent years to produce new, effective antibiotics against Mtb. New anti-TB drugs are desperately needed as a result of the emergence of drug-, multidrug- and extensively drug-resistant strains of Mtb, which have arisen as a consequence of factors including poor patient compliance with drug regimens. The “deadly synergy” between Mtb and the human immune deficiency virus (HIV) has also been an important factor contributing to the spread of TB and increase in human morbidity. The best studied form of Mtb is H37Rv, a virulent strain which has 20 genes encoding cytochrome P450 enzymes (P450s or CYPs). P450 monooxygenase enzymes are hemoproteins that typically catalyse the activation of molecular oxygen and the insertion of an atom of oxygen into a substrate bound close to the heme iron. The high proportion of the Mtb genome dedicated to P450s implicates these enzymes in multiple important biochemical functions, and the anticipation that Mtb P450s may be involved in the pathogenicity and survival of the bacterium in the host has led to a focus of drug discovery efforts on Mtb P450s for new anti-TB compounds. Characterization of P450 enzymes from Mtb has provided evidence for their involvement in the metabolism of lipids and sterol molecules, the oxidative modification of respiratory menaquinone and the production of cyclic dipeptide secondary metabolites.

Currently, six of the Mtb P450s have been crystallized and structures of these enzymes have been determined in both ligand-free and various substrate/inhibitor-bound forms. Physiological roles have been identified for CYP142A1 and CYP125A1 as host cholesterol catabolizing enzymes, and also for CYP124A1 in sterol and/or branched chain fatty acid
oxidation\textsuperscript{11}. CYP132A1 was also proposed to be involved in fatty acid metabolism based on its similarity to eukaryotic CYP4 family fatty acid hydroxylases\textsuperscript{19}. CYP51B1 is related to eukaryotic sterol 14α-demethylases (with fungal CYP51 enzymes being important targets for azole class drugs), while CYP121A1 is involved in secondary metabolite synthesis, catalysing the oxidative crosslinking of the aromatic side chains of the cyclic dipeptide cyclo-L-Tyr-L-Tyr (cYY) to form the product mycocyclosin. The \textit{CYP121A1} gene was reported to be essential for \textit{Mtb} viability, although the function of mycocyclosin remains unclear\textsuperscript{13,15,20}.

In view of the important roles demonstrated for the aforementioned P450 enzymes in \textit{Mtb} viability and pathogenicity, it is important to investigate the physiological roles of the remaining P450s and to explore their potential involvement in bacterial physiology and survival in the host.

The identification of the first prokaryotic sterol demethylase (CYP51B1) in \textit{Mtb} led to the proposal that azole antifungal drugs, which are effective, clinically-used inhibitors of fungal sterol demethylases, could be used as novel drugs for treating TB. The crystal structure of CYP51B1 was determined in complex with the antifungal drug fluconazole, revealing direct ligation of the P450 heme iron by a nitrogen atom from one of the inhibitor’s triazole rings. Although fluconazole is not very effective against \textit{Mtb}\textsuperscript{15}, the azole antifungals econazole and clotrimazole were found to be active against both latent and multidrug-resistant strains of \textit{Mtb} in murine model systems\textsuperscript{8,21}. However, the broad spectrum activity of these azole compounds has limited their application as systemic therapeutics. In addition to the potential of azole drugs as novel therapeutics against \textit{Mtb}, they are also useful tools for probing the structure and function of the \textit{Mtb} P450s\textsuperscript{8}. Recently, \textit{Mtb} CYP144A1, encoded by the H37Rv gene \textit{Rv1777}\textsuperscript{22,23}, was identified to bind tightly to several azole antifungals, raising the possibility that this P450 could be a target for novel anti-TB drug development.
In this paper, structural, biochemical, transcriptomic and bioinformatics data are presented to provide a detailed characterization of CYP144A1 and its potential as a *Mtb* drug target. The evolutionary ancestry of CYP144A1 is explored, and its conservation across the *Mycobacterium* genus is consistent with its importance to *Mtb*. Data from transcriptome annotation are reported, which reveal that alternative transcripts of CYP144A1 are produced, leading to the production of different forms of the P450 protein. The expression and purification of these different forms of CYP144A1 are described, as is the comparative analysis of their spectroscopic and ligand binding properties. Furthermore, the successful crystallization of the shorter form of CYP144A1 (CYP144A1-TRV) and the first X-ray crystal structure of a CYP144A1 protein are reported, enabling novel insights into this enzyme, its active site organization and its relationship to other *Mtb* P450 enzymes.

### 2.3 Results and Discussion

#### 2.3.1 Bioinformatics and evolutionary studies of CYP144A1

The *Mtb* gene *Rv1777* (*CYP144A1*) from the pathogenic strain H37Rv (which encodes the cytochrome P450 enzyme CYP144A1) is located in a region of the *Mtb* genome rich in genes encoding proteins of unknown function. Thus, the prediction of the physiological function of CYP144A1 has been difficult from its genetic context alone. The nearby gene *Rv1771* encodes a L-gulono-1,4-lactone dehydrogenase shown to catalyse the final step required in L-ascorbic acid biosynthesis, while the *Rv1781c* gene is predicted to encode a glucanotransferase enzyme (*MalQ* or amylomaltase). However, ligand-binding experiments with CYP144A1 using various sugars did not result in P450 heme spectral shifts that would be consistent with substrate-like binding (data not shown). The *eccB5* and *eccC5* genes (*Rv1782* and *Rv1783*) encode predicted components of the ESX-5 protein export system, related to the better known ESX-1 virulence factor secretion system that exports the
proteins ESX-1 (ESAT-6) and ESX-2 (CFP-10) into the host\textsuperscript{27}. A little further downstream of \textit{CYP144A1} are the \textit{Rv1785c} and \textit{Rv1786} genes, which encode the cytochrome P450 CYP143A1 and a likely 3Fe-4S ferredoxin redox partner for the P450. A BLAST search using the \textit{Mtb} CYP144A1 protein sequence revealed that CYP144 P450s are highly conserved within \textit{Mtb} strains (~99-100\% identity), and also amongst the closely related \textit{Mtb} complex (\textit{MTBC}) members, which includes \textit{M. bovis} (99\%) and \textit{M. canettii} (99\%). Apparent orthologues of CYP144A1 occur in other pathogenic mycobacterial species, such as in \textit{M. marinum} (80\%) and \textit{M. ulcerans} (79\%), though typically with lower sequence identity to CYP144A1. The majority of species containing CYP144A1 orthologues are directly associated with a human disease, including \textit{M. colombiense} (64\% identity to \textit{Mtb} CYP144A1), a member of the MAC (\textit{Mycobacterium avium} complex) which can infect HIV patients with low CD4 cell counts\textsuperscript{28-31}. CYP144A1 orthologues are also encoded by other mycobacteria isolated from patients with compromised immune systems (e.g. \textit{M. simiae}, 76\%; \textit{M. triplex}, 80\%; and \textit{M. lentiflavum}, 76\%). Among other CYP144A1-related sequences are CYP144A4 from the amphibian pathogen \textit{M. liflandii} (77\%), CYP144 orthologues from \textit{M. asiaticum} (73\%) which infects primates, and from the bacteria \textit{M. smegmatis} (66\%) and \textit{M. gastri} (82\%) which are typically considered to be non-pathogenic (\textbf{Figure 2.1}).
Phylogenetic tree illustrating the evolutionary relationship of CYP144A1 to other mycobacterial CYP144 family P450s.

Protein sequence alignment of CYP144A1 was made with CYP144 sequences from other mycobacteria. Protein sequence conservation between the CYP144 proteins in all species ranges from 58-100%, with higher conservation of CYP144A1 sequence identity within the *M. tuberculosis* complex (MTBC), and the lowest conservation in the *M. avium* complex (MAC). The 0.05 reference scale marks a 5% estimated sequence variance. The figure was generated using MEGA6.

No orthologues of CYP144 were observed outside the *Mycobacterium* genus to date. The apparent exclusive conservation of CYP144 P450 family enzymes in the mycobacteria suggests the retention of an important catalytic role. In previous studies, the *CYP144A1* gene was deleted from the *M. tuberculosis* H37Rv genome to investigate its effect on the growth and viability. It was found that the *M. tuberculosis* CYP144A1 deletion strain was viable and grew *in vitro*, but that the growth rate of the deletion strain was substantially lower than that of the wild-type *M. tuberculosis* H37Rv. In addition, the *CYP144A1* deletion strain was markedly more susceptible to growth inhibition by azole drugs than the wild-type *M. tuberculosis* H37Rv. These data point to important function(s) for CYP144A1 in *M. tuberculosis*, and possibly to its participation in the *M. tuberculosis* stress response.
mechanisms. Consistent with this conclusion is the finding that \textit{CYP144A1} is among the mycobacterial genes induced following bacterial growth arrest with vancomycin\textsuperscript{32}. Vancomycin inhibits \textit{Mtb} cell wall synthesis, and thus a potential function for this P450 may be related to the stress response to cell wall damage.

\subsection*{2.3.2 Identification of a truncated form of the CYP144A1 P450 from transcriptome analysis}

Analysis of the transcriptome from \textit{Mtb} H37R\textit{v} identified a weak transcription start site (TSS) at genome position 2010633, adjacent to a -10 consensus motif (TATTCT, 2010622-2010627) and 7 bp upstream from the annotated CYP144A1 Val1 start codon in Tuberculist\textsuperscript{23}. There is near complete amino acid sequence identity between the predicted 434 amino acid residue \textit{Mtb} H37R\textit{v} protein and the orthologues found in \textit{Mtb} CDC1551, \textit{M. canetti} CIPT, \textit{Mtb} T46, and \textit{M. bovis} AF2122/97. In previous studies, we expressed and purified this 434 amino acid full-length form of CYP144A1 (referred to hereafter as CYP144A1-FLV), demonstrating that it bound heme and displayed spectroscopic and ligand-binding properties consistent with a P450\textsuperscript{22}. However, the best protein sequence alignments of \textit{Mtb} CYP144A1 with many other mycobacterial CYP144 family P450 sequences are found using a sequence initiating at an internal residue around Met31. This highlighted the possibility that an alternative, shorter form of the \textit{Mtb} CYP144A1 protein might be produced. Consistent with this model, \textit{Mtb} H37R\textit{v} transcriptomics analysis identified a second, stronger TSS at position 2010745. The location of the second TSS would be consistent with the generation of a leaderless transcript with Met31 as the start codon. Interestingly, parallel studies of \textit{CYP144} transcripts from the \textit{Mtb} N145 strains (a Beijing isolate similar to \textit{Mtb} HN878) suggest that this organism favours the upstream TSS (\textbf{Figure 2.2}).
Figure 2.2. Alternative transcripts of the *CYP144A1* gene produced by *Mtb* H37Rv.

The figure shows that two alternative transcriptional start sites (TSS) are located at nucleotides 201063 (CYP144A1-FLV) and 2010745 (CYP144A1-TRV) in the *Mtb* H37Rv genome. These are shown as blue lines in the figure. A higher level of transcription occurs for the CYP144A1-TRV version of the gene. TSS are demarcated by black rectangles and show detailed views of the regions around the TSS for the two forms of the gene, identifying the relevant regulatory regions and the start codons (Val1 and Met31, respectively) for CYP144A1-FLV (VRRSPK…) and CYP144A1-TRV (…MTIAKD…).
Leaderless transcripts are common in *Mtb* and are characterized by the absence of a 5’-untranslated region and a Shine-Dalgarno ribosome binding site. They are considered to comprise up to 25% of the transcripts in *Mtb*, suggesting that “truncated” versions of the numerous proteins *Mtb* encodes might play crucial roles in mycobacterial physiology. TSS mapping thus identified both “full-length” and “truncated” versions of CYP144A1 as potential translation products from alternative transcripts. In light of these novel data, we generated plasmid expression constructs encoding both the full-length (CYP144A1-FLV) 434 amino acid protein and the 404 amino acid truncated version (CYP144A1-TRV) forms of CYP144A1. The expression and purification of the two forms of the CYP144A1 protein enabled comparative studies of their properties, and facilitated the crystallization of the proteins in order to obtain the first structural data for *Mtb* CYP144A1.

### 2.3.3 Expression and purification of CYP144A1 in its truncated and full-length forms

The N-terminal truncated version construct of CYP144A1 (CYP144A1-TRV) was generated using the previously prepared full-length version CYP144A1-FLV/pET15b plasmid vector, and by deletion of the region encoding the first 30 amino acids. The CYP144A1-TRV construct initiated from the first methionine codon (Met31) of the *CYP144A1* gene, corresponding to the second of the two CYP144A1 TSS detected (Figure 2.3A). CYP144A1-FLV and TRV were both expressed and purified as His6-tagged constructs in *E. coli* as described in the Materials and Methods. Both the CYP144A1-TRV and CYP144A1-FLV forms were readily purified using nickel affinity, anion exchange and size exclusion column chromatography steps. Both forms of the enzyme were soluble, indicating that the enzyme is almost certainly cytosolic in *Mtb*. This is characteristic for the majority of the *Mtb* P450s.
characterized to date, including the cholesterol oxidizing CYP125A1 and CYP142A1 enzymes, the branched chain fatty acid hydroxylase CYP124A1 and the cyclodipeptide oxidase CYP121A1<sup>7,11,13,16-18</sup>. A notable exception is CYP128A1, a likely dihydromenaquinone hydroxylase, which may be membrane associated in order to enable access to its lipophilic dihydromenaquinone substrate(s) that are retained in the <i>Mtb</i> cell membrane<sup>14</sup>. The CYP144A1-FLV and CYP144A1-TRV proteins were purified and shown to have approximate masses of 49 kDa and 46 kDa, respectively, as observed by SDS-PAGE analysis (<strong>Figure 2.3B</strong>). The different masses of the two CYP144A1 forms were confirmed using electrospray ionization mass spectrometry, with a mass of 49251 Da determined for CYP144-FLV and 46008 Da for CYP144-TRV (<strong>Figures 2.3C and 2.3D</strong>). These values are consistent with the expected masses from the amino acid sequences of the respective forms of the His<sub>6</sub>-tagged CYP144A1 proteins, minus the initial methionine residue (CYP144A1 gene and protein sequences are shown in <strong>Supplementary Figure S2.1</strong>), and are also consistent with our previous studies of CYP144-FLV<sup>22</sup>.
**Figure 2.3** Expression and purification of CYP144A1-FLV and CYP144A1-TRV.

**Panel A** shows the genetic truncation of the intact CYP144A1-FLV (red) to form CYP144A1-FLV – removing the gene segment encoding residues Val1 to Leu30, shown with the CYP144A1-FLV (Val) and -TRV (Met) initiator residues in yellow circles. This produces the CYP144A1-TRV (grey), with 30 amino acids fewer than CYP144A1-FLV (red). **Panel B** shows purification and approximate molecular weight determination of CYP144A1-FLV and -TRV by SDS-PAGE analysis, with markers of indicated mass (kDa) in the first lane. **Panels C and D** show accurate mass spectra of CYP144A1-FLV (49251 Da) and CYP144A1-TRV (46008 Da) obtained by high-resolution mass spectrometry under denaturing conditions using 40 μM protein in both cases.

### 2.3.4 UV-Visible spectroscopic properties of CYP144A1-FLV and CYP144A1-TRV

UV-Visible (UV-Vis) spectroscopy is a valuable method for characterizing P450s and for the quantitative analysis of their ligand-binding properties. The UV-Vis spectral features of the FLV and TRV forms of CYP144A1 were found to be identical for the two proteins in their oxidized and reduced states, and for these proteins bound to the gaseous ligands nitric oxide (NO, in the ferric form) and carbon monoxide (CO, in the ferrous form). As such, only UV-Vis spectral data for the newly constructed CYP144A1-TRV form are presented (**Figure 2.4A**). The consistency in the spectral properties of the two CYP144A1 proteins suggests that the 30 amino acid N-terminal “extension” on CYP144A1-FLV does not influence the environment of the heme prosthetic group or the coordination of its heme iron. The oxidized form of CYP144A1-TRV has a Soret peak at 420 nm, which is slightly red-shifted in comparison to the resting ferric states of other characterized *Mtb* P450s, such as CYP121A1 at 416.5 nm\(^{13}\), CYP51B1 at 419 nm\(^{15}\), and CYP130A1 at 418 nm\(^{35}\).

The CYP144A1-TRV resting state Soret peak of 420 nm is consistent with that reported previously by Driscoll *et al.*\(^{22}\) for CYP144A1-FLV (420.5 nm). Both the FLV and TRV forms of the CYP144A1 enzyme were dialysed extensively (~12 hours) after the nickel-IDA affinity chromatography step of protein purification to ensure the removal of any residual imidazole (used to elute the P450s from the nickel column) that remained bound to the heme.
iron. The affinity for imidazole is weak (with a $K_d$ of 3.0 mM for the CYP144A1-FLV form$^{22}$) and so this precautionary step should effect near-complete removal of the ligand. Thus, the 420 nm Soret peak for CYP144A1-TRV almost certainly reflects accurately the properties of the P450 protein in its ferric, low-spin (LS) state, with water as the 6th ligand to the heme iron. The ferric NO-bound CYP144A1-TRV complex showed distinct absorbance properties, with the Soret peak at 437 nm and major spectral changes in the alpha/beta band region with maxima at 576 nm and 544 nm. These features are consistent with those of the NO complexes of the aforementioned *Mtb* P450s CYP51B1, CYP130A1 and CYP121A1$^9$,$^{13}$,$^{15}$. The ferrous CO-bound CYP144A1-TRV produced a Soret peak at 420 nm, and thus forms almost completely the cysteine thiol-coordinated (P420) species. Only a minor peak for the cysteine thiolate-coordinated P450 species was detected at 448 nm. The consistency in the UV-Vis spectral properties of CYP144A1-FLV and -TRV in their aqua-ligated ferric, ferric-NO and ferrous-CO complexes provides further evidence that the 30 amino acid N-terminal truncation made to generate the CYP144A1-TRV protein does not alter its heme environment significantly$^{22}$.

### 2.3.5 EPR spectroscopic analysis of the CYP144A1-FLV and CYP144A1-TRV proteins

Electron paramagnetic resonance (EPR) X-band spectra were recorded for the oxidized (ferric) CYP144A1-FLV and -TRV proteins to provide further insights into the heme ligation environment and spin-state equilibrium in the two forms of CYP144A1. Spectra for both the FLV and TRV CYP144A1 proteins were similar, with g-values of $g_x = 2.41$, $g_y = 2.24$, and $g_z = 1.92$ (2.41/2.24/1.92) in both cases (Figure 2.4B). These data are indicative of a predominantly low-spin ferric and cysteinate-coordinated heme iron. No detectable signals for a ferric high-spin species were detected. The low-spin state of both forms of CYP144A1
is consistent with previous work on the CYP144A1-FLV P450 (g-values of 2.42/2.25/1.93), and these g-values are also similar to those for other *Mtb* P450s in their low-spin states, including the cholesterol hydroxylase CYP142A1 and the cyclodipeptide oxidase CYP121A1. The EPR spectrum of the CYP144A1-TRV protein bound to the azole drug econazole also shows the complete retention of a low-spin state of the enzyme, but with altered g-values of 2.43/2.25/1.90 (Figure 2.4B). Econazole binds tightly to the CYP144A1-FLV protein ($K_d = 0.78 \pm 0.29 \mu M$), and an optical titration of CYP144A1-TRV with econazole also revealed a similar $K_d$ value ($0.72 \pm 0.18 \mu M$). Econazole binding to CYP144A1-TRV occurs with a heme Soret red shift to 423 nm, which is again similar to that seen for CYP144A1-FLV (Soret $A_{max}$ of the CYP144A1-FLV/econazole complex is at 424 nm). The EPR g-values for the CYP144A1-FLV/econazole complex are at 2.45/2.26/1.89. These data are further confirmatory that the CYP144A1-TRV truncation does not influence the P450 heme environment significantly by comparison with the CYP144A1-FLV enzyme (Figure 2.4B).

**Figure 2.4** Spectroscopic properties of CYP144A1–TRV.

**Panel A.** UV-visible absorbance spectra of ligand-free CYP144A1-TRV (6 μM, black line), nitric oxide-bound (red line) and reduced/carbon monoxide-bound form (blue line). The Soret spectral maxima are at 420 nm, 437 nm and 420 nm, respectively. **Panel B.** X-band EPR spectra of ligand-free and econazole (400 μM)-bound CYP144A1-TRV (200 μM), with the g-values labelled.
In the econazole complexes of the CYP144A1 enzyme forms, it might be expected that the $g_z$ (and $g_x$) values would be more substantially changed from those in the resting, aqua-coordinated state if there was direct coordination of the heme iron by an econazole imidazole nitrogen (e.g. to $g_z \sim 2.50$ or beyond). Indeed, in previous studies of the CYP144A1-FLV protein a minor $g_z$ feature at 2.62 was observed that might indicate direct coordination of the heme iron by an econazole nitrogen in a small proportion of this P450 form\(^{22}\). However, the relatively small changes observed in the EPR spectra for the CYP144A1-TRV/FLV proteins may indicate that econazole ligates the heme iron indirectly. This could occur via a retained 6\(^{th}\) ligand water molecule in the majority of the P450 molecules for both these proteins. A similar phenomenon was reported for CYP121A1 in complex with fluconazole, and confirmed by X-ray crystallographic data\(^{36}\). In this case, the g-values for the CYP121A1-fluconazole complex are 2.45/2.26/1.90, compared to 2.48/2.25/1.90 in the resting form. A similar conclusion can be drawn from the rather less extensive Soret shift of the CYP144A1-FLV/TRV proteins on binding econazole (420.5/420 nm to 424/423 nm) than is observed typically in other P450s (e.g. from 416.5 nm to 423 nm on the binding of econazole to CYP121A1)\(^{20,37}\).

2.3.6 Differential Scanning Calorimetry (DSC) studies on the CYP144A1-FLV and CYP144A1-TRV enzyme forms

DSC was used to analyse the thermal stability of both the CYP144A1-FLV and -TRV proteins, in order to probe for any differences induced by truncation of the enzyme. The thermal unfolding profiles for both enzymes indicated a single unfolding midpoint temperature ($T_m$ value) at 47 °C (Figure 2.5A), confirming that the N-terminal truncation did not alter the thermal stability of the CYP144A1-TRV significantly. Further comparative analyses of the thermal stability of the clotrimazole- and econazole-bound forms of CYP144A1 again yielded highly similar results for the CYP144A1-FLV and -TRV proteins.
The binding of clotrimazole or econazole increased the $T_m$ for both forms of the protein to 50 °C, indicating that the tight-binding azoles stabilize both forms of CYP144A1 to similar extents (Figures 2.5B and 2.5C). The similarity in thermal stability of both CYP144A1 forms in their ligand-free and azole drug-bound states again indicates that the stability of the CYP144A1-TRV protein is not compromised by the deletion of the 30 amino acid N-terminal segment, and suggests that this part of the protein is not integral to the folding or stability of the structural core of the P450.

![Figure 2.5](image)

**Figure 2.5** Differential Scanning Calorimetry (DSC) analysis of CYP144A1-TRV.

Panels A-C show DSC thermograms for ligand-free, econazole-bound (30 μM), and clotrimazole-bound (30 μM) forms of CYP144A1-TRV (8 μM), respectively. The collected data (thick black line) and fitting (dotted line) are shown in each case. The $T_m$ values are at 47 °C, 50 °C and 50 °C, respectively.

### 2.3.7 Crystal structure determination of CYP144A1-TRV

Crystallization of both the FLV and TRV forms of CYP144A1 was achieved, producing crystals with a similar bipyramidal morphology and size (100 μm) in each case. CYP144A1-TRV crystals diffracted to a resolution of 1.55 Å and a ligand-free crystal structure of this form of the CYP144A1 protein was successfully solved using molecular replacement with the *Mtb* cholesterol oxidase enzyme CYP142A1 (PDB 2XKR) as a search model. The CYP144A1-FLV crystals yielded a similar structure (data not shown), with no electron density visible for the additional N-terminal amino acids. This could either reflect that
proteolytic cleavage of the N-terminal region of the CYP144A1-FLV form had occurred, or else be due to the fact that the 30 additional N-terminal residues present in CYP144A1-FLV are largely unstructured. Analysis of the CYP144A1-FLV protein sequence using the protein disorder prediction system (PrDOS)\textsuperscript{39} algorithm indicated a disorder probability nearing 90% for these 30 amino acids (Figure 2.6A).

The CYP144A1-TRV crystal structure contains two monomers in the asymmetric unit. However, a dimeric solution state is unlikely, since solution state MALLS experiments (Figures 2.6B and 2.6C) indicated that both the FLV and TRV forms of CYP144A1 are soluble, monomeric proteins (as are the majority of bacterial P450s). Both monomers in the crystal asymmetric unit have a similar overall conformation, with an rmsd of 0.55 Å for 393 C alpha atoms. The three dimensional structure and secondary structural organization of CYP144A1-TRV resembles those of previously solved \textit{Mtb} P450 structures, and is most similar to P450 CalO2 (PDB 3BUJ), a P450 involved in biosynthesis of the enediyne antitumor antibiotic calicheamicin\textsuperscript{40} (Z score 16.1, rmsd 1.96 Å for 346 C alpha atoms), and to the \textit{Mtb} cholesterol hydroxylase CYP142A1 (PBD code 2YOO) (Z score 15.9, rmsd 1.93 Å for 356 C alpha atoms) (Figures 2.7B and 2.7C)\textsuperscript{7,38}.

The structural elements (F/G-helices and BC-loop region) important in determining active site access and substrate specificity\textsuperscript{41} are those that are most distinct in CYP144-TRV from other P450 structures (Figure 2.7A). The relatively long linker region connecting the B-and C-helices folds into an extended beta-hairpin, that occupies the region between the N-terminal beta-sheet and the FG-loop. This creates a large access channel to the heme cavity (Figure 2.7D), approximately 10 Å deep from the protein surface and 10 Å wide. This suggests the binding of relatively large substrates in CYP144A1.
**Figure 2.6.** Protein disorder prediction and hydrodynamic features of CYP144A1.

**Panel A.** Protein disorder prediction analysis of CYP144A1-FLV (red), CYP144A1-TRV (green) and the *Mtb* cholesterol hydroxylase CYP142A1 (blue). The threshold (black line) measure of the prediction’s false positive rate was set at 5%. The numbers of disordered amino acids in the protein sequence are indicated on the top panel. The figure was generated using PrDOS software. **Panels B and C.** Multiangle Laser Light Scattering (MALLS) analysis of CYP144A1-TRV and CYP144A1-FLV, respectively. The refractive index (RI) data are shown as black lines in both cases, with the blue lines indicating the apparent molar mass of the relevant proteins. The data are consistent with both forms of CYP144A1 being monomeric in solution, and having molar masses of ~46,000 g/mol (CYP144A1-TRV) and ~49,000 g/mol (CYP144A1-FLV).

The active site of CYP144A1-TRV is predominantly hydrophobic, with Phe321 and His324 side chains located in close proximity of the heme 6th ligand. These residues are likely to be important in determining substrate selectivity (Figure 2.7E). I-helix residues Glu277 and Ser268 are probably required for protonation of P450 heme iron-oxo species during catalysis. While CYP121A1 also has a serine residue corresponding to Ser268 in CYP144A1-TRV, other *Mtb* P450s use a threonine. However, there are few other similarities in the amino acid composition of the active site of CYP144A1 compared to those of the other structurally characterized *Mtb* P450s, suggesting a unique role and substrate selectivity profile for CYP144A1. The substantial structural differences between the active site
organization in CYP144A1 and those of the other Mtb P450 isoform structures should enable the design of CYP144A1 isoform-selective inhibitors, a topic addressed in our ongoing fragment screening studies of this enzyme.

**Figure 2.7.** The crystal structure of CYP144A1-TRV.

**Panels A and D.** Ribbon and surface representations, respectively, of the Mtb P450 CYP144A1-TRV. The heme, BC-loop, FG-helices and I-helix are coloured red, blue, dark red and yellow, respectively. **Panel B and C.** Ribbon structure representations of *Mycobacterium smegmatis* CYP142A2 (PDB 2YOO) and the putative orsellinic acid oxidase P450 CalO2 (PDB 3BUJ) using similar colour coding as panel A. The orientation of both structures is similar to that of CYP144A1-TRV in panel A. **Panel E.** Active site structure of CYP144A1-TRV, with key residues colour coded as in panel A. Panels A-E were generated using PyMOL\(^48\).

### 2.4 Conclusions

There is growing recognition that alternative transcriptional start sites and leaderless transcripts play an important role in increasing the diversity of proteins produced from microbial genomes, with around 25% of mycobacterial transcripts being leaderless\(^34\). In the case of the *M. tuberculosis* cytochrome P450 CYP144A1, annotation of the relevant P450 (CYP) gene in the Tuberculist database indicates that the encoded protein has 434 amino
acids with a valine as the first residue (from a GTG codon). Previous studies indicated that this full length form of CYP144A1 (CYP144A1-FLV) could be purified using an *E. coli* expression system. However, a methionine is also located at residue 31 in the CYP144A1-FLV form of the P450, and this was considered as an alternative initiation codon for the protein, particularly in light of the predicted “truncated version” of CYP144A1 (CYP144A1-TRV) being only 404 amino acids in length – a size closer to the norm for a typical prokaryotic P450. Further studies from transcriptomics analysis resulted in the identification of an internal transcriptional start site from a leaderless transcript producing the CYP144A1-TRV form. Both the FLV and TRV forms of the CYP144A1 protein were successfully expressed and purified to homogeneity. Biophysical and biochemical studies revealed similar properties for the two proteins forms, confirming that the 30 N-terminal amino acid residues that were removed in the CYP144A1-TRV form did not affect protein stability, oligomerization state or heme environment. However, bioinformatics studies indicated that the N-terminal 30 amino acid region of CYP144A1-FLV was mainly disordered and that this might explain problems associated with the structural resolution of this region in the crystal structure of the CYP144A1-FLV form of the P450. Consistent with this conclusion, it was found that the CYP144A1-TRV form crystallized readily, allowing the determination of a high resolution (1.55 Å) X-ray structure of the ligand-free CYP144A1-TRV enzyme. This crystal structure shows a P450 with a large and predominantly hydrophobic active site, suggesting that CYP144A1’s natural substrate(s) are bulky hydrophobic molecules. These data provide the basis for biochemical and modelling studies to identify substrates and other ligands for the P450. In this respect, our ongoing work involves fragment based screening studies to identify novel, specific ligands for CYP144A1 that can be developed into useful reagents for use as inhibitors and as mechanistic probes of this *Mtb* P450.
2.5 Methods

2.5.1 Bioinformatics studies

All protein sequence alignments were performed using the Phylogeny web tool and the NCBI BLAST program\textsuperscript{42,43}. The individual protein sequences for the CYP144A1-FLV and CYP144A1-TRV P450s were interrogated on the database in searches for sequence neighbours. The resulting output sequences for the various species were further aligned and saved in CLUSTALW format. The phylogenetic tree was generated using MEGA6 software\textsuperscript{44}. The aligned CLUSTALW format sequences were imported into the MEGA6 phylogeny tool and a phylogenetic tree was generated using the “construct/test maximum likelihood” method. Structural alignments were performed using the ESPript web tool\textsuperscript{45}.

2.5.2 Identification of CYP144A1 transcriptional start sites

Transcriptional start sites (TSSs) for CYP144A1 were mapped at single-base resolution by sequencing of a preparation of RNA from an exponential culture of \textit{M. tuberculosis} H37Rv after enrichment for transcripts with intact 5’ triphosphate ends, as described by Cortes et al\textsuperscript{33}.

2.5.3 Cloning of the CYP144A1-FLV and TRV genes

The pCYP144A1-FLV construct encoding the full-length (434 amino acid) recombinant \textit{Mtb} CYP144A1-FLV enzyme was produced as previously described\textsuperscript{22}. The truncated \textit{Rv}1777 gene encoding CYP144A1-TRV was generated by PCR from a previously cloned construct of the full length \textit{Rv}1777 gene in the pET15b plasmid vector using the forward primer, 5’-CGATCACGCTGAA\underline{CATATG}\underline{GACAATTG}CC-3’ and the reverse primer, 5’-\underline{GGCAAT}TG\underline{TATGTTGACTG}GC-3’ (Merck-Millipore, Watford UK). The underlined letters in both the forward and reverse primers indicates an engineered NdeI
restriction endonuclease site. The bold letters in the forward primer indicate the start codon ATG. The PCR amplification reaction was carried out in a Techne TC-512 thermal cycler (Techne, Cambridge UK) using the proofreading Pfu Turbo DNA polymerase (Agilent, Cheadle UK). The amplification conditions were 95 °C for 2 min, 20 cycles of 95 °C for 30 s, 60 °C for 30 s, and 68 °C for 7 min. The PCR reaction was followed by a final polymerization step of 68 °C for 7 min. The generation of the truncated gene encoding CYP144A1-TRV was achieved by removing the DNA segment between the two NdeI restriction sites through a NdeI endonuclease digest reaction, followed by re-circularization carried out using a Quick Ligation Kit (NEB, Hitchin UK) to produce the pCYP144A1-TRV construct. The resulting CYP144A1-TRV gene construct encodes the CYP144A1-TRV protein with the initial 30 amino acids from the intact CYP144A1-FLV protein deleted.

2.5.4 Expression and purification of the CYP144A1-FLV and CYP144A1-TRV forms of CYP144A1

CYP144A1-FLV and CYP144A1-TRV proteins were produced by transforming the E. coli strain C41 (DE3) (Merck-Millipore, Watford UK) with the pCYP144A1-FLV or pCYP144A1-TRV plasmid constructs. Expression of the relevant gene constructs was done using the isopropyl β-D-thiogalactopyranoside (IPTG)-inducible T7 RNA polymerase/promoter system. This was achieved through IPTG-dependent expression of the T7 RNA polymerase from a chromosomally integrated gene copy in the C41 (DE3) strain, leading to T7 polymerase-dependent transcription of the CYP144A1-FLV and -TRV genes in pET15b. CYP144A1 protein production was typically done in ~15 litre cultures of 2xYT growth medium (ForMedium, Hunstanton UK). The culture medium was distributed between 24 x 2 litre conical flasks. Each flask contained 600 ml of growth medium and ampicillin (50 µg/ml), and was inoculated with 6 ml of transformant cells from an overnight culture in the
same medium under the same conditions. Cells were grown at 37 °C with agitation (200 rpm) until an OD$_{600}$ of 0.5 was reached, and then the growth temperature was decreased to 22 °C and bacterial cell growth was continued to an OD$_{600}$ of 0.7. 100 μM IPTG was then added to induce target gene expression, along with 100 μM delta-aminolevulinic acid (ΔALA) to promote heme synthesis and incorporation into the proteins. The transformant cells were then grown for a further 36 h. The cells were harvested by centrifugation at 6000 g for 10 min at 4 °C using a JLA-8.100 rotor in an Avanti J-26 XP centrifuge The supernatant was discarded and cell pellets were resuspended in ~300 ml of ice cold 50 mM potassium phosphate (KPi, pH 8.0) containing 250 mM NaCl and 10% glycerol. The protease inhibitors phenylmethanesulfonyl fluoride (PMSF, 1 mM), benzamidine hydrochloride (1 mM) and six cOmplete EDTA-free tablets (Roche Diagnostics Ltd, West Sussex UK) were added to inhibit protease enzymes. The cells were lysed on ice by ultrasonication (Bandelin Sonopuls sonicator) with 6 cycles of 30 s on and 60 s rest periods. The cell lysate was centrifuged at 40,000 g for 45 min at 4 °C and the supernatant collected.

The supernatant was loaded onto a Ni-IDA column (Generon, Maidenhead UK) pre-equilibrated with 50 mM KPi (pH 8.0) loading buffer containing 250 mM NaCl and 10% glycerol, using a peristaltic pump (GE Healthcare, Little Chalfont UK). The column was washed with ~80 ml of loading buffer and the flow-through discarded. Proteins were eluted from the column by washing consecutively with increasing concentrations of imidazole [10 mM (250 ml), 80 mM (150 ml) and 160 mM (100 ml) in the loading buffer]. Each eluted fraction was analysed spectrally (250-800 nm), and by SDS-PAGE. Fractions containing relatively pure (mainly the 80 mM and 160 mM) CYP144A1 samples were pooled and concentrated to ~100 ml using ultrafiltration with Amicon concentrators (Merck-Millipore) at 4 °C. The concentrated protein was further dialysed into 50 mM Tris HCl (pH 7.2, dialysis
buffer) containing 50 mM KCl and 1 mM EDTA to remove excess imidazole. The dialysed protein was then loaded onto a Q-Sepharose column (10 cm x 4 cm) pre-equilibrated with the dialysis buffer, and then washed and eluted with linear gradient of KCl (50-500 mM) in the dialysis buffer using an automated AKTA purification system (GE Healthcare). Fractions were analysed both spectrally and by SDS-PAGE as before. Samples with high A_{420}/A_{280} (or Reinheitszahl, Rz) ratios (≥1) were pooled and concentrated to ~200 μl by ultrafiltration using a Centriprep 30 concentrator (Merck-Millipore) at 1500 g. The concentrated protein was further dialysed into 10 mM Tris HCl (pH 7.5) containing 150 mM NaCl. The protein was then subjected to a final purification step via Sephacryl S-200 size exclusion chromatography column using an AKTA purification system. Fractions were again analysed both spectrally and by SDS-PAGE. Fractions with Rz values of ≥1.5 were pooled and concentrated as before, and dialysed into 50 mM Tris.HCl, pH 7.5 containing 50 mM NaCl and 20% glycerol, and stored at -80 °C until use. Purity of the CYP144A1 proteins was assessed by both SDS-PAGE and UV-Vis spectroscopy, with SDS-PAGE samples migrating as a single band at the appropriate molecular weight on SDS-PAGE gels and having an Rz ratio of ≥1.5 being considered as pure.

2.5.5 Mass spectrometry

Protein solutions (40 μM) were prepared by dilution of purified proteins (500-1000 μM) in 200 μM ammonium acetate buffer, pH 7.0. Liquid chromatography-mass spectrometry (LC–MS) was performed on a Xevo G2-S QTof UPLC instrument (Waters, Elstree UK) coupled to an Acquity UPLC system. Samples were eluted through an Acquity UPLC BEH300 C4 column (1.7 μm, 2.1 × 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⁻¹ 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 hr\(^{-1}\). 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.

### 2.5.6 Determination of molecular weight and aggregation state of CYP144A1-FLV and CYP144A1-TRV by multi-angle laser light scattering (MALLS)

An estimate of the molecular weight of both CYP144A1-FLV and CYP144A1-TRV proteins, as well as analysis of their aggregation states, was obtained using a MALLS detector (Wyatt DAWN Eos, Haverhill UK). This was immediately preceded by passing the sample through a Superdex 200 gel filtration step integrated into the detector (24 ml S200 10/300 GL, GE Healthcare). The column was run at a flow rate of ~0.8 ml/min, using 100 μl samples of 2 mg/ml CYP144A1 enzymes in 10 mM Tris.HCl buffer (pH 7.5) containing 150 mM NaCl. An Optilab rEX and Quasi Elastic Light Scattering (QELS) apparatus (both from Wyatt) were used to obtain the refractive index and the hydrodynamic radius of CYP144A1 enzymes, respectively. A K5 cell type and a laser wavelength of 690 nm were used to collect data at 1 s intervals.

### 2.5.7 EPR spectroscopic analysis of CYP144A1

Continuous wave X-band electron paramagnetic resonance (EPR) spectra of CYP144A1 proteins were obtained at 10 K using a Bruker ELEXSYS E500 EPR spectrometer equipped with an ER 4122SHQ Super High Q cavity. Temperature control was effected using an Oxford Instruments ESR900 cryostat connected to an ITC 503 temperature controller. Microwave power was 0.5 mW, modulation frequency was 100 KHz and the modulation amplitude was 5 G. EPR spectra were collected for the CYP144A1-FLV and CYP144A1-TRV proteins in the ligand-free state (200 μM) and at the same protein concentration following the addition of the azole inhibitor drug econazole (400 μM).
2.5.8 Determination of CYP144A1 thermal stability by Differential Scanning Calorimetry (DSC)

DSC Experiments were performed using a Microcal VP-DSC instrument (Malvern Instruments, Malvern UK). Parameters used for running samples were 20-80 °C temperature gradient, 10 min prescan thermostat and a 90 °C/h scan rate. Baseline scans were performed using assay buffer (10 mM KPi, 100 mM NaCl, pH 7.0) and proteins were near-saturated with the azole ligand clotrimazole and econazole prior to DSC analysis of ligand-bound samples. All samples were run using 8 μM protein and azole ligands were used at concentrations at least 20x those of the protein and according to the affinity for the azoles. Data analysis was performed using Origin Software (OriginLab, Northampton MA).

2.5.9 CYP144A1 crystallization and structure determination

Crystallization was performed using the sitting drop method using 20 mg/ml CYP144A1. Drops were prepared by the addition of 0.2 μl of CYP144A1-FLV and -TRV proteins to 0.2 μl of mother liquor, and by incubating at 4 °C. Following initial crystallogenesis using commercial screens, crystallization conditions were further refined using the sitting drop vapour diffusion technique to 0.8 or 1.0 M (NH₄)₂SO₄ with 0.1 M HEPES, pH 7.55, and 25% PEG 3350. Single crystals were flash cooled after addition of 10% PEG 200 as cryoprotectant, and data were collected at Diamond synchrotron beamline IO3 (Harwell, UK). The CYP144A1-TRV structure (PDB 5HDI) was solved by molecular replacement with the M. tuberculosis CYP142A1 structure (PDB 2XKR) as the template. Following automatic model building and refinement using Buccaneer from the CCP4 suite, the model was completed using iterative rounds of refinement using Refmac5 intercalated with manual model building using COOT. Data collection and refinement parameters for the CYP144-TRV crystal structure are presented in Table 2.1.
Table 2.1. Data reduction and final structural refinement statistics for *Mycobacterium tuberculosis* CYP144A1.

Data were collected using beamline IO3 at the Diamond synchrotron (Harwell, UK).

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray Source</td>
<td>Diamond, IO3</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P2,2,2,1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>58.30, 117.78, 122.09</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>2</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{merge}$ (%)</td>
<td>9.7 (80.7)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>10.4 (2.3)</td>
</tr>
<tr>
<td>$CC (I/2)$</td>
<td>99.9 (67.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.7 (97.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.4 (4.4)</td>
</tr>
<tr>
<td>Wilson B factor (Å$^2$)</td>
<td>15.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>88.58-1.54 (1.58-1.54)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>116903 (8401)</td>
</tr>
<tr>
<td>$R_{work}/R_{free}$</td>
<td>13.76/19.24 (21.4/28.3)</td>
</tr>
<tr>
<td>No. non-hydrogen atoms</td>
<td>6850</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean B factor (Å$^2$)</td>
<td>22.5</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.019</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.872</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Favourable regions (%)</td>
<td>96.7</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>2.9</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.4</td>
</tr>
</tbody>
</table>
References


3 Chapter Three

Biophysical characterization and ligand-binding profile of *Mycobacterium tuberculosis* CYP144A1: insight into protein function and inhibitor design

Madeline E. Kavanagh, a† Jude Chenge, b† Kirsty J. McLean, b Max D. Driscoll, b Andrew W. Munro b and Chris Abell a

a Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom; b Centre for Synthetic Biology of Fine and Specialty Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, Manchester M1 7DN, United Kingdom

†These authors are equal first authors on this article

Keywords: cytochrome P450 enzyme, CYP144A1, fragment-based, ligand profile, *Mycobacterium tuberculosis*, proteome annotation
3.1 Abstract

The ligand-binding profile of a protein can provide insight into substrate and function, as well as guidance for rational drug design. Chemoproteomic profiling has been aided in recent years by the rise of fragment-based methodologies. The ligand-binding profile of the orphan enzyme CYP144A1 from Mycobacterium tuberculosis H37Rv was elucidated using a combination of small-molecule chemical probes and fragments. A series of novel ligands and fragments covering a wide range of affinities ($K_D = 82$ - $3000 \mu M$) were identified and shown to exhibit different substrate-type, inhibitor-type and non-heme interacting modes of binding. Biophysical techniques including UV-Vis, NMR, and EPR spectroscopy, and nanoESI native mass spectrometry, were used to characterize the interaction of CYP144A1 with these ligands and to reveal important information about the properties of the enzyme as a potential drug target. These studies demonstrate the ability for CYP144A1 to simultaneously bind to multiple large (>300 Da) ligands, highlight the enzyme’s affinity for specific heme binding pharmacophores and characterize the structure-activity relationships that contribute to ligand binding affinity. Detailed information about the ligand profile of CYP144A1, in combination with the recently solved X-ray crystal structure of the ligand-free enzyme, and comparative structural and bioinformatics analysis with other P450s, provides insight into the potential substrate and function of the enzyme, as well as guidance for the design of novel CYP144A1 inhibitors.
3.2 Introduction

Advances in genome sequencing technologies over the past 20 years have produced a dramatic increase in the number of unique proteins that have been identified. Determining the biological substrate and function of these proteins presents a considerable challenge and a significant proportion of the proteome remains unannotated. For example, as of 2011, 25% of coding sequences in the genome of the pathogen Mycobacterium tuberculosis (Mtb) had no functional information available and were assigned as unknown or conserved hypothetical proteins. Typically, either the genetic context of the corresponding gene, the nucleotide or amino acid sequence similarity, or the subcellular localization of a protein have been used to guide functional annotation. However, these methods are frequently unsuccessful for prokaryotic proteins or for those that share low similarity (< 25-35% sequence identity) to proteins of known function. Furthermore, evidence that numerous functionally distinct proteins can be encoded from a single gene and that even proteins with high sequence similarity can have diverse biological roles, has meant that the development of new computational, chemical and biological techniques to identify protein function is a priority. A notable example of one such technique is chemotype or ligand profiling.

A chemotype profile describes the set of ligands that bind to a protein from a given screening library. Analysis of the size, chemical structure, and functional groups represented by these ligands, in addition to the orientation and location in which they bind to the target, are characteristic for individual proteins. Proteins that share similar substrates or functional roles tend to bind similar classes of ligands, and the clustering of orphan proteins with similar properties to those previously characterized has been used for functional annotation.
Chemotype profiling has been aided in recent years by the rise of fragment-based methodologies. Fragment-based lead discovery (FBLD) is now an established technique in both academic and industrial sectors. However, the utility of fragments as tools for interrogating biochemical systems is only beginning to be appreciated. The advantage of using fragments for ligand profiling arises from their low structural complexity, which allows them to probe macromolecular surfaces more efficiently than larger compounds. This can result in the identification of novel binding sites and has been used to assess the druggability of protein targets. Various in silico tools have been developed which map the clustering of solvent-sized fragments across a protein surface in order to identify “binding hotspots”, which, when exploited, can provide significant improvements in ligand potency. The small size of fragments mean that they inherently cover a greater proportion of the available chemical space using fewer scaffolds than large molecules, which allows screening libraries to be smaller. Finally, the various biophysical techniques used to detect fragment-macromolecule binding interactions typically do not require prior knowledge of protein structure or function, or the availability of other small molecule tool compounds. These attributes commended fragment-based ligand profiling to us when seeking to investigate the orphan cytochrome P450 enzyme (P450) CYP144A1 from \textit{Mtb}.

\textit{Mtb} is the bacterium that causes tuberculosis (TB), a disease that is responsible for the death of more than 1.5 million people, and an estimated annual cost of $12 billion to the global economy. The emergence of antibiotic resistant strains of \textit{Mtb} and the incompatibility of current TB treatments with co-morbid diseases such as HIV/AIDS means that new drugs are urgently required to prevent a global TB epidemic. Sequencing of the \textit{Mtb} H37Rv genome in 1998 revealed a large number of genes (20) encoding P450 enzymes. The prevalence of these enzymes in \textit{Mtb} suggested that they might have important roles for the pathogenicity
and survival of the bacterium. A number of the \textit{Mtb} P450s that have been characterized are essential for bacterial growth and survival, while others are vital for the establishment of infection within the host macrophage. Others perform important functions such as the catabolism of host cholesterol, the synthesis of cell-wall components or secondary metabolites.

CYP144A1 is an orphan \textit{Mtb} P450 that has emerged as a promising drug candidate for TB. CYP144A1 binds tightly to a series of azole antifungal drugs that have potent antmycobacterial activity against \textit{Mtb} H37Rv, and knockout of the \textit{CYP144A1} (\textit{Rv1777}) gene increases the sensitivity of \textit{Mtb} to killing by these compounds. CYP144A1 was shown to be important for normal bacterial growth, and is postulated to have a role in cell wall synthesis, stress response and drug resistance mechanisms. The expression of the \textit{CYP144A1} gene is up-regulated when \textit{Mtb} is cultured in the oxidative environment of the dendritic cell phagosome, and also in response to treatment with the antibiotic vancomycin, which targets cell wall synthesis. However, the true endogenous function and substrate of the enzyme remain unknown. Traditional methods of proteome annotation have been hampered by a lack of information regarding the proteins encoded by genes immediately surrounding \textit{CYP144A1} and also by the low sequence similarity of the CYP144A1 protein (< 33% sequence identity) to other P450s of known function. Consequently, chemoproteomic profiling, using a combination of small molecule tool compounds and fragment screening, was proposed as a means to characterize the enzyme and potentially to identify lead compounds for drug development.

In this study, a series of novel compounds and fragments that bind to CYP144A1 are identified. The preferred heme binding pharmacophore of CYP144A1 is characterized by
screening a focused library of heme binding fragments and analyzing the affinity of various different *Mtb* P450s for compounds containing specific metal binding functional groups. These studies provide insight into structural features of ligands and the P450 active site that effect heme coordination. The SAR of ligands that bind to the distal part of the CYP144A1 active site have been characterized using ligand-observed NMR experiments and competition UV-Vis assays. Displacement NMR and UV-Vis experiments, and nanoESI native mass spectrometry are used to demonstrate the cooperative and competitive binding modes of the ligands, and provide evidence that multiple ligands can bind simultaneously to the CYP144A1 active site. EPR and DSC experiments demonstrate the stabilizing effect of ligands on the enzyme, and provide evidence for the location of ligand binding interactions. This information, in combination with the recently elucidated X-ray crystal structure of the ligand-free CYP144A1 protein, is used to rationalize the experimentally determined binding mode of ligands. The CYP144A1 ligand profile developed in this study provides insight into the potential role of CYP144A1, the development of hypotheses about the enzymes potential substrate(s) and provides guidance for the development of novel CYP144A1 inhibitors.
3.3 Results

3.3.1 Bioinformatic and structural analysis

A BLAST search of non-redundant protein sequences in the UniProtKB/SwissProt and PDB databases revealed a maximum 33% amino acid sequence identity between CYP144A1 and other P450 enzymes of known function.38,39 The proteins with the highest amino acid identity scores (31-33%) were the macrolide oxidases CYP107L1 from *Streptomyces venezuelae*, MycG from *Micromonospora griseorubida* and CYP113A1 (EryK) from *Saccharopolyspora erythraea*, which are required for the biosynthesis of the antibiotics pikromycin, mycinamicin and erythromycin respectively.40–42 Amino acid sequence identities of less than 35% are generally considered insufficient to infer functional information between proteins.9 However, the consistency in the biological role of the P450s identified, in addition to evidence for the upregulation of *CYP144A1* expression in response to TB treatment with the macrolide antibiotic vancomycin,36 suggests that comparative sequence analysis might provide insight into the ligand preference and substrates of CYP144A1.

The recently solved X-ray crystal structure of CYP144A1 (PDB 5HDI)35 was overlaid with the X-ray crystal structures of the *Mtb* P450s CYP121A1 (PDB 3G5F and 1N40), CYP125A1 (PDB 3IVY) and CYP142A1 (PDB 2XKR) (Figure 1a).44 The CYP121A1 and CYP125A1 structures represented the P450s in the water-ligated, ferric resting state, while CYP142A1 was crystallized with tetraethylene glycol bound to the heme cofactor. The overall secondary structure of CYP144A1 was similar to that of the other P450s, with root-mean-squared-deviations (RMSD) of 1.88-2.22 Å calculated for all comparisons (PDBeFOLD, European Molecular Biology Laboratory ([EMBL]-EBI)).45 The greatest secondary structural similarity was found to CYP142A1, which was used as a template when solving the CYP144A1 X-ray crystal structure. Pairwise alignment of residues using
PDBeFold indicated that structural differences between isoforms were predominantly located in the β2-sheets and at the termini of the F- and G-helices, regions known to be important in substrate specificity. In addition, the B-helix, which is prominent in CYP121A1 and CYP125A1, is notably absent in CYP144A1. The lack of distinct secondary structure in this region of CYP144A1, and replacement of the B-helix with a long loop region, gives rise to a more “open” active site. This structure could suggest broader substrate tolerance, or indicate conformational flexibility to accommodate large substrates. Despite having closer sequence similarity to CYP142A1, a comparison of the active sites and substrate entry channels of the four P450s indicated that greater parallels could be drawn between CYP144A1 and CYP121A1 than with the cholesterol oxidases CYP125A1 and CYP142A1 (Figure 3.1B,C).

SiteMap was used to calculate the properties of the P450 active sites within the vicinity (specifically, a 6 Å radius) of the heme cofactor (Figure 3.1). The volume and surface area of the CYP144A1 and CYP121A1 active sites were significantly larger than those of CYP125A1 and CYP142A1, and they both had wide, shallow cavities forming the substrate entry channel. In contrast, the substrate channels of CYP125A1 and CYP142A1 are long and narrow, which is reflective of their sterol (cholesterol and 4-cholesten-3-one) substrates. Despite the larger active site volume of CYP144A1, the surface area available for metal binding interactions was calculated to be significantly smaller (5.1 Å²) compared to the other P450s (9.5-16.1 Å²), which is expected to influence the specificity of the enzyme for particular metal binding pharmacophores. All P450s were predicted by SiteMap to be druggable (D-score > 0.85), with the large active site of CYP144A1 contributing to a high D-score of 0.99. Only CYP142A1 had a larger D-score, reflecting the greater hydrophobicity and enclosure of its active site.
In addition to similarity in general active site geometry, CYP121A1 also had the highest similarity to CYP144A1 in terms of the specific amino acid residues lining the active site (Figure 3.1C). Unlike the majority of P450s, which catalyze proton transfer to iron-oxo intermediates in the P450 catalytic cycle using an I-helix threonine residue, CYP144A1 and CYP121A1 both contain a serine residue (Ser246) at this position (Figure 3.1D). The other I-helix residues of CYP144A1 and CYP121A1 are also largely conserved, except for Ala241, which is replaced by glycine in CYP121A1. The CYP144A1 active site is dominated by a bulky phenylalanine residue (Phe289), which projects further out over the heme than does Phe280 of CYP121A1, and by a prominent histidine residue (His292), which is replaced by a leucine (Leu284) in CYP121A1. Other key differences between these P450 isoforms include the much smaller F/G-helix residues Thr175 and Leu194 present in CYP144A1, which are replaced by phenylalanine (Phe168) and tryptophan (Trp82) residues in CYP121A1, and the greater prevalence of hydrophobic residues (Leu386 and Val387) in place of the polar glutamine (Gln385) and arginine (Arg386) residues which are involved in the catalytic mechanism of CYP121A1.27
Figure 3.1. Structural overlay of CYP144A1 with *Mt* CYP121A1, CYP125A1 and CYP142A1, and comparative structural data. (a) Cartoon representation of CYP144A1 (green) overlaid with CYP121A1 (orange), CYP125A1 (pink) and CYP142A1 (blue). Secondary structures that differ between the two isoforms have been annotated and highlighted in yellow; (b) Panel 1: Surface representation of CYP144A1 (green) overlaid with CYP121A1 (orange) (PDB 1N40). Panel 2: Surface representation of CYP125A1 (pink) (PDB 3IVY) overlaid with CYP142A1 (blue) (PDB 2XKR). Heme cofactor (magenta sticks) and axial water ligand (red sphere) are displayed for the ferric P450 resting state; (c) Similarity of amino acid residues within 6 Å of the heme cofactor of CYP144A1 to *Mt* CYP121A1, CYP125A1 and CYP142A1. Similarity scores and alignment were performed using PDBeFOLD software (EMBL-EBI); (d) Panel 1: Amino acid residues within a 6 Å radius of the heme axial water ligand of CYP144A1 (PDB 5HDI) (green). Panel 2: Overlay with CYP121A1 (PDB 3G5F) active site residues that are conserved (orange) or that differ (yellow) from those in CYP144A1. Residues are labelled for CYP144A1 and CYP121A1 in panel 1 and 2 respectively. Heme and secondary structure are shown in grey. The water ligand oxygen and heme iron are shown in red and pink spheres respectively. Table of structural data for P450 isoforms.

<table>
<thead>
<tr>
<th>Property</th>
<th>RMSD /α-C (Å)</th>
<th>AS Vol (Å³)</th>
<th>AS S.A. (Å²)</th>
<th>MB S.A. (Å²)</th>
<th>D-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP144A1</td>
<td>1.95/352</td>
<td>86</td>
<td>298</td>
<td>16.1</td>
<td>1.17</td>
</tr>
<tr>
<td>CYP121A1</td>
<td>2.28/327</td>
<td>181</td>
<td>452</td>
<td>5.1</td>
<td>0.99</td>
</tr>
<tr>
<td>CYP125A1</td>
<td>2.21/335</td>
<td>143</td>
<td>386</td>
<td>12.8</td>
<td>0.86</td>
</tr>
<tr>
<td>CYP142A1</td>
<td></td>
<td>86</td>
<td>298</td>
<td>16.1</td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Ligand screening using UV-Vis spectrophotometry

Ligands that bind in close enough proximity to the heme cofactor of a P450 to perturb the coordination sphere can be detected from the change they cause in the optical spectrum of the enzyme. Ligands which directly coordinate to the ferric iron (or which indirectly coordinate by interacting with the retained axial water ligand), or which displace the axial water ligand without Fe$^{3+}$ coordination cause a red (Type II) or blue (Type I) shift in the maximum wavelength of the Soret band ($\lambda_{\text{max}}$), respectively. The magnitude of the $\Delta\lambda_{\text{max}}$ (nm) is dependent on the chemical nature of the ligand, the binding affinity ($K_D$), and the concentration of the ligand and enzyme under specific assay conditions.\textsuperscript{56} Binding interactions between ligands and active site residues distal to the heme cofactor may also be detected by monitoring the decrease or increase in the $\Delta\lambda_{\text{max}}$ caused by a known Type II ligand when the two compounds are assayed in combination. This UV-Vis assay provided a rapid method by which to identify and rank CYP144A1 ligands, and was used to screen a series of different small molecule and fragment libraries (Figure 3.2).
Figure 3.2. Ligand profiling and characterization strategy for the orphan Mtb enzyme CYP144A1.

DSC, differential scanning calorimetry; DSF, differential scanning fluorimetry; EPR, electron paramagnetic resonance spectroscopy; $K_D$, dissociation constant; Mw, molecular weight; NanoESI, Nano-electrospray ionization mass spectrometry; UV-Vis, UV-visible spectroscopy.

**Azole antifungal drugs** - Eight azole antifungal drugs that were previously shown to bind to a number of P450s were screened for binding to CYP144A1 using UV-Vis spectrophotometry. All compounds with structures that contained imidazole rings caused a red (Type II) shift ($\geq +2$ nm, 50 μM ligand) in the $\lambda_{max}$ (Figure 3.3A) of the CYP144A1 Soret band. In contrast, triazole-containing compounds, such as fluconazole and voriconazole, produced only minor perturbations in the optical spectrum, and no significant $\Delta \lambda_{max}$ was observed for itraconazole ($\Delta \lambda_{max} < \pm 1$ nm). The binding affinities ($K_D$ values) of the antifungal drugs for CYP144A1 were determined by optical titration (Table 3.1). Low nanomolar to micromolar $K_D$ values ($K_D = 0.095 - 15.3$ μM) were found for the imidazoles clotrimazole, econazole and miconazole, while ketoconazole had a weaker $K_D$ value of 103 μM. Triazole-containing drugs exhibited weak to negligible binding affinity, with calculated $K_D$ values of 6.5 - 10 mM. Other Mtb P450s that have been characterized to date also show weaker affinity for triazole-
containing antifungal drugs, which is consistent with the lower pKa of the heterocycle relative to imidazole.\textsuperscript{57} However, the differences in affinity for imidazole versus triazole compounds for these P450s are not as large as those observed for CYP144A1. Sigmoidal titration curves, which are suggestive of cooperative binding interactions, were observed for the binding of CYP144A1 to ketoconazole and voriconazole.\textsuperscript{30} Consequently, data for these compounds were fitted using the Hill equation, while the $K_D$ values for the tighter binding antifungals were obtained from data fitting using a modified version of the Morrison equation.\textsuperscript{30}

**Amino-pyrazole compounds** - A library of amino-pyrazole compounds was synthesized during our previous efforts to develop inhibitors of *Mtb* CYP121A1.\textsuperscript{58} Eight out of 21 of these compounds that were screened were also identified to bind to CYP144A1, of which 1-3 were the most potent (Figure 3.3b, and Figure S3.1, Supporting Information). Compounds 1 - 3 each exhibited a different mode of binding to CYP144A1, as characterized by their effect on the $\lambda_{max}$ of the Soret band. Compound 1 (50 μM) caused a 1 nm blue shift in the Soret, while compound 2 (50 μM) caused a 2 nm red shift, identifying them as Type I and Type II ligands, respectively. Compound 1 is the first Type I “substrate-like” ligand identified for CYP144A1. It was calculated to have a binding affinity of 82 μM for CYP144A1, which is in the same order of magnitude as that of CYP121A1 for its substrate cyclo-L-Tyr-L-Tyr (cYY) ($K_D = 21$ μM).\textsuperscript{27} Compound 2 was found to have a slightly weaker binding affinity of 201 μM. Compound 3 was identified as a non-heme binding ligand ($\Delta\lambda_{max} = 0$ nm), which competitively decreased the red shift of the CYP144A1 Soret band by 1.5 nm, compared to that observed in the presence of econazole alone ($\lambda_{max} = 423$ nm, 50 μM). The remaining 5 CYP144A1 ligands identified from the amino-pyrazole compound library were also found to have a non-heme mode of binding. A $K_D$ value of 136 μM was determined
for compound 3 using ITC, which is intermediate to the binding affinities of compounds 1 and 2 determined by optical titration (Figure 3.4). However, a 10 - 100 fold difference in the $K_D$ values determined by different biophysical techniques has been previously noted,$^{58-60}$ and, as such, a $K_D$ value of $\approx 14 \mu M$ for 3 might provide a better comparison between the relative affinities of the amino-pyrazoles 1 - 3. This would also be consistent with other methods used to rank the binding affinity of the ligands, such as by native mass spectrometry (Figure 3.9).

**Metal-binding fragment library** - A library of 80 “heme binding” fragments was screened against CYP144A1 and 6 other *Mtb* P450s using UV-Vis spectrophotometry. All of the fragments in this library contained functional groups demonstrated to bind to heme in the literature and in X-ray crystal structures deposited in the PDB. Only one fragment, the internal positive control in the library 4-phenylimidazole (4-PIM), was identified as a hit for CYP144A1. In contrast, other *Mtb* P450s bound to 10-24% of fragments in the library (data not included). A $K_D$ value of $\approx 3 \text{mM}$ was calculated for CYP144A1 binding to 4-PIM, which was significantly weaker that that calculated for other P450s (Table 3.1). A small library of 17 fragments structurally related to 4-PIM was compiled to explore the effect of imidazole substituents on heme coordination and to establish whether other heterocycles could be substituted in place of the imidazole ring. Two of these fragments, 1-phenylimidazole (1-PIM) and 4-(4-bromophenyl)imidazole 4 bound to CYP144A1, producing Type II shifts in the Soret band. The related 2-phenylimidazole isomer (2-PIM) produced only minor perturbations in the Soret band and no fragments demonstrated Type I binding interactions. The binding affinity of CYP144A1 for 1-PIM ($K_D = 280 \mu M$) was 10-fold stronger than 4-PIM, showing good correlation with the basicity of the respective isomers.$^{61}$ Despite the similar pKa values of the heterocycles, the weaker affinity of 2-PIM ($K_D \approx 5 \text{mM}$) relative to 4-PIM, highlighted the geometric restrictions on heme coordination.
for the respective isomers. Fragments substituted at C2 of the imidazole ring with alkyl, aromatic or polar functional groups showed negligible binding interactions as linear N-Fe coordination would likely result in steric clashing of the remainder of the ligand with the porphyrin (Figure 3.3C). In contrast to 4-PIM, fragment 4 ($K_D = 227 \mu M$) had comparable binding affinity to 1-PIM (Figure 3.4). This significant improvement in binding affinity from the addition of a 4-bromo substituent highlighted the relative importance of forming interactions with CYP144A1 active site residues and suggested that elaboration from the phenyl ring of the fragments could be pursued to develop more potent or selective CYP144A1 ligands.

![Figure 3.3. Structures of CYP144A1 ligands and fragments.](image)

(a) Structures ofazole antifungal drugs arranged in order of binding affinity. (b) Novel Type I 1, Type II 2 and non-heme binding 3 amino-pyrazole ligands. (c) Imidazole fragments arranged according to binding affinity.
Figure 3.4. Binding affinity titrations of CYP144A1 with amino-pyrazole ligands and fragments.

(a, b, c) Difference spectra (top panel) and concentration-dependent change in heme absorbance (bottom panel) for the titration of CYP144A1 (6 μM) with (a) compound 1, (b) compound 2, or (c) fragment 4, respectively. Data were fitted using the Hill equation to obtain $K_D$ values for compounds 1 ($h = 1.8$) and 4 ($h = 2.4$), or using a hyperbolic function for compound 2; (d) ITC binding isotherm for the titration of compound 3 (1 mM) with CYP144A1 (60.9 μM). The residual heat of ligand dilution was subtracted as a constant value. The integrated enthalpy change was fitted using a one-site binding model setting $N = 2$. The ITC trace and data fitting for $N = 1$ are provided in the Supporting Information (Figure S3.2).
Type II (red), Type I (blue) and non-heme (green) modes of binding are indicated by shading. Non-binding (NB) and untested (-) ligands are also identified. All binding affinity ($K_D$, μM) values were calculated by optical titration except for CYP144A1-3, CYP121A1-1 and CYP121A1-2, which were determined using ITC. All $K_D$ values are provided for the truncated construct of CYP144A1, except for voriconazole and fluconazole, which were determined for the full-length CYP144A1 protein. *Values determined using ITC.

<table>
<thead>
<tr>
<th>Compound</th>
<th>CYP144A1</th>
<th>CYP121A1&lt;sup&gt;26&lt;/sup&gt;</th>
<th>CYP125A1&lt;sup&gt;27&lt;/sup&gt;</th>
<th>CYP142A1&lt;sup&gt;44&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_D$ (μM)</td>
<td>$K_D$ (μM)</td>
<td>$K_D$ (μM)</td>
<td>$K_D$ (μM)</td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>0.095 ± 0.054</td>
<td>0.073 ± 0.008</td>
<td>5.3 ± 0.6</td>
<td>3.8 ± 0.9</td>
</tr>
<tr>
<td>Econazole</td>
<td>0.307 ± 0.075</td>
<td>0.024 ± 0.006</td>
<td>11.7 ± 0.7</td>
<td>4.6 ± 0.2</td>
</tr>
<tr>
<td>Miconazole</td>
<td>15.3 ± 0.9</td>
<td>0.136 ± 0.021</td>
<td>4.6 ± 0.4</td>
<td>4.0 ± 0.5</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>103.7 ± 5.7</td>
<td>3.44 ± 0.31</td>
<td>27.1 ± 0.9</td>
<td>21 ± 4</td>
</tr>
<tr>
<td>Fluconazole</td>
<td>6510 ± 470&lt;sup&gt;30&lt;/sup&gt;</td>
<td>16.3 ± 2.11</td>
<td>NB</td>
<td>NB</td>
</tr>
<tr>
<td>Itraconazole</td>
<td>NB</td>
<td>-</td>
<td>30.2 ± 4.3</td>
<td>NB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>CYP144A1</th>
<th>CYP121A1&lt;sup&gt;26&lt;/sup&gt;</th>
<th>CYP125A1&lt;sup&gt;27&lt;/sup&gt;</th>
<th>CYP142A1&lt;sup&gt;44&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81.9 ± 3.8</td>
<td>45 ± 5*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>201 ± 6</td>
<td>270 ± 18*</td>
<td>41 ± 3</td>
<td>0.30 ± 0.03</td>
</tr>
<tr>
<td>3</td>
<td>136 ± 27*</td>
<td>0.29 ± 0.04</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-PIM</td>
<td>2980 ± 300</td>
<td>32.2 ± 2.2</td>
<td>216 ± 5</td>
<td>12 ± 1.5</td>
</tr>
<tr>
<td>1-PIM</td>
<td>293 ± 74</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-PIM</td>
<td>5180 ± 1380</td>
<td>102 ± 10</td>
<td>345 ± 4</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>227 ± 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
3.3.3 Fragment screening using ligand-observed NMR

A library of 81 randomly selected, Rule of 3 compliant fragments was screened in cocktails of 3 compounds against CYP144A1 and a panel of other Mtb P450s by relaxation edited, ligand-observed NMR, 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 (Figure S3.13A, Supporting Information). Twenty six of the 81 fragments screened were shown to bind to CYP144A1 (Figure S3.14, Supporting Information). This hit rate of 32% was low compared to that of the other Mtb P450s screened, which typically bound 60-80% of this fragment library. CYP121A1 was the only P450 screened which had a comparable hit rate to CYP144A1, binding only 40% of the fragment library (unpublished data). Analysis of the chemical structures of fragment hits revealed a high proportion of bulky (av. Mw = 198 Da) bi-aryl (31%) and benzyl-phenyl (23%) scaffolds, in addition to fused benz-azole (19%), substituted phenyl (15%) and quinoline (12%) scaffolds (Figure 3.5A).

To increase the diversity of scaffolds identified, a second set of 51 fragments was screened as singletons using a combination of saturation transfer difference (STD), WaterLOGSY and CPMG ligand-observed NMR experiments. These 51 fragments were selected at random from a larger 800 compound library which had itself previously been screened against CYP144A1 by differential scanning fluorimetry (DSF). Inconsistencies in the results from DSF fragment screening with CYP144A1 led us to pursue other biophysical techniques, including the UV-Vis and ligand-observed NMR fragment screening experiments reported here. Twenty fragments were identified to bind to CYP144A1 using at least 2 out of the 3 experiments, CPMG, STD and WaterLOGSY, including fragment 4, which was cross-represented in this fragment library (Figure S3.15, Supporting Information). Indole (40%),
indazole (10%) and imidazole (15%) heterocycles were overrepresented amongst the chemical scaffolds of the fragment hits identified, which were slightly smaller on average (av. Mw = 164 Da) than those in the initial NMR screening library (Figure 3.5B). The detection of these smaller fragments could result from the higher concentration of fragments (1 mM) used in the experiments, which might also account for the slightly higher, albeit comparable, hit rate of 39% that was obtained.

Figure 3.5. Structural composition and hit rates from fragment screening by ligand-observed NMR spectroscopy.

(a) Fragments identified to bind CYP144A1 (20 μM) when screened as cocktails of 3 fragments (each 200 μM) using CPMG NMR experiments. Fragment binding was defined by a cut-off of >80% reduction in ligand 1H signal intensity relative to identical cocktails not containing protein. The composition of the fragment hits identified, the library screened, and the relative proportion of a given scaffold type identified as a hit compared to library composition has been broken down according to scaffold type; (b) Fragments identified to bind CYP144A1 (20 μM) when screened in singleton (1 mM) using a combination of WaterLOGSY, STD and CPMG experiments. The composition of the fragment hits identified has been broken down according to scaffold type.

3.3.4 Characterization of fragment binding mode using NMR spectroscopy

Competition NMR experiments were performed to validate that fragment hits bound to CYP144A1 within the active site, and to exclude false-positive results arising from non-specific interactions with the protein surface or from compound aggregation.63 Either clotrimazole or econazole were selected as displacer probes and were screened in
combination with fragments using CPMG NMR experiments. Only 5 fragments showed competitive binding interactions with azole probes, as indicated by a return of the $^1$H signal intensity of fragments in $^1$H-NMR spectra (Figure 3.6A and Figure S3.16a, Supporting Information). Four fragments, in addition to 4-PIM, showed cooperative binding interactions with displacer probes, with a further reduction in proton signal intensity (Figure 3.6B and Figure S3.16B, Supporting Information). The remainder of the fragments did not interact with the displacer probes. However, competition CPMG NMR experiments do not take into account the potential for a target to accommodate multiple ligands simultaneously, or for those ligands to acquire multiple binding modes. This could occur if the active site is sufficiently large or if the protein undergoes conformational change in response to an initial binding event. In support of this possibility, competition CPMG NMR experiments between compounds 1 and 2, both of which had been shown to bind to the heme cofactor of CYP144A1 by UV-Vis and EPR spectroscopy (vide infra), also failed to produce the expected return of proton signal intensity that would be indicative of competitive binding interactions (Figure S3.13B, Supporting Information). We thus proceeded to characterize the binding mode of CYP144A1 fragments and ligands using a variety of other spectroscopic techniques.
Figure 3.6. CPMG ligand-observed NMR spectra of CYP144A1 with competitive and cooperatively binding fragments.

(a) Fragment 4 (1 mM), and (b) Fragment 5 (1 mM) in buffer, with CYP144A1 (20 μM), and after the addition of a displacer (clotrimazole, 500 μM). Buffer in (a) and (b) is “buffer and ligand“. Binding interactions are indicated by a decrease in the signal intensity of ligand protons in the presence of protein, and have been highlighted with red arrows. Displacement of fragment 4 binding by clotrimazole restored ligand signal intensity and is highlighted by blue arrows. Cooperative binding interactions further reduce the signal intensity of fragment 5. Results were consistent between CPMG, WaterLOGSY and STD experiments.

3.3.5 Characterization of ligand binding mode by competition UV-Vis spectroscopy

None of the fragments identified to bind to CYP144A1 in ligand-observed NMR experiments, except for fragment 4, directly perturbed the Soret band of the CYP144A1 optical spectrum when analyzed by UV-Vis spectrophotometry. Therefore, competition UV-Vis experiments were conducted to investigate whether the fragments might bind elsewhere within the active site and perturb the interaction of validated Type II heme binding ligands, so modifying their effect on the Soret $\lambda_{\text{max}}$. Econazole, miconazole and compound 2 were each employed as Type II “indicator” ligands for competition assays, providing a wide range of binding affinities ($K_D = 0.3 - 200 \mu M$) and steric volumes to detect fragment interactions. Indicator ligands were used at the minimum concentration required to perturb the CYP144A1 Soret band ($\Delta \lambda_{\text{max}}$) by $\geq 1.5$ nm, in order to increase the possibility of their displacement by weakly binding fragments. Changes in the $\lambda_{\text{max}}$ of $< 1$ nm were considered within experimental error. Seven of 26 fragments showed competitive binding interactions with
miconazole, decreasing the red shift in the CYP144A1 Soret band compared to that observed with miconazole alone (Figure 3.7A and Figure S3.17, Supporting Information). The same 7 fragments either bound non-competitively with compound 2, having no effect on the red shift of the Soret (ΔΔλ_{max} ≈ 0 nm), or showed cooperative effects, further increasing the Δλ_{max} compared to that found with compound 2 alone. The chemical scaffolds represented by fragments that interacted with indicator ligands were all large (av. M_W = 213 Da) lipophilic bi-aryl, benzyl-phenyl or halogenated phenyl structures. Both competitive and cooperative optical interactions increased our confidence that fragments were binding within the CYP144A1 active site. However, aggregation with the indicator ligand, or solubility equilibrium-driven changes in binding affinity could produce similar optical effects and cannot be excluded. No competitive or cooperative interactions were observed for the remaining 19 fragments, which could be a consequence of their small size or weak binding affinity relative to the Type II indicator ligands.

The binding interactions of compounds 1 and 3 with CYP144A1 were also analyzed in competition by UV-Vis, employing econazole (50 μM) and compound 2 (50 μM) as Type II indicator ligands. An equimolar addition of either compound 1 or 3 caused a 1-1.5 nm decrease in the red-shift of the CYP144A1 Soret band compared to that observed with econazole alone. Competitive interactions were also observed between compounds 1 or 3 and compound 2, although a higher concentration of compound 1 (100 μM) was required to detect a decrease in Δλ_{max} (Figure 3.7B). As UV-Vis assays demonstrated that compounds 1 - 3 and azole compounds compete for interactions with the heme cofactor, but these ligands did not mutually displace each other in ligand-observed NMR experiments (vide supra), it is proposed that the compounds must occupy a secondary binding location within the CYP144A1 active site.
Figure 3.7 CYP144A1 UV-Vis absorption spectra from competition binding assays.

(a) CYP144A1 (4-5 μM) absorbance spectra collected for ligand-free (black) protein or CYP144A1 in the presence of fragment 6 (1 mM, black, dashed), miconazole (12.5 μM) (orange), compound 2 (50 μM, red), both 6 (1 mM) and miconazole (12.5 μM, orange, dashed), or both 6 (1 mM) and compound 2 (50 μM, red, dashed). Black dots indicate no change in λ_max with addition of the fragment alone; (b) CYP144A1 (4-5 μM) absorbance spectra collected for the ligand-free protein (black) or CYP144A1 in the presence of compound 1 (100 μM, blue), compound 2 (50 μM, red), or both compound 1 (100 μM) and compound 2 (50 μM) (green). Ligands were prepared as stock solutions in d_6-DMSO. Concentrations of d_6-DMSO did not exceed 1% v/v in competition assays and were not found to alter the λ_max of the ligand-free CYP144A1 Soret band (λ_max = 421 nm) within this concentration range. Arrows indicate the direction of the change in Δλ_max produced when fragment 7 or compound 1 were combined with CYP144A1 in the presence of indicator ligands miconazole or compound 2.

3.3.6 Characterization of ligand binding mode using EPR spectroscopy

X-band EPR spectra of CYP144A1 were collected to characterize the binding modes of ligands, including their interactions with the heme coordination sphere and distal active site residues. Ligand-free CYP144A1 (200 μM) generated a characteristic low-spin P450 rhombic spectrum. The low-spin g-values of 2.41 (g_z), 2.24 (g_y) and 1.92 (g_x) observed are consistent with a water-bound, cysteine thiolate-coordinated ferric P450 resting state.\textsuperscript{30,35} The imidazole antifungal compounds clotrimazole, econazole, miconazole and ketoconazole all bound to
CYP144A1, producing a new set of g-values (e.g. 2.44/2.25/1.89 in the case of both ketoconazole and miconazole). The perturbation to the azole-bound EPR spectra (specifically, the increase in \( g_z \)-values and decrease in the \( g_x \)-values relative to the resting form of CYP144A1) are consistent with the ligands interacting with the heme iron through a strong-field ligand, such as the imidazole nitrogen atom (Figure 3.8B and Figure S3.18, Supporting Information).\(^{57,64}\) The shifts in the g-values that were observed on the binding of the azole compounds and all other ligands (amino-pyrazole compounds and fragments) to CYP144A1 were rather small, and might reflect indirect coordination of the heme iron through a retained 6\(^{th}\) axial water ligand, mediated by an aromatic nitrogen atom.\(^{65,66}\)

Compounds 1, 2 and 3 generated similar low-spin EPR spectra with CYP144A1 to those observed with the azole drugs, e.g. 2.44/2.24/1.89 with compound 1. These data again suggest direct or indirect (via retained axial water) interaction of the ligands with the heme iron through a nitrogen atom, likely from the imidazole, indazole and aniline functional groups present in these compounds (Figure 3.8A and Table S3.2, Supporting Information).

However, as observed in the spectra generated by the more weakly binding azole compounds miconazole and ketoconazole, compounds 1 - 3 produced a heterogeneous set of g-values. These spectra indicated a minimum of two different coordination states of the CYP144A1 heme iron, possibly indicating that the novel ligands are able to acquire different binding modes with respect to the heme cofactor.\(^{67}\) However, there clearly remains a proportion of the resting (water-coordinated) form in the spectra for CYP144A1 bound to compounds 1 - 3. For instance, the 2.41/2.24/1.92 set of g-values seen for the CYP144A1-compound 1 sample. The EPR spectra indicate that the proportion of the novel, ligand-bound CYP144A1 species decreased in the order of compound 2>1>3. Compound 3 also produced a slightly less extensive shift in the g-values (2.43/2.24/1.90) compared to compounds 1 and 2, likely
reflecting the “non-heme interacting” mode of binding of this compound, as observed from UV-Vis spectroscopy.

Despite compound 1 producing substrate-like, Type I optical spectra in UV-Vis assays, no formation of high-spin CYP144A1 was observed in the compound 1-bound EPR spectrum, or indeed for CYP144A1 complexes with any of the other ligands tested. However, this is likely a consequence of the conditions required for heme EPR experiments.

Nineteen fragments that had been identified to bind CYP144A1 using a combination of UV-Vis and NMR spectroscopy were also analyzed by EPR. All spectra collected for CYP144A1 (200 μM) in complex with fragments (500 μM) were typical of low-spin, rhombic P450 spectra. Eight of the 19 fragments bound to CYP144A1 to produced shifts in the g-values consistent with fragments binding in the active site by either direct or indirect coordination of the heme iron (Figure 3.8B and Figure S3.19, and Table S3.3, Supporting Information). Four of the eight fragments induced large spectral shifts, including the imidazole-containing fragment 4, which generated a near-homogeneous set of g-values (2.46/2.26/1.89) that is consistent with direct heme coordination by an imidazole nitrogen atom. All other fragments generated a heterogeneous set of g-values indicating weak binding or multiple binding modes. The g-values for CYP144A1 bound to fragment 5, which was identified as non-heme interacting in UV-Vis assays, were identical to those for the CYP144A1-compound 3 complex, consistent with the proposed modes of binding of these compounds. Notably, all fragments identified to interact with the heme coordination sphere in EPR spectra also showed inter-ligand interactions with clotrimazole in competition NMR experiments and shared a common indole or imidazole-like scaffold.
Figure 3.8. X-band EPR spectra of CYP144A1 with a selection of heme binding and non-heme binding ligands.

(a) EPR spectra of ligand-free CYP144A1 (200 μM in each case), and for CYP144A1 bound to compounds 1, 2 and 3 (each 500 μM); (b) EPR spectra of ligand-free CYP144A1, and CYP144A1 bound to clotrimazole and to fragments 4 and 5 (each 500 μM). The g-values of ligand-free CYP144A1 (g_z = 2.41, g_y = 2.24, g_x = 1.92) and the novel g-values generated by ligand binding interactions have been indicated. Tabulated g-values for each species are provided in Table S3.1, Supporting Information.

3.3.7 Characterization of ligand binding using nanoESI native mass spectrometry

Nano-electrospray ionization (NanoESI) mass spectrometry was used to determine the stoichiometry of ligand-CYP144A1 binding interactions and to rank ligands according to their relative binding affinities. NanoESI mass spectra of ligand-free CYP144A1 (10 μM) confirmed previous reports that the protein is predominantly monomeric (Figure S3.20, Supporting Information). Peaks in the range 3200-4400 m/z were assigned to the CYP144A1 monomer and corresponded to a protein with a mass of 46611 ± 75 Da, which is in close agreement with the mass calculated from the amino acid sequence of CYP144A1, plus 1 equivalent of heme-β (calculated mass = 46625 Da). The addition of d_6-DMSO (5% v/v), which can alter the charge state distribution of a protein by suppressing the basic sites available for protonation, caused protein peaks to shift to slightly higher m/z values.
NanoESI spectra of ligand-bound CYP144A1 were collected for clotrimazole, econazole, miconazole and compounds 1 - 3, at 1:1 and 1:5 ratios of protein-to-ligand (Figure 3.9). The appearance of 1-2 additional peaks, species B and C respectively, at higher m/z values was observed in nanoESI spectra for all compounds. The peaks were calculated to have masses corresponding to CYP144A1 binding to one and two molecules of ligand per protein monomer, respectively. The relative intensity of the ligand-bound peaks correlated reasonably well with the binding affinities of the respective ligands. However, greater consistency was observed between ligands that shared a similar chemotype. For example, the proportion of CYP144A1 bound to two ligand molecules (species C) was greater for compounds 1 and 3, than for compound 2. For clotrimazole, essentially none of the peaks corresponding to the ligand-free CYP144A1 protein (species A) remained at a 1:5 protein-to-ligand ratio, under which conditions there was near-complete formation of a CYP144A1 complex bound to two molecules of clotrimazole. In contrast, only one molecule of econazole or miconazole was observed to bind within the concentration range tested. Variation in the gas phase binding affinities of the amino-pyrazole and azole antifungal drugs could be a consequence of their different chemophysical properties, especially clogP. The trends were consistent with the contrasting effect that the gas phase of the mass spectrometer has on polar versus hydrophobically driven intermolecular interactions, and suggest that the binding interactions of econazole and miconazole to CYP144A1 are predominantly hydrophobic in nature.
<table>
<thead>
<tr>
<th>Compound</th>
<th>M_W (Da)</th>
<th>K_D (μM)</th>
<th>Stoichiometry</th>
<th>Species Intensity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>L:P (1:1)</td>
<td>L:P (5:1)</td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>345</td>
<td>0.095</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Econazole</td>
<td>382</td>
<td>0.307</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Miconazole</td>
<td>416</td>
<td>15</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>358</td>
<td>136</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>387</td>
<td>82</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>347</td>
<td>201</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**Figure 3.9.** NanoESI native mass spectra of CYP144A1-ligand binding interactions.

Top panel: NanoESI spectra of ligand-free CYP144A1 (10 μM), and CYP144A1 (10 μM) in the presence ofazole antifungals and compounds 1 - 3 (50 μM). All spectra were collected in ammonium acetate (200 mM) buffer, pH 7.0, containing a final d_6-DMSO concentration of 5% (v/v). Major peaks m/z = 4000-5400 correspond to the CYP144A1 monomer. Minor peaks correspond to an unidentified protein impurity. Dashed lines A, B, C mark the m/z corresponding to the ligand-free protein, and CYP144A1 bound to one molecule and two molecules of ligand, respectively. Tabulated molecular weight (Mw), binding affinity (K_D), binding stoichiometry at 1:1 and 5:1 ligand-to-protein (L:P) ratios, and relative intensity of peaks corresponding to the different ligand-bound species in the spectra collected at 5:1 L:P ratios are shown.
3.4 Discussion

The predicted role of CYP144A1 in essential pathogenic processes and drug resistance mechanisms of *Mtb* raised interest in the biochemical characterization of this orphan protein. The realization that CYP144A1 exhibits tight binding affinity for a series of antimicrobial azole drugs, second only to CYP121A1 from among of all the characterized *Mtb* P450s, established the enzyme as a potential drug target for *Mtb*. Early attempts to identify CYP144A1 substrates or inhibitors by screening a diverse collection of lipids, sterols and other known P450 substrates were unsuccessful (see Methods S3.12, Supporting Information). Furthermore, the lack of a characterized operon structure surrounding the *CYP144A1* gene and the low sequence identity of the protein to other P450s of known function presented a significant challenge for determining enzyme function. As such, a chemoproteomics approach was pursued to develop a ligand profile of CYP144A1 by using a combination of fragment screening, comparative bioinformatics, structural analysis and detailed biophysical characterization.

Alignment of the amino acid sequence of CYP144A1 with functionally characterized enzymes indicated a potential role for CYP144A1 in macrolide biosynthesis or metabolism. While the low sequence identity scores (33%) of the best alignments restricted our confidence in predicting protein function, the consistency in the function of the enzymes identified hinted that large macromolecules might fit the CYP144A1 substrate profile. Analysis of the X-ray crystal structure of CYP144A1 using SiteMap software supported this hypothesis, revealing the protein to have a large active site volume and surface area (Figure 3.1). Greater similarities in the active site and substrate entry channels of CYP144A1 and CYP121A1, compared to those of CYP125A1 and CYP142A1, correlated with the relative binding affinities of the different P450 isoforms for a panel of azole drugs. The
structure of the active sites also reflected the substrate chemotype of each enzyme; with the narrow, linear substrate channels of CYP125A1 and CYP142A1 poised to accommodate their cholesterol/4-cholesten-3-one substrates,\textsuperscript{29,50,52,71} while the CYP121A1 substrate cyclo-dityrosine (cYY) binds favorably in the shorter, wider substrate channel of this P450 isoform.\textsuperscript{27} These comparisons point towards a CYP144A1 substrate profile comprised of large, globular ligands and demonstrates the insight into the biological function of orphan proteins that can be gained from ligand profiling.

Classification of CYP144A1 as druggable by SiteMap calculations provided confidence in our ability to establish a ligand profile of the enzyme and to develop potent ligands as drug leads, as has recently been achieved against CYP121A1 (Figure 3.1).\textsuperscript{58} However, the low hit rates obtained for CYP144A1 compared to other P450s, from screening various fragment libraries by UV-Vis and ligand-observed NMR, correlates well with the difficulty experienced in trying to identify potential substrates and/or inhibitors of this enzyme. The \textit{Mtb} P450 with the next lowest hit rate was CYP121A1, which correlates with the similar active site structure of the two enzymes. The similarities also suggest that, like CYP121A1 and its substrate-product pair cYY and mycocyclosin, the endogenous CYP144A1 substrate could have a complex or unusual chemical structure. Fragment screening is a common experimental method to assess the druggability of protein targets, and results obtained here suggest that CYP144A1 may have unique properties.\textsuperscript{18}

The unusually small surface area available for metal binding interactions that was calculated for CYP144A1 provided some insight into the narrow profile of heme binding pharmacophores or functional groups found to be tolerated by the enzyme. Fragment-screening has previously been used to characterize the preferred metal-binding
pharmacophore of various classes of metalloproteins, and the elaboration of fragments identified from these studies has aided the development of more potent or selective ligands.\textsuperscript{72–74} In our own experience, other \textit{Mtb} P450s were found to bind 10-24\% of fragments from a focused compound library which was enriched with fragments containing metal binding functional groups (\textit{unpublished data}). Distinct preferences between specific isoforms and particular functional groups were apparent. For example, the preference of CYP121A1 for primary anilines or CYP125A1 for \textit{para}-substituted pyridine rings, trends which are consistent with the structures of CYP121A1 and CYP125A1 inhibitors reported in the literature.\textsuperscript{58,75,76} Furthermore, fragment binding preferences reflect the topology of the active sites of these particular P450s. For example, the wide active site of CYP121A1 enables sp\textsuperscript{3}-hybridized aniline ligands to approach the heme at an acute (\(\approx 32^\circ\)) angle and thus benefit from additional aromatic stacking interactions with the heme porphyrin, while \textit{I}-helix residues disfavor the orthogonal ligation of heterocyclic N-ligands.\textsuperscript{58,66,77} In contrast, the narrow architecture of the CYP125A1 and CYP142A1 active sites favors sp\textsuperscript{2}-hybridized N-donors which can coordinate orthogonal to the porphyrin plane. CYP144A1 was only identified to bind to one fragment from this library, the internal positive control 4-PIM, and screening of a library of structurally related 4-PIM analogues only resulted in two additional hits; 1-PIM and fragment 4. The high selectivity of CYP144A1 for imidazole was rationalized with respect to the small surface area available for metal binding interactions, which might disfavor larger heterocyclic or aniline groups. Additionally, the close proximity and orientation of active site residues His292 and Phe289 to the heme cofactor in CYP144A1 might favor T-shaped aromatic stacking interactions with imidazole heterocycles (\textbf{Figure 3.1D} and \textbf{Figure 3.10B}). While the identification of 4-PIM, 1-PIM and 4 as fragment hits is unsurprising when considering that the tight binding azole drugs clotrimazole, econazole, miconazole and ketoconazole also contain 1-N substituted imidazole rings, the \(K_D\) values of
these ligands differ by more than 4 orders of magnitude ($K_D = 95 \text{ nM} - 3 \text{ mM}$). Furthermore, a wide range of other substituted imidazole fragments that were screened throughout our ligand profiling campaign failed to perturb the optical spectrum of CYP144A1 (Figure 3.3). These data indicate that, in addition to heme coordination, interactions made by ligands with residues in the CYP144A1 active site contribute significantly to binding affinity and selectivity. Furthermore, it suggests that even “promiscuous” metal binding pharmacophores, such as imidazole, might be structurally modified to develop selective P450 ligands.

Similarities in the active site structure and binding affinities of CYP144A1 and CYP121A1 for azole drugs prompted us to screen a library of amino-pyrazole ligands, which were originally developed against CYP121A1,\cite{58} in order to identify tool compounds to help characterize the properties of CYP144A1, and also to aid our subsequent fragment-based ligand profiling studies. The identification of a Type I, Type II and a number of non-heme binding ligands from this library, exemplified by compounds 1, 2 and 3 respectively, validated our comparative structural and chemotype profiling approach to ligand identification. Compound 1 is the only ligand that has been identified to date that produces a “substrate-like”, Type I optical spectrum with CYP144A1 (Figure 3.7B). Whilst no catalytic turnover of 1 was observed in biochemical studies (Methods S3.1, Supporting Information), and EPR spectroscopy did not indicate any formation of the catalytically-poised high-spin enzyme (Figure 3.8A), this could be a consequence of the non-native conditions employed in these experiments. The low temperature (ca 10 K) required for X-band EPR of heme proteins has been shown to stabilize the low-spin state of ferric P450s. For example, CYP125A1, an enzyme that is predominantly high-spin at ambient temperature, generates a rhombic EPR spectrum consistent with a low-spin P450.\cite{54} It was also shown that the validated CYP121A1 substrate cYY produces a predominantly low-spin EPR spectrum even at saturating ligand
concentrations. Furthermore, limited characterization and/or the low catalytic efficiency of the redox partner proteins utilized by \textit{Mtb} P450s mean that heterologous redox partner enzymes are commonly employed during turnover studies. While these systems might support CYP144A1 activity \textit{in vitro}, the oxidation products may alter depending on the nature of the interactions between the P450 and redox partner. As such, the possibility of compound 1 being a \textit{bona fide} CYP144A1 substrate cannot be excluded.

More than 43\% of the fragment hits from the two different fragment libraries that were screened using ligand-observed NMR contained indole, indazole and (benz)-imidazole scaffolds. The high proportion of these heterocyclic scaffolds was notable in light of the structures of compounds 1 and 2, and illustrates the potential that fragment screening has in identifying preferred ligand chemotypes. Indole, indazole and (benz)-imidazole scaffolds were also over-represented amongst the set of fragments that were characterized to bind to CYP144A1 in the vicinity of the heme prosthetic group. Displacement of bound fragments by clotrimazole in competition NMR experiments and the perturbation of g-values in EPR spectroscopic studies by these fragments suggests a preference for CYP144A1 to bind these heterocycles in close proximity to the heme. However, the lack of direct perturbation of the Soret $\lambda_{\max}$ of the CYP144A1 optical spectrum suggests either insufficiently strong binding affinity to disrupt the water-coordination network of the cofactor, or else binding modes that do not directly impact the environment of the heme axial ligand. The consistency in the binding mode characterized for this fragment scaffold supported the docking models that were generated in the absence of X-ray crystallographic data to explore the binding interactions of compounds 1 and 2 \textit{(vide infra)} (Figure 3.10A-C).
In contrast, the majority of fragments identified to bind in the distal part of the CYP144A1 active site shared structural similarities with the non-heme binding amino-pyrazole compound 3. Ligand-observed NMR screening of fragment cocktails by CPMG resulted in the identification of a large number of bi-aryl or benzyl-phenyl fragment scaffolds, substituted with amino groups or halogens (Figure 3.5A and Figure S3.14, Supporting Information). Unlike the smaller heterocyclic fragments, the non-heme interacting bi-aryl and benzyl-phenyl fragments competitively displaced azole drugs in UV-Vis assays, as had also been observed for compound 3. This is likely a result of both the greater steric bulk of the bi-aryl fragments relative to the smaller heterocycles and also their inherently stronger binding affinity. The strong affinity of non-heme binding ligands for CYP144A1 complemented our prediction that, based on the large variation in the \( K_D \) values of imidazole antifungal drugs and fragments 4-PIM, 1-PIM and 4, interactions formed by ligands with CYP144A1 active site residues were likely to make a significant contribution to binding affinity. The comparable binding affinity of non-heme binding compound 3 (\( K_D = 136 \mu M \)) to ketoconazole (\( K_D = 103 \mu M \)) and miconazole (\( K_D = 15 \mu M \)) (taking into account a consistent ~ 10-fold variation in \( K_D \) values calculated using ITC and optical titration methods), suggests that potent CYP144A1 inhibitors could be developed without the need to employ a heme-binding functional group. Ligands which contain metal-binding functional groups are widely considered to be more promiscuous, although recent evidence suggests that careful modification of ligand scaffolds can substantially improve selectivity.\textsuperscript{58,72,79} The 50-fold difference in the \( K_D \) values of econazole and miconazole, and the 13-fold difference in the \( K_D \) values of 4-PIM and fragment 4, resulting from the addition of a single halogen atom, illustrates the sensitivity of ligand binding affinity to ligand elaboration into the CYP144A1 active site. These SAR data, in addition to the narrow profile of heme binding
pharmacophores found to be tolerated by the enzyme, indicate that inhibitors with high selectivity could be developed against CYP144A1.

Cooperative binding effects observed between fragments and amino-pyrazole compound 2, or between fragments andazole drugs in UV-Vis (Figure 3.7) and ligand-observed NMR experiments (Figure 3.6), respectively, indicated that multiple ligands might bind simultaneously within the large CYP144A1 active site. This hypothesis was supported by evidence that while compounds 1 and 2 competed for mutually exclusive ligation of the heme cofactor in UV-Vis competition experiments (Figure 3.7B), they did not “displace” each other when assessed by ligand-observed NMR (Figure S3.13B, Supporting Information). As such, both compounds must remain in sufficient contact with the enzyme to acquire the slow T2-relaxation time of the macromolecule, one in proximity to the heme and the second binding closer to the entrance of the active site. The heterogeneity in EPR g-values generated by all ligands - azole drugs, amino-pyrazole compounds and fragments - was also consistent with the proposal that ligands could adopt different binding modes, or that different ligand-bound CYP144A1 species were present in solution. Furthermore, the sigmoidal titration curves and Hill values of 1.4-2.5 calculated when fitting the binding titrations of miconazole, ketoconazole, compounds 1, 3 and fragment 4, suggest cooperative binding interactions. Finally, nanoESI native mass spectrometry experiments confirmed a 2:1 ligand-to-CYP144A1 binding stoichiometry for azole drugs and compounds 1 - 3. The correlation between the proportion of ligand-bound protein species and the $K_D$ values of the ligands supported the conclusion that the spectra observed were the result of specific binding interactions. The 5-fold excess ligand-to-protein concentrations required to observe this 2:1 binding stoichiometry for clotrimazole and compounds 1 - 3 indicates that similar binding
stoichiometries could be expected in other biophysical assays, which typically employ higher concentrations of ligand-to-protein.

The ability for P450s to bind multiple ligands simultaneously within the active site is not unprecedented. An X-ray crystal structure of the human drug-metabolizing enzyme CYP3A4 revealed two molecules of ketoconazole bound in a head-to-tail arrangement in the active site.81 We have also previously reported the X-ray crystal structures of CYP121A1 in complex with multiple fragments, or small ligands (Mw = 250 Da).58,75 The ability for P450s to accommodate multiple ligands at once has been used to explain the broad substrate specificity and atypical kinetics of some isoforms, and has been attributed, in part, to the significant conformational flexibility in the tertiary structure of these proteins.59,81

Having identified a series of novel CYP144A1 ligands and fragments and explored their interactions with the enzyme using a variety of biophysical methods, co-crystallization and soaking experiments were pursued in the hope of obtaining a ligand-bound CYP144A1 structure to further validate the ligand binding profile. All compounds were analyzed by DSC prior to crystallization and shown to stabilize CYP144A1 by 0.9 - 3.3 °C (Tm1) relative to the ligand-free protein (Tm = 47.3 °C, 5% v/v DMSO). Thirteen of 21 fragments also produced a secondary melting point (Tm2) occurring 10.4 - 15.9 °C higher than the Tm of the ligand-free protein. Biphasic melting transitions would be consistent with different populations of CYP144A1 in solution and could correspond to different P450-ligand binding stoichiometries (Figure S3.21 and Table S3.4, Supporting Information).82,83 Unfortunately, no ligand density was observed in the CYP144A1 crystal diffraction data sets collected from ligand soaking or co-crystallization attempts. The combination of the open-structure of the CYP144A1 active site and the conformational flexibility of the protein, implied from the
results of the ligand profiling experiments reported here and predicted in previous studies, might account for these results. Ligand-docking studies were consequently employed in the absence of co-crystal structures to model the experimentally determined binding mode of the azole drugs, compounds 1-3 and fragments (Figures 3.10A-C). Docking was performed using parameters that summarized the culmination of data generated from UV-Vis, NMR, EPR and NanoESI mass spectrometry experiments. Displacement of the axial water ligand, or water-bridged hydrogen bonding interactions between the heme iron and indazole ring of compound 1, was predicted to account for the optical and EPR spectra generated by this compound. Direct N-Fe$^{3+}$ heme coordination was simulated for the imidazole-containing drugs and compound 2, again reflecting the Type II optical spectra and more significant perturbation in the g-values generated by these compounds in EPR spectroscopy. Compound 3 was allowed to dock using no constraints and posed favorably in a non-heme binding orientation that would be consistent with experimental data. Analysis of the binding poses generated through docking models indicated intermolecular interactions with active site residues, such as His292 and Phe289, which could be used to rationalize some of the SAR observed in the current study, and to subsequently guide the development of CYP144A1 inhibitors.
Figure 3.10. Rationalization of experimentally characterized CYP144A1-ligand binding interactions using ligand docking.

(a-c) Compounds 1, 2, and 3, respectively, in complex with CYP144A1. Compounds were docked into prepared protein grids to simulate (a) displacement of the axial water ligand or hydrogen bonding between the axial heme water and indazole ring of compound 1 (cyan), (b) direct N-Fe ligation (Type II binding) between the imidazole ring of compound 2 (salmon) and heme cofactor, or (c) non-heme binding interactions between compound 3 (green) and distal active site residues. The heme cofactor (magenta), active site residues (grey), heme axial water ligand (red sphere) and proposed intermolecular interactions (yellow dashes) are shown in each figure. Docking was performed using GLIDE v6.5 (Schrödinger, LLC, New York, NY). Details are provided in Experimental Methods. Figures were prepared using PyMOL v1.7.4 (Schrödinger, LLC).

In summary, a series of novel compounds and fragments that bind to the orphan *Mtb* P450 enzyme CYP144A1 have been identified by using a combination of fragment-based techniques and compound screening, and guided by comparative structural and bioinformatics analysis. The ligands cover a range of binding affinities (\(K_D = 82 - 3000 \, \mu M\)) and exhibit diverse modes of interaction with the CYP144A1 active site. The combination of these novel compounds and a collection of azole drugs describes the chemotype profile of CYP144A1 and provides a unique set of chemical tools with which to study the enzyme. Fragment screening demonstrated the strong preference of CYP144A1 for imidazole as a heme coordinating pharmacophore, and for indole/indazole scaffolds to bind in close proximity to the heme prosthetic group, in a mode that is reminiscent of P450 substrates. These preferences in fragment scaffold were used to rationalize the binding modes of compounds 1 and 2 in docking studies, and to provide insight into the intermolecular interactions that might be significant for ligand affinity and selectivity.
Fragment screening by ligand-observed NMR revealed a preference for large bi-aryl and benzyl-phenyl scaffolds, which shared a similar structure and binding mode to the non-heme binding compound 3. The ability for ligands to attain different binding modes in the CYP144A1 active site, and for the enzyme to bind to multiple large (>300 Da) ligands simultaneously, support a prediction of bulky endogenous substrates for CYP144A1. These results are consistent with the large size and open structure of the CYP144A1 active site, and also point towards a conformationally flexible protein structure. The relatively low hit rates obtained throughout our fragment screening campaign indicate the unique properties of this enzyme and suggest a distinct role for CYP144A1 within the family of \textit{Mtb} P450s. The accumulation of this detailed ligand profile will provide insight into the potential substrate and function of CYP144A1. The trends in binding affinity and SAR elucidated in this study provide direction for the subsequent development of more potent and selective ligands, which might have potential for optimization into therapeutic compounds.
3.5 Supporting Information

Figure S3.11. Aminopyrazole compounds identified using UV-Vis spectroscopy to bind to CYP144A1.

Ligands (0.1 - 1 mM) were screened against CYP144A1 (5 μM) using UV-Vis spectroscopy and classified as Type I or Type II ligands according to whether a blue or a red shift occurred in the $\lambda_{\text{max}}$ of the CYP144A1 Soret band, respectively. Ligands which did not directly perturb the Soret absorbance, but reduced the $\Delta \lambda_{\text{max}}$ caused by econazole (50 μM), were classified as non-heme binding ligands.

Figure S3.12. ITC binding isotherm and integrated enthalpy plot for the titration of CYP144A1 (50 μM) with compound 3 (1 mM).

A $K_D$ value of 230 ± 19 μM was calculated for compound 3 when titration data were fitted using a one-site equilibrium binding model, setting the stoichiometry as N=1.
Figure S3.13. Ligand observed NMR screening and binding mode characterization.

(a) CPMG spectra of cocktail of 3 fragments (each 200 μM) in buffer and in the presence of CYP144A1 (20 μM). Red arrows indicate proton signals corresponding to the fragment hit identified as methyl 4-hydroxy-3-iodobenzoate 6. (b) CPMG spectra of compound 1 (200 μM) and compound 2 (200 μM) tested independently in buffer and with CYP144A1 (20 μM); and also compound 1 and 2 tested in competition in the presence of CYP144A1 (20 μM).

Figure S3.14. Twenty-six fragment hits identified from cocktails screened by ligand-observed NMR CPMG experiment.
Figure S3.15. Twenty fragment hits identified by singleton ligand observed NMR screening using STD, CPMG and WaterLOGSY experiments.

Figure S3.16. Fragments showing competitive or cooperative binding interactions with clotrimazole in CPMG NMR experiments.

(a) Fragments (1 mM) which were displaced by clotrimazole (500 μM) in the presence of the CYP144A1 (20 μM); (b) Fragments (1 mM) which showed stronger binding interactions (further reduction in proton signal intensity) when combined with clotrimazole (500 μM) in the presence of CYP144A1 (20 μM).
Figure S3.17. Fragments (0.5-1 mM) which displaced miconazole (12.5 μM) in competition UV-Vis spectroscopy experiments.

Figure S3.18. X-band EPR spectra of CYP144A1 bound toazole antifungal drugs.

EPR spectra were collected for ligand-free CYP144A1 (200 μM in each case), and for CYP144A1 bound to econazole, miconazole and ketoconazole (each 500 μM). The g-values generated for the ligand-bound forms are indicated and compared to those of ligand-free CYP144A1 (g_z = 2.41, g_y = 2.24, g_x = 1.92). Tabulated g-values for each species are provided in Table S2.

Table S3.2. The g-values from X-band EPR spectra of CYP144A1 with selection of heme binding and non-heme binding ligands.

EPR spectra were recorded for ligand-free CYP144A1 (200 μM), and for CYP144A1 (200 μM) bound to clotrimazole and to compounds 1, 2, 3, 4 and 5 (each 500 μM). A minor species appearing as a shoulder on a larger feature is annotated with a superscript “sh”. The g_z feature for clotrimazole-bound CYP144A1 is broad.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Ferric low-spin g-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g_z</td>
</tr>
<tr>
<td>Ligand-free</td>
<td>2.41</td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>2.43</td>
</tr>
<tr>
<td>1</td>
<td>2.44, 2.41</td>
</tr>
<tr>
<td>2</td>
<td>2.44, 2.41</td>
</tr>
<tr>
<td>3</td>
<td>2.43, 2.41</td>
</tr>
<tr>
<td>4</td>
<td>2.46</td>
</tr>
<tr>
<td>5</td>
<td>2.43, 2.41</td>
</tr>
</tbody>
</table>
Figure S3.19. Fragments identified to bind CYP144A1 by EPR spectroscopy.

Fragments which produced large shifts in the g-values relative to ligand-free CYP144A1 (g_z = 2.41, g_y = 2.24, g_x = 1.92) are colored blue.

Table S3.3. The g-values generated by fragments and azole antifungal drugs binding to CYP144A1 in X-band EPR spectroscopy experiments.

EPR spectra of ligand-free CYP144A1 (200 μM), and CYP144A1 (200 μM) bound to fragments or azole drugs (each 500 μM). The g-values for the complexes of CYP144A1 with fragments 4 and 5 are provided in the main paper and in Table S1. A minor species appearing as a shoulder on a larger feature is annotated with a superscript “sh”.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Ferric low-spin g-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g_z</td>
</tr>
<tr>
<td>Ligand-free</td>
<td>2.41</td>
</tr>
<tr>
<td>NMR062</td>
<td>2.45, 2.42&lt;sup&gt;sh&lt;/sup&gt;</td>
</tr>
<tr>
<td>NMR263</td>
<td>2.48&lt;sup&gt;sh&lt;/sup&gt;, 2.44&lt;sup&gt;sh&lt;/sup&gt;, 2.41</td>
</tr>
<tr>
<td>NMR314</td>
<td>2.44&lt;sup&gt;sh&lt;/sup&gt;, 2.41</td>
</tr>
<tr>
<td>NMR642</td>
<td>2.47&lt;sup&gt;sh&lt;/sup&gt;, 2.41</td>
</tr>
<tr>
<td>NMR729</td>
<td>2.48&lt;sup&gt;sh&lt;/sup&gt;, 2.43&lt;sup&gt;sh&lt;/sup&gt;, 2.41</td>
</tr>
<tr>
<td>NMR779</td>
<td>2.44&lt;sup&gt;sh&lt;/sup&gt;, 2.41</td>
</tr>
<tr>
<td>Econazole&lt;sup&gt;5&lt;/sup&gt;</td>
<td>2.43</td>
</tr>
<tr>
<td>Miconazole</td>
<td>2.44, 2.41</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>2.43, 2.41&lt;sup&gt;sh&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Figure S3.20. NanoESI mass spectra of ligand-free CYP144A1 demonstrating the change in charge state with $d_6$-DMSO.

Spectra of CYP144A1 (10 μM) were collected in 200 mM ammonium acetate buffer (pH 7.0), containing either 0% v/v or 5% v/v $d_6$-DMSO. In the absence of $d_6$-DMSO, peaks in the range 3200-4400 m/z were assigned to monomeric (Mw = 46611 ± 75 Da) CYP144A1 (bottom panel). Peaks for the monomeric protein shifted to 4000-5400 m/z when $d_6$-DMSO (5% v/v) was added.

Figure S3.21. DSC traces demonstrating the effect of ligand binding on the thermal stability of CYP144A1.

(a) Monophasic unfolding transition ($T_{m1} = 48.6^\circ C$) observed from the combination of CYP144A1 (10 μM) with compound 1 (200 μM); (b) Biphasic unfolding transition ($T_{m1} = 48.2^\circ C, T_{m2} = 57.7^\circ C$) observed from the combination of CYP144A1 (10 μM) with fragment 4 (200 μM). The melting temperature ($T_m = 47.3^\circ C$) of ligand-free CYP144A1 (5% v/v DMSO) is indicated by a grey dashed line. The first ($T_{m1}$) and second ($T_{m2}$) unfolding transitions are colored blue and red, respectively.
**Table S3.4.** Melting transitions measured for CYP144A1-ligand complexes by DSC.

<table>
<thead>
<tr>
<th>Compound</th>
<th>$T_m1$ (°C)</th>
<th>$T_m2$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand-free CYP144A1</td>
<td>47.26</td>
<td>-</td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>50.55</td>
<td>-</td>
</tr>
<tr>
<td>Econazole</td>
<td>49.68</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>49.07</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>49.78</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48.55</td>
<td></td>
</tr>
<tr>
<td>4PIM</td>
<td>49.34</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>48.16</td>
<td>57.65</td>
</tr>
<tr>
<td>729</td>
<td>49.42</td>
<td>62.4</td>
</tr>
<tr>
<td>422</td>
<td>49.53</td>
<td>62.83</td>
</tr>
<tr>
<td>244</td>
<td>49.94</td>
<td>62.95</td>
</tr>
<tr>
<td>207</td>
<td>49.79</td>
<td>62.53</td>
</tr>
<tr>
<td>131</td>
<td>49.78</td>
<td>-</td>
</tr>
<tr>
<td>062</td>
<td>49.44</td>
<td>-</td>
</tr>
<tr>
<td>263</td>
<td>50.16</td>
<td>-</td>
</tr>
<tr>
<td>003</td>
<td>49.41</td>
<td>-</td>
</tr>
<tr>
<td>034</td>
<td>49.61</td>
<td>-</td>
</tr>
<tr>
<td>074</td>
<td>49.68</td>
<td>-</td>
</tr>
<tr>
<td>166</td>
<td>49.35</td>
<td>61.73</td>
</tr>
<tr>
<td>224</td>
<td>49.54</td>
<td>-</td>
</tr>
<tr>
<td>257</td>
<td>49.4</td>
<td>63.17</td>
</tr>
<tr>
<td>327</td>
<td>48.85</td>
<td>-</td>
</tr>
<tr>
<td>352</td>
<td>47.12</td>
<td>58.6</td>
</tr>
<tr>
<td>642</td>
<td>50.31</td>
<td>-</td>
</tr>
<tr>
<td>251</td>
<td>49.79</td>
<td>-</td>
</tr>
<tr>
<td>779</td>
<td>50.63</td>
<td>-</td>
</tr>
<tr>
<td>241</td>
<td>49.93</td>
<td>-</td>
</tr>
</tbody>
</table>
3.6 Experimental Methods

3.7 General

All fragment screening experiments by UV-Vis spectroscopy and ligand-observed NMR, as well as ligand binding competition experiments by UV-Vis and NMR, nanoESI native mass spectrometry and protein crystallography were performed using the truncated version of the CYP144A1 protein (CYP144A1-TRV, from which the first 30 amino acids, predicted to form a disordered region, are deleted). Ligand-binding studies using EPR and DSC were repeated on both the full-length CYP144A1 protein (CYP144A1-FLV) and CYP144A1-TRV. Both versions of the CYP144A1 enzyme were previously shown to have near-identical optical, electronic, thermal and ligand-binding properties. No differences between the TRV and FLV CYP144A1 forms were observed in the experiments reported here and consequently only data for CYP144-TRV have been included. All ligands were obtained from commercial sources except amino-pyrazole compounds, which were synthesized according to the previously reported procedures.

3.7.1 Bioinformatics, modelling and ligand docking studies

Protein sequence similarity. Protein sequence alignments of the full-length CYP144A1 (CYP144A1-FLV) enzyme were performed using the online NCBI tool BLAST, restricting alignments to proteins in the Swiss-protKB/UniProt databases (https://blast.ncbi.nlm.nih.gov/)

Tertiary structure overlays and residue comparison. Structure overlays of \( Mtb \) CYP144A1 (PDB 5HDI), CYP121A1 (PDB 3G5F), CYP125A1 (PDB 3IVY) and CYP142A1 (2XKR) were performed using the PDBbFOLD online tool accessible at: http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver. Structure overlays were viewed and images were prepared using
the PyMOL Molecular Graphics System, Version 1.3, 2010, Schrödinger, LLC. Amino acid residues within 6 Å of the PEG-ligand that is co-crystallized in the structure of CYP142A1, were compared for structural and chemical similarity between the P450s, using the default similarity index (0-5) generated by PDBeFOLD for calculations.

**Active site properties and druggability.** Proteins structures for *Mtb* CYP144A1 (PDB 5HDI), CYP121A1 (PDB 3G5F), CYP125A1 (PDB 3IVY) and CYP142A1 (PDB 2XKR) were prepared using the inbuilt Protein Preparation Wizard in Maestro v10.0 (Schrödinger LLC, NY). Ionization states were generated to be compatible with metal-binding interactions and the heme-iron was manually adjusted to the ferric (+3) oxidation state. All water molecules except for the axial heme water ligand were removed. The PEG ligand co-crystallized in CYP142A1 (PDB 2XKR) was manually removed and replaced with a water molecule to prevent distortion in the size of the active site being analyzed. Prepared protein structures were submitted to SiteMap, selecting to evaluate a single binding site within 6 Å of the axial heme water ligand. All other default parameters were retained. The calculated D-score was used as an assessment of protein druggability.

**Ligand docking.** Ligands were prepared for docking using the LigPrep, v3.2 and Epik v3.0 functions of Schrödinger suite software (Schrödinger LLC, NY). Duplicate energy minimized (OPLS 2005) protein structures were prepared from the X-ray crystal structure of ligand-free CYP144A1 (PDB 5HDI) using the internal Protein Preparation Wizard in Maestro v10.0. Ionization states were generated to be compatible with metal-binding interactions and the heme-iron was manually adjusted to the ferric (+3) oxidation state. All water molecules were removed from one structure, while the axial heme water ligand was retained in the second protein structure. Docking grids were prepared using the coordinates of the axial heme water ligand as a central point. Ligands were allowed to dock into both grids under a range of scenarios, either employing no constraints, enforcing hydrogen bonding
interactions with the axial heme water ligand, or metal coordination to the ferric iron. Docking images were generated using the PyMOL Molecular Graphics System, Version 1.3, 2010, Schrödinger, LLC.

3.7.2 Expression and purification of CYP144A1

CYP144A1 was expressed and purified using a pET15b/Rv1777 vector transformed in E. coli C41 (DE3) cells according to the previously reported procedure. Protein identification was confirmed by SDS-PAGE and ESI mass spectrometry. Protein purity was assessed by SDS-PAGE, amino acid analysis and UV-Vis spectroscopy, with samples having an A420/A280 ratio of ≥1.5, and/or >90% agreement with expected amino acid composition being considered as pure.

3.7.3 Optical titrations

Optical titrations to determine $K_D$ values were carried out on a Cary 60 UV-visible spectrophotometer (Varian, UK) according to a previously described procedure. All titrations were performed using 1-cm pathlength quartz cuvettes. Azole antifungal compounds, novel compounds 1-3 and fragment 4 were prepared as stock solutions (typically 0.1-100 mM) in DMSO and added directly as 0.2 μL aliquots to cuvettes containing either a solution of CYP144A1 (~4-8 μM) in 100 mM KPi (pH 7.0), containing 10 mM KCl; or, to buffer alone. DMSO concentrations did not exceed 1% v/v of the final titration volume (1 mL) and the absorbance spectrum of CYP144A1 was not affected within this range. Spectra were recorded continuously between 800-250 nm at 25 °C. Spectra collected from the buffer control cuvette were subtracted from protein spectra to remove any optical interference from small-molecule absorbance. Difference spectra were generated by subtraction of the initial ligand-free protein spectrum from each successive ligand-bound protein spectrum. The maximum change in absorbance for each difference spectrum was then plotted against ligand
concentration. Data were fitted using either a modified version of the Morrison equation for tight-binding ligands (Eq. 3.1), the Hill function for cooperative binding (Eq. 3.2), or using the Michaelis-Menten equation (Eq. 3.3), depending on the affinity of the ligand and type of binding curve observed.

Equation 3.1 \[ A_{\text{obs}} = \frac{(A_{\text{max}}/2Et) \times ((L + Et + K_D) - ((L + Et + K_D)^2 - (4 \times L \times Et))^{0.5})}{(L + Et + K_D)^2 - (4 \times L \times Et)} \]

Equation 3.2 \[ A_{\text{obs}} = \frac{A_{\text{max}} \times [L]^n}{(K_D + [L])^n} \]

Equation 3.3 \[ A_{\text{obs}} = \frac{A_{\text{max}} \times [L]}{(K_D + [L])} \]

In Equation (3.1-3.2), \( A_{\text{obs}} \) represents the observed difference in Soret absorption of the P450 at each addition of ligand, \( A_{\text{max}} \) is maximal absorption difference at ligand saturation, Et is the total amount of enzyme, L is the ligand or substrate concentration, n is the extent of cooperativity (potentially providing information on the number of binding sites for a given ligand per enzyme), \( E_t \) is the total P450 concentration, and LnEt is the concentration of the ligand-enzyme complex. In Equation (3.3), \( A_{\text{obs}} \) is the absorbance change observed at ligand concentration [L]. \( K_D \) is the dissociation constant for the P450-ligand complex. All data fitting and analysis were performed using Origin software (OriginLab, Northampton, MA). All titrations were repeated in triplicate and data from representative experiments are reported.

3.7.4 UV-Visible spectroscopy, ligand screening and competition assays

Ligand screening of azole drugs, aminopyrazole compounds, fragments and \( Mtb \) metabolites (see Supporting Information) was performed using a CARY400 UV-Vis spectrophotometer (Varian, UK). Fragments and ligands were prepared as \( d_6 \)-DMSO stock solutions (1-100 mM) and added directly to 1-cm pathlength quartz cuvettes containing solutions of either
CYP144A1 (5 μM) in 50 mM Tris-HCl (pH 7.5), containing 100 mM KCl buffer, or buffer alone. Concentrations of \( d_6\)-DMSO did not exceed 1% v/v. Spectra were recorded continuously between 800-250 nm at 25 °C. Any interference from the inherent absorbance of small molecules was removed by subtracting spectra from buffer control cuvettes from samples containing protein. The difference in maximum wavelength of the CYP144A1 Soret band (\( \Delta \lambda_{\text{max}} \)) observed in the presence of ligands compared to that obtained a presence of 1% v/v \( d_6\)-DMSO (\( \lambda_{\text{max}} = 420.5 \) nm) was used to identify compounds as Type I or Type II heme-binding ligands. Changes in the \( \lambda_{\text{max}} \) of the Soret band of less than ± 1 nm were considered within experimental error.

Competition assays to detect the binding interactions of CYP144A1 with non-heme binding ligands were performed as above, except that a Type II compound (econazole 25-50 μM, miconazole 12.5 μM or compound 2 50 μM) was subsequently added to assay solutions containing CYP144A1 (or buffer alone) and the non-heme binding ligand. Concentrations of \( d_6\)-DMSO did not exceed 1% v/v of the total assay volume. A decrease or increase in the \( \Delta \lambda_{\text{max}} \) caused by the known Type II compound in the presence of the non-heme binding ligands, compared to that observed for the Type II ligand alone, was used to categorize competitive or cooperative ligand binding interactions, respectively. Spectra of ligands collected in buffer were subtracted from CYP144A1-containing spectra to control for the effect of small molecule optical interference. All UV-Vis spectra were generated using Origin software (OriginLab, Northampton, MA) and data were processed using Microsoft Excel (Microsoft Office, 2010) and GraphPad Prism v5.01 (GraphPad Software, San Diego, USA).
3.7.5 Isothermal titration calorimetry

ITC binding isotherms were recorded on a MicroCal ITC200 instrument (Malvern Instruments). Titrations were conducted at 25 °C by injecting aliquots (2.0 μL) of ligand solutions (1 mM) into protein samples (60.9 μM), both of which had been diluted in 50 mM Tris-HCl buffer (pH 7.5), containing 100 mM KCl, and adjusted to give a final concentration of 10% v/v d₆-DMSO. Titrations were typically 20 injections at 90-second intervals. Small background heats from dilution of the ligand were subtracted after performing a second control titration of ligand samples into buffer without protein. Binding isotherms were integrated to give the enthalpy change of each injection and plotted against the molar ratio of ligand added to the sample cell. Titrations were fitted using a one-sites binding model, using Origin Analysis Software provide by the manufacturer. The binding stoichiometry was set at N=1, or to N=2 to reflect binding stoichiometries observed in nanoESI native mass spectra of compound 3.

3.7.6 Ligand-observed NMR

Fragment screening of an 81 compound Rule-of-3 compliant library was conducted as cocktails of 3 fragments by ¹H-NMR employing a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence. Solutions (200 μL) of fragments (each 200 μM) were prepared in 50 mM Tris-HCl buffer (pH 7.5), containing 100 mM KCl and a final d₆-DMSO concentration of 0.6% v/v. d₄-3-(Trimethylsilyl)propionic acid (20 μM) in D₂O (10 % v/v) was added as an internal standard. Samples were prepared both with and without CYP144A1 (20 μM) and then loaded into 3 mm diameter capillaries for analysis inside 524-PP-8 NMR tubes (Wilmad-LabGlass, NJ, USA). ¹H-NMR experiments were performed on a Bruker Avance 500 MHz NMR, equipped with a TXI cryoprobe and a BACS-60 autosampler. All experiments were performed at 278 K employing a relaxation delay of 200 ms. Water signal suppression was
achieved using a W5 Watergate gradient spin-echo pulse sequence. Spectra were processed using TopSpin 3.5 software (Bruker, Coventry, UK) and scaled relative to the $d_4$-TSP peak intensity.

Additional screening of a 50 compound fragment library was performed as singlets using a combination of STD, CPMG and WaterLOGSY NMR experiments. Samples (200 μL) of individual fragments (1 mM) were prepared in 100 mM MES-NaOH buffer (pH 6.0), containing 50 mM NaCl, 2% v/v $d_4$-MeOD and a final $d_6$-DMSO concentration of 2% v/v. $d_4$-3-(Trimethylsilyl)propionic acid (20 μM) in D$_2$O (10 % v/v) was added as an internal standard. Samples were prepared both with and without CYP144A1 (20 μM) and then loaded into 3 mm diameter capillaries for analysis in 3 mm diameter capillaries inside 528-PP-8 NMR tubes (Wilmad-LabGlass, NJ, USA). Spectra were acquired on a Bruker Avance 500 MHz NMR, equipped with a TXI cryoprobe and a BACS-60 autosampler using pulse sequences provide by Dr Glyn Williams (Astex Therapeutics, UK). STD experiments utilized a 40 ms selective Gaussian 180° pulse at a frequency alternating between ‘on-resonance’ (0.4 ppm) and ‘off-resonance’ (40 ppm) after every scan. WaterLOGSY experiments employed a 20 ms selective Gaussian 180° shaped pulse at the water signal frequency and an NOE mixing time of 1 s. CPMG experiments employed a relaxation delay of 200 ms. Water signal suppression was achieved using a W5 Watergate gradient spin-echo pulse sequence. All experiments were performed at 278 K. Spectra were processed using TopSpin 3.5 software (Bruker, Coventry, UK), phasing WaterLOGSY spectra relative to the non-binding buffer and scaling CPMG spectra relative to the $d_4$-TSP peak intensity.
3.7.7 **EPR spectroscopy**

EPR spectra of both ligand-free CYP144A1 (200 μM), and CYP144A1 (200 μM) bound to ligands (500 μM) were recorded using a Bruker ER-300D series electromagnet and microwave source interfaced with a Bruker EMX control unit, and fitted with an ESR-9 liquid helium flow cryostat (Oxford Instruments), and a dual-mode microwave cavity from Bruker (ER-4116DM). Spectra were recorded at a temperature of 10 K, a microwave power of 2.08 mW, and an amplitude of 1 mT. Samples were prepared 100 mM KPi (pH 7.0), 10 mM KCl, from stock solutions of ligands in DMSO. Data analyses were performed using Origin Software.

3.7.8 **Differential scanning calorimetry**

DSC experiments were performed using a Microcal VP-DSC instrument (Malvern Instruments, Malvern UK). Samples were run using a continuous temperature gradient between 20-80 °C at a 90 °C/h scan rate after an initial 10 min pre-scan thermostat. Samples of ligand-free CYP144A1 (10 μM) and CYP144A1 (10 μM) bound to ligands (200 μM) were prepared in 10 mM KPi (pH 7.0), containing 100 mM NaCl. Data analyses were performed using Origin Software.

3.7.9 **Nano-ESI native mass spectrometry**

Protein stock solutions (10 - 40 μM) were prepared by dilution of purified proteins (500 - 1000 μM) in 200 μM ammonium acetate buffer, pH 7.0. Samples were buffer exchanged by size exclusion chromatography using Micro Biospin 6 columns, molecular weight cut-off 6 kDa (BioRad, Hemel Hempstead, UK). Ligands were prepared as stock solutions in d6-DMSO at 0.2 - 2 mM concentrations. Ligand-protein samples were prepared by diluting
protein stocks (10 μL) and ligand stocks (1 μL) with ammonium acetate buffer (9 μL) to give final concentrations of 10 μM CYP144A1, 10-50 μM ligand and 5% v/v d₆-DMSO. Mass spectra were recorded on a Synapt HDMS instrument (Waters UK Ltd., Manchester, U.K.). Capillaries for nano-ESI were purchased from ThermoFisher, Hemel Hempstead, UK. Capillary tips were cut under a stereo microscope to give inner diameters of 1 - 5 μm and then loaded with 2.5 μL of sample solutions. Given below are the general instrumental conditions used to acquire the reported spectra. However, parameters were recorded and varied over the course of each experiment to observe the strength of protein-ligand complexes under different ionizing strengths. All measurements were carried out in a positive ion mode with ion source temperature of 20 °C. A capillary voltage of 1.5 kV, cone voltage of 40 V and extraction cone voltage of 4.8 V was applied to perform nanoESI. All reported spectra were collected with a trap collision energy 30-60 V, transfer collision energy 12-30 V, IMS pressure 5.02 × 10⁻¹ mbar, TOF analyzer pressure 1.17 × 10⁻⁶ mbar. External calibration of the spectra was achieved using cesium iodide at 100 mg mL⁻¹ in water. Data acquisition and processing were performed using Micromass MassLynx v4.1. Mass differences resulting from ligand binding were calculated from the unbound protein peak internal to each spectrum. The unbound protein peak was compared to the relevant 5% v/v d₆-DMSO control spectra for consistency. Mass differences were divided by the molecular weight of the ligand to calculate binding stoichiometry.

Supplementary Methods S1. Biochemical turnover analysis of compound 1.

Biochemical turnover of the Type I ligand compound 1 was assessed according to the previously published procedure. Assays were prepared in 50 mM KPi (pH 7.0), employing 1 μM compound 1, 0.5 μM CYP144A1, 10 μM spinach ferredoxin, 10 μM spinach ferredoxin reductase. 1 mM NADPH, and an internal NADPH regenerating system of catalytic amounts
of glucose-6-phosphate/glucose-6-phosphate dehydrogenase was used. Control assays omitting either compound 1 or CYP144A1 were also run. No oxidation products were detected from assays by LC-MS analysis.

**Supplementary Methods S2. Potential substrates screened against CYP144A1 using UV-Vis spectroscopy.** Select molecules, including sterol and lipid substrates present in *Mtb*, were screened against CYP144A1 by UV-Vis titration according to the experimental protocol detailed in the main paper. None of the compounds were found to perturb the λ<sub>max</sub> of the Soret band.\(^{30}\) None of the tested compounds (cholesterol, 4-cholesten-3-one, testosterone, vitamin D, phytanic acid, biotin, vitamin K, myristic acid, pentadecanoic acid, palmitoleic acid, arachidonic acid, cis-parinaric acid, arachidic acid, tridecanoic acid, D-sphingosine, D-glucose, D-mannose, D-fructose, D-galactose, maltose and starch) produced any significant shift in the CYP144A1 Soret band that could be associated with the binding of a substrate-like molecule.
References


49. SiteMap.
59. Ouellet, H., Podust, L. M. & Ortiz de Montellano, P. R. *Mycobacterium tuberculosis*
80. Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W. One-dimensional relaxation- and
diffusion-edited NMR methods for screening compounds that bind to macromolecules. 


4 Chapter Four

Biochemical and structural characterization of Mycobacterium tuberculosis CYP126A1

Jude Chenge†, Duyet V. Le†, Shalini Swami, Kirsty J. McLean, Madeline E. Kavanagh, Anthony Coyne, Stephen E. J. Rigby, Myles R. Cheesman, Hazel M. Girvan, Colin W. Levy, Jens P. von Kries, Chris Abell, David Leys, Andrew W. Munro*

1Manchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. 2Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom. 3School of Chemistry, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom. 4Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-Strasse 10, 13125, Berlin, Germany

Running title: Novel structural properties of M. tuberculosis CYP126A1

*To whom correspondence should be addressed: A. W. Munro Tel: +44 161 3065151; Fax: +44 161 3068918; E-mail: Andrew.Munro@Manchester.ac.uk. †These authors contributed equally to the research presented in this manuscript.

Keywords: Cytochrome P450, CYP126A1, Electron Paramagnetic Resonance, enzyme structure, high-throughput screening, mass spectrometry, Mycobacterium tuberculosis, redox potentiometry
4.1 Abstract

*Mycobacterium tuberculosis* CYP126A1 is representative of a P450 enzyme family widely distributed in mycobacteria and other bacteria. CYP126A1 was purified, revealing a propensity to dimerize in solution. CYP126A1 is inhibited strongly by azole drugs with imidazole groups, but not by selected triazole group-containing azoles. High-throughput screening (HTS) identified CYP126A1 substrate and inhibitor molecules, using characteristic type I (blue) and type II (red) P450 Soret spectral shifts, respectively, to recognize and classify hits. Polycyclics dominated the hits, including type I nitroaromatic-containing compounds. Spectroelectrochemical measurements on ligand-free and type I nitroaromatic compound 1-bound forms of CYP126A1 revealed a 155 mV heme iron potential increase with compound 1, consistent with P450 substrate behaviour. Crystal structures were solved for native CYP126A1 and its complexes with ketoconazole, compound 1 and the type II hit compound 7. Structural data reveal compound 1 to “moonlight” as a substrate by displacing the CYP126A1 heme distal water using its nitroaromatic group and inducing high-spin heme iron formation, but acting instead as an inhibitor. The compound 1- and 7-bound species, and native CYP126A1 form crystallographic dimers, with interface structure influenced by ligand-binding. Ketoconazole binds orthogonally to the paths taken by compounds 1 and 7 within a crystallographic monomer, causing structural changes that disrupt the dimer interface. CYP126A1 has a relatively polar active site and a different specificity profile from its most closely related, sterol-binding P450s in *M. tuberculosis*. Novel HTS substrates and inhibitors for CYP126A1 provide insights into its molecular preference as well as providing probes of its catalytic function.
4.2 Introduction

The human pathogen *Mycobacterium tuberculosis* (*Mtb*) remains a major global cause of mortality as the infectious bacterium that causes tuberculosis (TB)\(^1\). Recent data from the World Health Organization indicate that TB is the leading cause of human death worldwide among infectious diseases\(^2\). Mortality rate in TB victims may be increased through their co-infection with the human immunodeficiency virus (HIV). Moreover, the development of *Mtb* strains resistant to leading drugs compounds the problem and, once diagnosed, usually results in extended treatment times\(^2\). The more recent advent of multidrug (MDR) and extensively drug resistant (XDR) strains poses further major issues in treating TB. MDR and XDR *Mtb* strains are resistant to, respectively, at least the two leading TB drugs (rifampicin and isoniazid), or to both these drugs as well as to any one of the quinolone drugs, and to at least one of the second-line injectable TB drugs amikacin, capreomycin and kanamycin\(^3,4\). Consequently, there is increased need for development of new TB drugs with novel modes of action; a need which has been partially met recently by the development of drugs such as delamanid (which inhibits cell wall mycolic acid synthesis) and bedaquiline (a *Mtb* ATPase proton pump inhibitor), both of which have been authorized for use in MDR TB treatment\(^5\).

One of the revelations from the first genome sequence of *Mycobacterium tuberculosis* (that for the virulent H37Rv strain) was that twenty different cytochrome P450 (CYP or P450) enzymes were encoded\(^1\). The unexpectedly large number of P450s suggested important functions for these enzymes, and research has uncovered key roles in metabolism of host cholesterol/cholest-4-en-3-one (CYP125A1 and CYP142A1) and branched chain lipids (CYP124A1), oxidative tailoring of cyclic dipeptides (CYP121A1), hydroxylation of respiratory menaquinone (CYP128A1), and sterol demethylation (CYP51B1)\(^6-14\). The *CYP125A1* and *CYP142A1* genes are both located in a gene regulon associated with host
cholesterol metabolism and uptake, and were shown to be important for the oxidative metabolism of the steroid side chain to initiate catabolism of the molecule by Mtb in the macrophage\textsuperscript{7,8,15}. CYP128A1 has yet to be expressed and characterized, but is implicated in the synthesis of a virulence-associated sulfolipid (S881) through hydroxylating menaquinone 9, (MK9H\textsubscript{2}), which is considered the sole quinol electron carrier in the Mtb respiratory chain. CYP128A1 is thought to catalyze terminal hydroxylation of MK9H\textsubscript{2} to enable sulfation at the hydroxyl group by the sulfotransferase Stf3 encoded by the gene Rv2267c immediately downstream of CYP128A1\textsuperscript{1,12}.

The first Mtb P450 enzyme to be structurally and biochemically characterized was CYP51B1, the first member of the CYP51 (sterol demethylase) gene family identified in a prokaryote\textsuperscript{13,16,17}. The CYP51B1 Fe\textsuperscript{II}-CO complex was found to be unstable, and to collapse over time from the cysteine thiolate-coordinated P450 form to the thiol-coordinated P420 state. However, the thiolate-coordinated form could be stabilized by the binding of estriol\textsuperscript{14}, and later studies on both the Mtb cholesterol hydroxylase CYP142A1 and the Sorangium cellulosum epothilone C/D epoxidase EpoK showed that binding of substrates (cholest-4-en-3-one and epothilone D, respectively) could regenerate the P450 state when added to the substrate-free Fe\textsuperscript{II}-CO P420 form\textsuperscript{8,18}. Importantly, the soluble CYP51B1 enzyme catalyzes oxidative 14α-demethylation of lanosterol, 24,25-dihydrolanosterol and the plant sterol obtusifoliol, and was shown to bind azole drugs used clinically for inhibition of fungal CYP51 enzymes\textsuperscript{13,17}. These findings inspired research to examine the potency of azole drugs against mycobacteria. In vitro studies revealed that a number of azoles had good MIC values against M. smegmatis, particularly econazole (<0.1 µg/ml) and miconazole (0.1 µg/ml). The same two azoles were found to be most effective against Mtb H37Rv, albeit with higher MIC values (8 µg/ml for both drugs)\textsuperscript{19,20}. This is possibly due to lower azole penetration into the Mtb cells, or e.g. efflux of the drugs from the cells\textsuperscript{21}. Studies in mice also showed that
econazole reduced bacterial burden by 90% in lungs and spleen, and was also effective against MDR \textit{Mtb} strains\textsuperscript{22,23}. Thus, regardless of issues surrounding cross-reactivity of azole drugs with human P450s, various azoles are clearly potent inhibitors of several \textit{Mtb} P450s, as demonstrated from \textit{in vitro} and structural studies, and are important tools in the characterization of these enzymes\textsuperscript{13,24}.

Several of the \textit{Mtb} P450s remain structurally uncharacterized. Among these is CYP126A1, a P450 with \textasciitilde35\% amino acid identity to the cholesterol oxidizing \textit{Mtb} CYP142A1 and CYP125A1. The \textit{CYP126A1} (\textit{Rv0778}) gene is located in a genomic region close to genes encoding the sterol demethylase P450 CYP51B1 (\textit{Rv0764c}) and the uncharacterized CYP123A1 (\textit{Rv0766c}), which was predicted to be associated with the \textit{Mtb} H37Rv cell membrane using 2D LC-MS analysis\textsuperscript{25}. It is also located between genes of the purine synthesis pathway (\textit{purB}, \textit{Rv0777}; \textit{purC}, \textit{Rv0780}; and \textit{purD}, \textit{Rv0772}), although there is no evidence for its involvement in this pathway. However, \textit{CYP126A1} is highly conserved across pathogenic and non-pathogenic \textit{Mycobacterium} species. In recent research, we introduced the concept of “biofragment” screening, in which a library of biologically relevant fragments was assembled based on the structures of known substrates of diverse P450 enzymes molecules. Members of this library had greater 3-dimensional character compared to those of typical fragment libraries, and resulted in a higher CYP126A1 fragment hit rate compared to previous fragment screening studies on \textit{Mtb} CYP121A1 and CYP125A1\textsuperscript{26,27}.

In this paper, we present biochemical, spectroscopic and structural data for CYP126A1, including studies of its binding to novel substrate and inhibitor molecules identified from a large compound library screen. The crystal structures of CYP126A1 in complex with the azole drug ketoconazole and with inhibitor and substrate-like molecules from the screen are presented, illustrating the capacity of CYP126A1 to bind bulky, polycyclic compounds. CYP126A1 was also found to form a crystallographic dimer in which both open and closed
forms of the protein (with respect to P450 conformation and active site access) are observed. These data provide important insights into the structural properties and the molecular selectivity of a new *Mtb* P450 enzyme, highlighting its capacity to bind bulky polycyclic compounds, and the ability of molecules with nitroaromatic groups to “moonlight” as substrates through inducing formation of high-spin (HS) heme iron in CYP126A1. The crystal structures of CYP126A1 are formed in both dimeric and monomeric states, with structures determined in complex with the azole drug ketoconazole and with compounds identified from a high-throughput screening (HTS) process that either inhibit CYP126A1 by coordinating the heme iron, or produce a substrate-like heme spectral shift on binding. This is the first report of structural data for a widely conserved *Mtb* P450 enzyme, and data provide important insights into CYP126A1 molecular selectivity and structural adaptation to the binding of large compounds.

4.3 Results

4.3.1 Expression and Purification of CYP126A1

CYP126A1 was purified from *E. coli* expression cell extracts using Ni-IDA chromatography followed by ion exchange chromatography (IEC) using Q-Sepharose. Using the Rz (A$_{418}$/A$_{280}$) value as a guide to relative CYP126A1 protein purity, it was found that a typical Ni-IDA step resulted in an ~45-50-fold purification fold, with a further 3-4-fold purification achieve using Q-Sepharose IEC. At this stage, purified CYP126A1 was analyzed by SDS-PAGE gel electrophoresis and found to be highly purified, with an apparent molecular mass of ~47 kDa (Figure 4.1A, inset). For protein destined for crystallography and P450 structural studies, either a second Q-Sepharose purification step under the same conditions or a gel filtration step using Sephacryl S-200 SEC (both on an AKTA purifier) was used to ensure near-homogeneity of the CYP126A1 sample.
Native mass spectrometry analysis of the purified CYP126A1 produced a mass of 47986 Da that is consistent with that predicted for the His$_6$-tagged form of the heme-free apoprotein. A small component with a mass of 48601 Da is likely due to the heme-bound form of CYP126A1 (heme $b$ has a mass of 616.5 Da) (Figure 4.1A, main panel). The determined mass of the His$_6$-tagged CYP126A1 matches exactly that predicted from the P450 amino acid sequence after removal of the N-terminal methionine (47986.12 Da). Analysis of CYP126A1 by nanoESI mass spectrometry revealed the enzyme to be predominantly monomeric (Figure 4.1B) at a relatively low concentration (10 μM, m/z = 48687), albeit with a small proportion of dimer (m/z = 97352). Similar analysis of CYP126A1 at a higher concentration (20 μM) confirmed propensity of CYP126A1 to self-associate, with small amounts of P450 trimer and tetramer also detected (m/z = 146052 and 194752, respectively).
Figure 4.1. Purification and mass spectrometry of CYP126A1.

Panel A. The main image shows the native mass spectrum of CYP126A1. The major feature at 47986 Da is consistent with the predicted mass of the His-tagged CYP126A1 apoprotein following removal of the initiator methionine (47986.12 Da). The minor feature (48601 Da) is consistent with the mass for the heme-bound CYP126A1 holoprotein. The inset shows an SDS-PAGE gel with molecular weight markers (sizes in kDa) in lane 1, and a sample of purified CYP126A1 in lane 2. Panel B shows nanoESI mass spectrometry data for CYP126A1 at concentrations of 10 μM (upper spectrum) and 20 μM (lower spectrum). For the 10 μM sample, differently charge species are labelled A (for monomer) and B (for a minor dimer species). For the 20 μM sample, signals are labelled A-D for monomer through to tetramer forms. Apparent masses of these different species are indicated.

4.3.2 UV-Visible Spectroscopic Properties of CYP126A1 and Identification of Novel CYP126A1 Ligands and their Binding Properties

The UV-visible spectrum of the oxidized (FeIII) CYP126A1 has major heme absorbance features at approximately 418.5 nm (Soret), 568.5 nm (alpha) and 539 nm (beta). Reduction of CYP126A1 with sodium dithionite results in a Soret band shift to ~412 nm with a Q-band feature at ~544 nm. The binding of carbon monoxide to reduced CYP126A1 results in a FeII-
CO complex with a Soret peak at 448.5 nm and a Q-band feature at 549 nm (Figure 4.2). The spectral features of these ferrous and Fe^{II}-CO forms of CYP126A1 are consistent with the retention of cysteine thiolate as the proximal ligand to the P450 heme iron. However, the CYP126A1 Fe^{II}-CO complex converts slowly to the cysteine thiol-coordinated P420 state (with Soret maximum at ~423 nm) after ~15 minutes, indicating a propensity for protonation of the cysteine thiolate in this protein state, as is also observed in various other P450s, particularly in the absence of substrate\textsuperscript{8,18} (Figure 4.2).

![Figure 4.2. UV-visible spectral properties of CYP126A1.](image)

The ferric form of CYP126A1 (6.5 μM) has absorbance band maxima at 418.5 nm (Soret), 568.5 nm (alpha) and 539 nm (beta) (thick, solid line). Reduction with sodium dithionite results in a Soret shift to ~412 nm with a Q-band feature at ~544 nm (dashed line). Binding of carbon monoxide results in a Soret shift to 448.5 nm with a Q-band feature at ~549 nm, consistent with retention of cysteine thiolate coordination. Over a period of ~10-15 minutes, this P450 form converts to a P420 (thiol-coordinated) form with Soret maximum at 423 nm.

Consistent with other \textit{Mtb} P450s characterized to date, CYP126A1 binds a range of antifungal compounds containing imidazole heterocycles\textsuperscript{7,10,13,14,28}. These compounds all produced Type II (red) shifts in the wavelength maximum ($\lambda_{\text{max}}$) of the CYP126A1 Soret band. For example, ketoconazole induces a Soret band $\lambda_{\text{max}}$ shift from 418.5 nm to 423.5 nm at apparent saturation, in addition to inducing smaller absorbance changes in the heme alpha
and beta band peaks (minor shifts to ~569.5 nm and 540.5 nm, respectively, with notably decreased intensity of the alpha band) (Figure 4.3A). These spectral changes are consistent with ligation of the heme iron by a heterocyclic nitrogen atom. The CYP126A1 $K_d$ for ketoconazole was $0.34 \pm 0.07 \, \mu M$, while imidazole itself has a substantially weaker $K_d$ of $2.59 \pm 0.06 \, mM$. However, a common feature in the spectral titrations of CYP126A1 with imidazole group-containing azole drugs (ketoconazole, miconazole, econazole, clotrimazole and imidazole itself) is the decrease in intensity of the red-shifted Soret band by ~15-20% as the titration progresses to completion (Figure 4.3A). The $K_d$ values for the binding of these compounds to CYP126A1 are given in Table 4.1, and compared to previously reported values for their binding to other $Mtb$ P450s. It was also notable that none of the tested azole drugs containing triazole groups (fluconazole, voriconazole and itraconazole) gave any convincing CYP126A1 Soret shifts, suggesting negligible or very weak binding to this P450.

A high throughput screen (HTS) was also used to identify novel compounds that bind to CYP126A1. A series of ligands, compounds 1-9 (Figure 4.4), were selected using UV-Vis spectroscopy on the basis of their ability to induce either a substrate-like Type I Soret shift (a blue shift, indicative of conversion of the low-spin [LS] CYP126A1 towards HS) or a type II Soret shift (a red shift, usually indicative of the coordination of the heme iron by a nitrogen or other atom from the compound). Compounds 1-6 each produced Type I CYP126A1 P450 spectra, with the Soret band shifting from 418.5 nm to ~395 nm. This is accompanied by a decrease in absorbance in the alpha/beta band region from ~530-585 nm, along with an increase in absorbance at ~650 nm characteristic for a cysteine thiolate-to-HS ferric heme iron charge transfer (CT) complex. Figure 4.3B shows data for a spectral titration of CYP126A1 with compound 1, with data fitted to give a $K_d$ of $5.62 \pm 0.67 \, \mu M$. Compounds 1-6 are all polycyclic, and compounds 1 and 2 both have a nitro-aromatic group at the end of the molecules. Affinity for compounds 1-6 varies in the range from ~3-300 $\mu M$ (Table 4.2).
**Compounds 7-9** induce Type II spectra and all contain groups that can ligate to ferric heme iron through nitrogen atoms (e.g. imidazole, aromatic amine and quinolone groups), with CYP126A1 $K_d$ values from ~4-16 $\mu$M (Table 2). These compounds all induce heme spectral shifts similar to those seen with the imidazole antifungal drugs (Figure 4.3).

**Figure 4.3.** Binding of the azole inhibitor ketoconazole and HTS compound 1 to CYP126A1.

**Panel A** shows a CYP126A1 (4.5 $\mu$M) titration with the azole drug ketoconazole. Ketoconazole coordinates the P450 ferric heme iron and induces a type II red shift of the heme Soret band from 418.5 nm to 424 nm at near-saturation. Isosbestic points are located at ~373 nm, 502 nm and 549 nm. The inset shows a plot of the maximal ketoconazole-induced heme absorption change ($\Delta A_{433}$ minus $\Delta A_{414}$) versus ligand concentration, with data fitted using a tight-binding (Morrison) quadratic function to give a $K_d$ of 0.34 ± 0.07 $\mu$M$^{29}$. In both panels, arrows indicate directions of heme absorbance change at different wavelengths during the progression of the titrations. Structures of these molecules and other CYP126A1 ligands identified in this study are shown in Figure 4.4.

**Panel B** shows a UV-visible absorbance titration of CYP126A1 (4.6 $\mu$M) with compound 1 (N-isopropyl-N-((3-(4-methoxyphenyl)-1,2,4-oxadiazol-5-yl)methyl)-2-(4-nitrophenyl)acetamide). Compound 1 induces a P450 substrate-like type I blue shift of the heme Soret band from 418.5 nm (ligand-free) to 396 nm (near-saturated with compound 1). Isosbestic points in the titration are located at ~ 405 nm, 462 nm and 524 nm. The inset shows a plot of the maximal compound 1-induced heme absorption change ($\Delta A_{388}$ minus $\Delta A_{419}$) versus ligand concentration, with data fitted using a hyperbolic (Michaelis-Menten) function to give a $K_d$ of 18.3 ± 1.3 $\mu$M.
Several ligands and potential substrates for CYP126A1 were identified using HTS methods. The figure (i) imidazole and imidazole-based azole drug inhibitors that coordinate the CYP126A1 heme iron, and (ii) compounds identified by HTS that induce type I (compounds 1-6, inducing HS heme iron accumulation) or type II (compounds 7-9, coordinating the P450 heme iron) spectral shifts. **Compounds 1-9** are as follows: **compound 1**: \(N\)-isopropyl-\(N\-((3-(4-methoxyphenyl)-1,2,4-oxadiazol-5-yl)methyl)-2-(4-nitrophenyl)acetamide; **compound 2**: 4-nitro-\(N\-(7-oxo-6b,7,8,9,10,10a\)-hexahydropyran[1,2-b]benzofuran-5-yl)benzenesulfonamide; **compound 3**: \(N\)-(p-tolyl)phenazine-1-carboxamide; **compound 4**: 3-(4-(benzo[d][1,3]dioxol-5-ylmethyl)piperazin-1-yl)-1-(4-chlorobenzyl)pyrrolidine-2,5-dione; **compound 5**: 5-chloro-\(N\)-(5-methyl-3H-1\(\lambda^2\),3,4-thiadiazol-2-yl)-2-(3-methylbenzyl)thio)pyrimidine-4-carboxamide; **compound 6**: 5-chloro-\(N\)-(6-fluorobenzo[d]thiazol-2-yl)-2-(2-fluorobenzyl)thio)pyrimidine-4-carboxamide; **compound 7**: 1-\(N\-(3-(1H-imidazol-1-yl)propyl)-3-(3,5,7\)-adamantan-1-yl)urea; **compound 8**: 2-(4-(1H-benzo[d]imidazol-2-yl)phenyl)-1-methyl-1\(H\)-benzo[d]imidazol-5-amine; and **compound 9**: 2-(2-chloroquinolin-3-yl)thiazolidine-4-carboxylic acid.
Table 4.1. Binding constants for azole drugs with CYP126A1 and other *M. tuberculosis* cytochrome P450 enzymes.

The dissociation constants ($K_d$ values) for the binding of imidazole and several azole drugs to CYP126A1 and to various other *Mtb* cytochrome P450 enzymes were determined by UV-visible absorbance titration. Data for ligand binding are taken from the following references: CYP126A1 (this work), CYP144A1, CYP125A, CYP142A1, CYP121A1, CYP130A1, and CYP51B1. NB indicates that no significant P450 spectral perturbation was observed on addition of the relevant azole drug. ND indicates that there are no published data for the interaction of the relevant azole drug with the indicated P450 enzyme.

<table>
<thead>
<tr>
<th>Azole inhibitor</th>
<th>CYP126A1 ($K_d$ (µM))</th>
<th>CYP144A1 ($K_d$ (µM))</th>
<th>CYP125A1 ($K_d$ (µM))</th>
<th>CYP142A1 ($K_d$ (µM))</th>
<th>CYP121A1 ($K_d$ (µM))</th>
<th>CYP130A1 ($K_d$ (µM))</th>
<th>CYP51B1 ($K_d$ (µM))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Econazole</td>
<td>4.0 ± 0.6</td>
<td>0.78 ± 0.29</td>
<td>11.7 ± 0.7</td>
<td>4.6 ± 0.2</td>
<td>0.024 ± 0.006</td>
<td>1.93 ± 0.03</td>
<td>0.31 ± 0.04</td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>5.9 ± 0.4</td>
<td>0.37 ± 0.08</td>
<td>5.3 ± 0.6</td>
<td>3.8 ± 0.9</td>
<td>0.073 ± 0.008</td>
<td>13.3 ± 0.6</td>
<td>0.18 ± 0.02</td>
</tr>
<tr>
<td>Miconazole</td>
<td>1.3 ± 0.2</td>
<td>0.98 ± 0.22</td>
<td>4.6 ± 0.4</td>
<td>4.0 ± 0.5</td>
<td>0.136 ± 0.021</td>
<td>1.70 ± 0.21</td>
<td>0.20 ± 0.04</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>0.34 ± 0.07</td>
<td>1.34 ± 0.5</td>
<td>27.1 ± 0.9</td>
<td>21 ± 4</td>
<td>3.41 ± 0.31</td>
<td>48.0 ± 1.3</td>
<td>3.57 ± 0.25</td>
</tr>
<tr>
<td>Fluconazole</td>
<td>NB</td>
<td>&gt;10,000</td>
<td>43.2 ± 0.8</td>
<td>860 ± 108</td>
<td>8.61 ± 0.21</td>
<td>ND</td>
<td>5.82 ± 0.12</td>
</tr>
<tr>
<td>Voriconazole</td>
<td>NB</td>
<td>6510 ± 450</td>
<td>NB</td>
<td>16.3 ± 2.1</td>
<td>ND</td>
<td>2.10 ± 0.16</td>
<td></td>
</tr>
<tr>
<td>4-Phenylimidazole</td>
<td>NB</td>
<td>280 ± 18</td>
<td>216 ± 5</td>
<td>12.0 ± 1.5</td>
<td>32.3 ± 2.2</td>
<td>ND</td>
<td>452 ± 27</td>
</tr>
<tr>
<td>Imidazole</td>
<td>2590 ± 60</td>
<td>2965 ± 275</td>
<td>536 ± 7</td>
<td>ND</td>
<td>ND</td>
<td>11700 ± 900</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2. Binding constants for compounds identified to bind CYP126A1 through high-throughput screening studies.

The dissociation constants ($K_d$ values) were determined by UV-visible absorbance titration. HTS compounds 1-6 induce type I (accumulation of HS heme iron) spectral shifts, while HTS compounds 7-9 reinforce the LS state of CYP126A1 through distal coordination of the heme iron via a nitrogen ligand in each case.

<table>
<thead>
<tr>
<th>High-throughput screening hits</th>
<th>Type I ($K_d$ (µM))</th>
<th>Type II ($K_d$ (µM))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 1</td>
<td>18.3 ± 1.3</td>
<td>Compound 7</td>
</tr>
<tr>
<td>Compound 2</td>
<td>2.85 ± 0.15</td>
<td>Compound 8</td>
</tr>
<tr>
<td>Compound 3</td>
<td>56.1 ± 2.2</td>
<td>Compound 9</td>
</tr>
<tr>
<td>Compound 4</td>
<td>65.2 ± 6.6</td>
<td></td>
</tr>
<tr>
<td>Compound 5</td>
<td>6.15 ± 0.31</td>
<td></td>
</tr>
<tr>
<td>Compound 6</td>
<td>25.3 ± 0.5</td>
<td></td>
</tr>
</tbody>
</table>

4.3.3 EPR Spectroscopy of CYP126A1 and Ligand-Bound forms

X-band EPR spectroscopy was used to analyze the heme coordination environment of ferric CYP126A1 in its ligand-free state and when in complex with (i) Type I compounds 1-6 identified from HTS, and (ii) Type II compounds 7-9 from HTS, as well as with a group of small compounds (compounds 10-13) shown to bind with modest affinity (~150-550 µM $K_d$) to the CYP126A1 heme iron through amine, pyridine or pyrazole nitrogens. The ligand-free CYP126A1 has a rhombic LS spectrum with a single major species, with g-values at 2.41
(g_z), 2.24 (g_y) and 1.92 (g_x), consistent with one dominant coordination state from cysteine thiolate (proximal) and water (distal) ligands. These g-values are similar to those reported for other Mtb P450s such as CYP125A1 (2.40/2.25/1.94), CYP142A1 (2.40/2.23/1.92) and CYP51B1 (2.44/2.25/1.91)\textsuperscript{7,8,14}.

EPR spectra were also collected for CYP126A1 complexes with HTS compounds and the small compounds 10-13. In several cases, substantial changes occurred in the EPR spectra, consistent with changes in distal heme iron coordination. For several of the HTS type II compounds and compounds 10-13, these molecules produced new LS EPR spectra consistent with distal ligation of CYP126A1 heme iron by a nitrogen atom. Similarly, the binding of selected HTS type I molecules resulted in new spectral signals indicative of the development of HS ferric heme iron, as well as in some perturbation to the LS ferric spectra. Selected EPR spectra are shown in Figure 4.5.

CYP126A1 heme coordination by the four small nitrogen-containing molecules gave new LS spectra with g-values at 2.44/2.24/1.92 for 6-phenoxy pyridin-3-amine (compound 10); 2.47/2.25/1.90 for 3-(pyridin-3-yl)propanoic acid (compound 11); 2.50/2.24/1.91 (minor) and 2.44/2.24/1.91 (major) for 4-(1H-pyrazol-1-yl)aniline (compound 12); and 2.48/2.25/1.89 for (4-(pyridin-4-yl)phenyl)methanol (compound 13). These EPR data suggest that direct coordination of heme iron may occur mainly through primary amine nitrogens in compound 10 and in the minor species in compound 12 (both with g_z = 2.44). For the major species in compound 12 (g_z = 2.50), pyrazole nitrogen may be the ferric iron ligand. The similar EPR spectral features for the CYP126A1 complexes with compounds 11 and 13 (g_z = 2.47 and 2.48, respectively) suggests that pyridine nitrogen ligates heme iron in both these complexes (Figure 4.5A).

The type II hits from HTS studies also perturbed the LS ferric CYP126A1 EPR spectra, producing new species with g-values at 2.49/2.24/1.88 (compound 7); 2.44/2.24/1.90
(compound 8); and 2.46/2.25/1.90 (compound 9) (Figure 4.5A). For the type I HTS compounds, the development of HS ferric heme iron spectra was observed for compound 1 ($g_z = 7.92$, $g_y = 3.57$, $g_x = 1.69$) and to a smaller extent for compounds 2 (7.79/3.77), 3 (7.85/3.63) and 4 (7.92/3.60). For compounds 2-4, the HS $g_x$ signal was too small to be accurately assigned. No significant EPR spectral changes were seen for complexes with compounds 5 and 6. EPR spectra for CYP126A1 complexes with compounds 1, 2 and 4 are shown in Figure 4.5B. The relatively small accumulation of a HS EPR signal (compared to data collected by UV-visible titrations at ambient temperature) for many of the type I HTS compounds is consistent with data for several other P450s on binding substrate-like compounds, and is likely due to the requirement for collection of heme EPR spectral data at 10 K$^{32}$. A P450 in which a much larger HS EPR signal is seen in the substrate-bound form is the alkene producing fatty acid decarboxylase OleT from a Jeotgalicoccus sp. OleT is converted to ~95% ferric HS on binding arachidic acid (C20:0), and retains a large HS signal in the EPR spectrum, probably due to the presence of a rigid binding site for the fatty acid carboxylate close to the heme iron, leading to effective displacement of the water distal ligand from the OleT heme iron$^{32}$. For the HTS compounds that did induce formation of a HS ferric heme EPR signal, there was also an effect on the LS spectrum. As seen in Figure 4.5B, there are shifts in the LS g-values to 2.38/2.22/1.96 (major), 1.92 (minor) for compound 1; 2.40/2.23/1.92 (major), 1.96 (minor) for compound 2; and 2.39/2.22/1.96 (major), 1.92 (minor) for compound 4.
Figure 4.5. Electron Paramagnetic Resonance Spectra of CYP126A1.

Panel A shows EPR spectra for CYP126A1 in its ligand-free form and for complexes with HTS compounds (compounds 7 and 9) and other small molecules that ligate CYP126A1 heme iron via nitrogen ligands (compounds 11-13). Panel B shows EPR spectra for HTS screening hits (compounds 1, 2 and 4) that induce HS heme iron accumulation in CYP126A1. The g-values for the LS and HS forms of CYP126A1 are indicated in each case.

4.3.4 MCD spectroscopy of CYP126A1

Magnetic circular dichroism (MCD) spectroscopy was used as a complementary method to EPR in order to probe electronic structure of CYP126A1. MCD data were collected in both the UV-visible (250-800 nm) and near-infrared (near-IR, 800-1400 nm) regions. MCD features at different wavelengths in the UV-visible and near-IR regions can be used to probe heme structure and heme iron coordination\textsuperscript{33,34}. UV-visible MCD spectra can be diagnostic for the spin- and oxidation-states of metal ions, and in heme proteins the positions of the porphyrin macrocycle absorption bands are influenced by the heme iron. Porphyrin $\pi$-to-$\pi^*$ transitions give rise to MCD absorption features in the ~300-600 nm region, and the mixing of the porphyrin-$\pi$ and iron-$d$ electronic levels enables UV-visible MCD to provide data on
heme iron oxidation and spin-state\textsuperscript{35}. **Figure 4.6A** shows the UV-visible MCD spectrum of CYP126A1. The sharp feature at 292 nm is due primarily to contributions from tryptophans in CYP126A1. Other major features in the heme spectrum (values in units of mM\textsuperscript{-1} cm\textsuperscript{-1}) are located at 359 nm (-43.4) and 407.5 nm (+148.8), with a crossover at 418.5 nm in the Soret region; and at 520 nm (+48.7), 556.5 nm (+56.5), 546.5 nm (0), and 575.5 nm (-83.8) in the heme Q-band region. These features in the CYP126A1 UV-vis MCD spectrum resemble those of proteins with LS ferric hemes. For example, the UV-vis MCD spectrum of CYP126A1 is highly similar to that reported for the \textit{Mtb} cyclodipeptide oxidase CYP121A1. A minor trough feature at \textasciitilde655 nm in the CYP126A1 UV-vis spectrum MCD indicates a very small proportion of HS ferric cysteine thiolate-coordinated heme iron\textsuperscript{36}.

A CYP126A1 MCD spectrum was also collected in the near-IR region, to report on the position of a ligand-to-metal CT transition that can be characteristic for heme iron ligation state in the ferric form\textsuperscript{35} (**Figure 4.6B**). The CYP126A1 CT band is located at approximately 1097 nm. This position is consistent with those seen in other P450 enzymes, e.g. the fatty acid/fatty acyl-ACP oxidizing P450 BioI (CYP107H1) at 1090 nm, and CYP121A1 at 1125 nm\textsuperscript{36,37}. 
Figure 4.6. Magnetic Circular Dichroism analysis of CYP126A1.

The MCD spectra for CYP126A1 are shown in the UV-visible region (panel A) and in the near-IR region (panel B). Spectra were collected at room temperature using a CYP126A1 concentration of 200 μM. Spectral features are consistent with a predominantly LS ferric heme iron with cysteine thiolate coordination.

4.3.5 Determination of the heme iron reduction potentials of substrate-free and ligand-bound forms of CYP126A1

The substrate-free form of CYP126A1 is predominantly LS in its resting ferric form. However, the binding of the type I HTS compounds 1-6 results in shifts towards HS heme iron, consistent with substrate-like properties. In many other bacterial P450s, the substrate-induced HS shift is associated with a substantial increase in the P450 heme iron Fe\textsuperscript{III}/Fe\textsuperscript{II} reduction potential, e.g. from \(-368\) mV to \(-239\) mV (vs. NHE) for the \textit{Bacillus megaterium} CYP102A1 (P450 BM3) heme domain on binding arachidonic acid; and from approximately \(-300\) mV to \(-170\) mV for the \textit{Pseudomonas putida} P450cam on binding \textit{D}-camphor\textsuperscript{38,39}. 
To analyze the influence of type I compound binding to CYP126A1, the heme iron redox potentials were determined for both the substrate-free CYP126A1 form and for the CYP126A1/compound 1 complex. The spectroelectrochemical redox titration for substrate-free CYP126A1 displays a Soret band transition from the oxidized form at 418.5 nm to a broad feature of lower intensity, centred at ~409 nm, as the heme iron is reduced by progressive additions of sodium dithionite (Figure 4.7A). There are clear isosbestic points in the titration (at ~410 nm and 437 nm), indicating a simple conversion between the oxidized and reduced states without accumulation of an intermediate species. By comparison with other P450 enzymes, the blue shift of the Soret band and the spectral fusion of the alpha- and beta-bands into a single Q-band feature at 550 nm is consistent with the retention of cysteine thiolate coordination in the ferrous heme iron form \(^{40,41}\). A plot of absorbance at 402 nm vs. applied potential was fitted using the Nernst equation to give a midpoint potential for the CYP126A1 heme Fe\(^{III}\)/Fe\(^{II}\) transition of \(-332 \pm 4\) mV vs. NHE (Figure 4.7A, inset).

In the case of the CYP126A1/compound 1 complex, the P450 is extensively HS in the ferric state, with a Soret maximum at 394 nm and a cysteine thiolate-to-ferric heme CT signal at 647 nm. The reduction of the heme iron occurs with a Soret band shift to 421 nm, and with loss of the CT and spectral bleaching in the Q-band region. Isosbestic points in the redox titration at 406.5 nm, 479 nm and 521 nm are again consistent with a simple two state transition between oxidized and reduced forms of the CYP126A1/compound 1 complex (Figure 4.7B). A plot of absorbance at 393 nm vs. applied potential was again fitted using the Nernst equation, giving a heme Fe\(^{III}\)/Fe\(^{II}\) midpoint potential of \(-177 \pm 3\) mV vs. NHE (Figure 4.7B, inset). Thus, binding of compound 1 results in a 155 mV increase in the CYP126A1 midpoint potential, consistent with heme iron reduction being favoured in a substrate-bound form of the P450\(^{42}\).
Figure 4.7. Determination of the heme reduction potentials of ligand-free and compound 1-bound CYP126A1.

Panel A. The main panel shows UV-visible absorbance spectral data from a spectroelectrochemical redox titration of ligand-free CYP126A1 in the potential range from ~ -170 to -395 mV (versus the normal hydrogen electrode, NHE). The UV-visible Soret spectral maximum shifts from 418.5 nm towards ~409 nm as the reductive titration approaches completion. In the Q-band region, the alpha- and beta bands appear at ~568 and 539 nm in the oxidized enzyme, with the reduced CYP126A1 exhibiting a single absorption band at 550 nm. The absorbance versus potential data at 402 nm were fitted using the Nernst equation to give a midpoint potential for the heme iron Fe$^{III}$/Fe$^{II}$ couple as -332 ± 4 mV vs NHE (inset). Panel B shows a similar data set for the compound 1-bound form of CYP126A1. In this case, the main panel shows P450 spectral data collected in the potential range from ~ -30 to -265 mV vs NHE. In the ferric state, the compound 1-bound CYP126A1 has a Soret maximum at 394 nm, with Q-band maxima at 538 and 568 nm, and with a low-intensity CT-band at ~647 nm. The reduced form has a Soret spectral maximum at 421 nm, with loss of absorbance intensity in the Q-band region and for the CT species. Absorbance versus potential data were plotted at 393 nm and fitted using the Nernst equation to give a heme iron midpoint potential of -177 ± 3 mV vs NHE.
4.3.6 Analysis of CYP126A1-mediated oxidation of HTS molecules

Studies to identify products of oxidation from several of the type I HTS compounds were done using two bacterial type NADPH-dependent redox systems comprising (i) *E. coli* flavodoxin reductase and (ii) *E. coli* flavodoxin or spinach ferredoxin. No turnover was detected with **compounds 1 and 2**, both of which contain nitroaromatic groups. As discussed below, it is likely that these compounds act as inhibitors of CYP126A1, while simultaneously inducing a type I CYP126A1 heme spectral shift that is characteristic of substrate binding in most other P450s. While evidence for oxygen incorporation (+16 addition to the parent species molecular mass) into a minor proportion of **compounds 3 and 6** could be obtained from the small amounts of products resolved using LC-MS, there was insufficient product to identify the position of oxidation. However, in the case of **compound 4** (3-(4-(benzo[d][1,3]dioxol-5-ylmethyl)piperazin-1-yl)-1-(4-chlorobenzyl)pyrrolidine-2,5-dione) a higher amount of oxidized product was formed. While the precise position of oxidation could not be resolved fully by molecular fragmentation, it was clear that oxygen addition (hydroxylation) occurs on the 1,3-benzodioxole portion of the molecule (Figure 4.8).
**Figure 4.8.** CYP126A1-dependent oxidation of Compound 4.

**Panel A.** LC extracted ion chromatogram of CYP126A1 extracts following enzymatic incubations with compound 4. Major peaks with retention times at 2.791 and 3.231 min are observed, with minor species at 2.916, 3.074 and 3.912 min. The peak at 2.791 min (shaded in grey) corresponds to the (+16) hydroxylation of compound 4. **Panel B.** Mass spectrum following LC separation of the extracted major product at 3.231 min, with m/z 442.1528 (M+H) diagnostic of the parent substrate ion of compound 4 ([C$_{23}$H$_{24}$ClN$_3$O$_4$]+H) and a smaller species with m/z 464.1348 for the compound 4 sodium ion adduct. **Panel C.** Mass spectrum of the extracted 2.791 min peak showing the (+16) hydroxylation of compound 4 with m/z 458.1477 ([C$_{23}$H$_{25}$ClN$_3$O$_5$]+H). Inset is an amplification of the peak m/z 458.1477, showing the ionization pattern for the hydroxylated form of compound 4.

### 4.3.7 Crystal structure of CYP126A1 in the imidazole-bound form

The structure of the ligand-free form of CYP126A1 was determined by molecular replacement using the structure of the P450 EryF (CYP107A1), a 6-deoxyerythronolide B hydroxylase (forming erythronolide B) in the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (PDB code 1OXA)$^{43}$. The CYP126A1 structure exhibits the characteristic P450 fold, and the ligand-free structure contains a dimer in the asymmetric unit (Figure 4.9). The dimer interface is made by the substrate binding BC- and FG-loop regions, and while one monomer (chain A, a “closed” form with respect to active site access) has
water as a 6\textsuperscript{th} heme ligand, with a relatively solvent inaccessible active site, the other monomer is bound to imidazole as the 6\textsuperscript{th} heme ligand (likely retained through the nickel affinity purification procedure in a proportion of the P450 molecules). In contrast to the ligand-free active site, the imidazole-bound form contains a clear solvent access channel (chain B, the “open” form). The access channel to the bound imidazole is opened due to disorder of the FG-loop region (residues 177-194). In comparison, the FG-loop region of chain A is ordered, but explores in part the chain B active site region. Hence, it is possible that ligand binding to chain B affects the conformation of chain A. An overlay of both monomers reveals that imidazole binding is linked to a reorientation of the N-terminal region of the I-helix at residues 240-257, moving Ala253 by 2.7 Å to avoid severe steric clashes with the heme 6\textsuperscript{th} ligand. As the F- and G-helices are involved in hydrophobic contacts with this particular region of the I-helix, this I-helix relocation leads to an accompanying shift in orientation of the FG region. While the chain B FG-loop is disordered, the chain A FG-loop chain trace makes contact with the chain B I-helix N-terminal region at residues 194-196.

The asymmetric behavior of the CYP126A1 dimer can be rationalized by the fact that a putative A-A symmetric dimer (where both chains adopt the chain A conformation) is not possible due to a severe steric clash of the respective FG-loop regions. In other words, the ordered FG-loop region in chain A at the dimer interface obscures the putative FG-loop docking region for chain B. As imidazole binding leads to the I-helix and consequently the FG helix reorientation, it is possible that imidazole binding to one monomer leads to the active site closure of the second monomer. It is interesting to note that a putative B-B symmetric dimer (where both chains adopt the chain B conformation) also appears possible.

Given the flexibility of the P450 fold, it is difficult to predict substrate identity from P450 structures. In the case of CYP126A1, the ligand binding pocket is not exclusively lined by hydrophobic residues (Met190, Leu249, Ala253), but also contains Thr83, Asn96, Ser300,
Lys303, Arg400, and His401, suggesting that the physiological substrate likely exhibits hydrophilic properties. For chain B, the imidazole ligand is positioned in between I-helix residues Leu249 and Thr257, and is stacked against Ala253. In this case, the side chain of Arg400 has reoriented and points away from the active site.

Figure 4.9. The crystal structure of CYP126A1.

Panel A. A stereoview of the CYP126A1 dimer structure with the N-terminal region of the I-helix, as well as the F- and G-helices and the BC-loop region, colored in blue for the imidazole-bound monomer and in cyan for the ligand-free monomer. Panel B. A cross-section of the CYP126A1 dimer through both heme groups, revealing that the imidazole-bound monomer has an open active site with access to solvent, while the opposite monomer has a closed active site. Color coding as in panel A. Panel C. A stereoview of an overlay of both active sites of the CYP126A1 dimer. Key residues are shown in atom colored sticks, with blue carbons for the imidazole-bound monomer and cyan carbons for the ligand-free monomer.
4.3.8 Crystal structure of CYP126A1 in complex with inhibitor and substrate-like molecules from compound screens

In order to further establish the nature of CYP126A1 protein-ligand interactions, and the effects these have on protein oligomerization/conformation, we determined the crystal structures of three distinct inhibitor complexes through co-crystallization. The crystal structure of CYP126A1 in complex with theazole drug ketoconazole revealed that a large scale reorientation of the BC-loop region has occurred in response to ligand binding, disrupting the putative dimer interface observed in the imidazole-bound form (Figure 4.10). Furthermore, the N-terminal region of the I-helix and the associated FG-helices display comparatively more modest reorientations as a consequence of ligand binding. The ketoconazole dichlorobenzene moiety occupies the region previously filled by Leu249, while the extended tail of the molecule displaces the BC-loop residues Val84 and Leu85, resulting in their forming hydrophobic contacts with the FG-helices.

In contrast, the binding of the smaller imidazole-containing compound 7 does not lead to BC-loop reorientation, and a putative CYP126A1 dimer in complex with compound 7 in both its active sites is observed in the asymmetric unit (Figure 4.11). The overall conformation is similar to the proposed B-B symmetric dimer, with both chains adopting a conformation resembling the imidazole-bound monomer. Additional contacts with compound 7 are limited to hydrogen bonds with Asn96 and Ser252, and with the ligand urea moiety, while the adamantanyl unit remains largely solvent exposed.

Unlike the previous ligands, compound 1 does not contain an azole group, and instead elicits a type I heme shift suggestive of substrate-like ligand binding. The crystal structure of CYP126A1 in complex with compound 1 reveals that the ligand is bound with the nitrobenzene moiety placed directly above and parallel to the heme porphyrin plane (Figure 4.11). This effectively ensures that the heme iron 6th ligand water is displaced, explaining the
observed spectroscopic effect with HS heme iron formation. As observed for the azole-containing ligands, the binding of compound 1 elicits a shift in the N-terminal region of the I-helix to accommodate the nitrobenzene moiety, accompanied by the reorientation of the FG-helix region. While the BC-loop region appears unperturbed by compound 1 binding, as we also observed for binding of compound 7, the putative dimer interface observed in the CYP126A1:compound 1 complex is distinct in nature. While still composed of interactions between the FG- and BC-loop regions, the respective orientation of both monomers is altered. Although the compound 7 occupies a binding pocket similar in nature to that observed for compound 1, it would seem that small changes in the conformations of residues lining the compound 7 binding pocket (Gln86 and Arg400) lead to distinct packing interactions with the FG-helices of the opposite dimer.
Figure 4.10. Structures of CYP126A1-ligand complexes.

Overlays of the various CYP126A1-ligand complex structures are shown. Panel A. A stereoview of an overlay of the CYP126A1 ligand complex structures together with the ligand-free monomer structure (in cyan). The ligand complexes are colored blue for imidazole, yellow for compound 7, red for compound 1 and green for ketoconazole. Only those regions affected in conformation by ligand binding are in color. Panel B. A stereoview of an overlay of the various CYP126A1-ligand complex active sites. Color coding as in panel A. The N-terminal region of the I-helix is shown, revealing that highly ligand-specific changes occur. These are transmitted through to the F- and G-helices that are in direct contact with this region.

Figure 4.11. Active site structures of CYP126A1-ligand complexes.

Detailed views of the active sites of CYP126A1 are shown from the structures of the complexes with ketoconazole (Panel A), the type II inhibitor compound 7 (Panel B) and the type I substrate mimic compound 1 (Panel C). Key residues are shown in atom colored sticks, with the omit electron density for the various ligands contoured at 3 sigma. The N-terminal region of the I-helix is shown in cartoon. Compounds 1 and 7 bind in a channel that follows the general direction of the I helix, while ketoconazole extends away from the heme in an orthogonal direction.
4.4 Discussion and Conclusions

Orthologues of the \textit{Mtb} \textit{CYP126A1} gene are widely distributed among both pathogenic (e.g. \textit{M. ulcerans} Agy99, 80% amino acid sequence identity) and non-pathogenic (e.g. \textit{M. smegmatis} mc\textsuperscript{3}155, 79% identity) mycobacteria. This suggests an important conserved function in these organisms. Related P450s are also found in other actinobacteria (e.g. \textit{Streptomyces} sp. Eco86, 43% identity), and also among the myxobacteria (e.g. \textit{Sorangium cellulosum} So ce56, 42% identity; and \textit{Stigmatella aurantiaca} DW4/3-1, 41% identity), again suggesting that a common function may be retained in diverse bacteria\textsuperscript{44}. Phylogenetic studies indicate that CYP126A1 is most closely related to CYP125A1, CYP142A1 and CYP124A1 from \textit{Mtb} \textit{H37Rv}, all of which have been implicated in host cholesterol metabolism\textsuperscript{6-9,45,46}. However, there is no evidence for cholesterol binding to CYP126A1. While there are relatively few data available on \textit{CYP126A1} from transcriptomics and microarray studies, the \textit{CYP126A1} gene is not essential for the \textit{in vitro} growth of the virulent \textit{Mtb} \textit{H37Rv} strain. However, its importance in the infective state remains uncertain\textsuperscript{47}. CYP126A1 protein is clearly identified in \textit{Mtb} \textit{H37Rv} cell lysates by mass spectrometry, and is not identified in the bacterial culture filtrate or membrane fractions\textsuperscript{48,49}. Moreover, a computational analysis of predicted \textit{Mtb} target proteins and their druggability ranks CYP126A1 (and other \textit{Mtb} P450s) in the top 50 potential targets among more than 1500 candidate \textit{Mtb} proteins\textsuperscript{50}.

In order to characterize in more detail the structural and biochemical properties of CYP126A1, we purified the P450 using an \textit{E. coli} expression system and demonstrated by mass spectrometry that the protein was intact. Nano-ESI MS studies indicated a propensity for the enzyme to dimerize, and (at higher concentrations) to form small amounts of trimeric and tetrameric species (\textbf{Figure 4.1}). This property likely underpins the formation of CYP126A1 dimer crystals (see below). UV-visible spectroscopy confirms typical heme
spectral features in CYP126A1. A slow collapse of the P450 form of the CYP126A1 Fe\textsuperscript{II}-CO complex at 448.5 nm to the P420 state at 423 nm occurs, and this phenomenon is also observed in other Mtb P450s\textsuperscript{8,13,14}. On the basis of previous studies, it is expected that the thiolate-coordinated P450 state would be stabilized by the binding of its natural substrate\textsuperscript{8,18}. Supporting this conclusion, the stabilization of the thiolate-coordinated ferrous form of CYP126A1 is evident in the redox titration of the P450 bound to the type I screening hit \textbf{compound 1}. The blue Soret shift seen on CYP126A1/\textbf{compound 1} heme iron reduction, along with spectral “fusion” of the Q-bands, is characteristic of retention of thiolate coordination in the ferrous state, whereas a Soret red shift (e.g. to 423 nm in the case of \textit{Mtb} CYP51B1) with shifted $\alpha$- and $\beta$-band features is typical for cysteine thiol-coordinated ferrous heme iron\textsuperscript{14,38}. The redox titration also reveals an $\sim$155 mV increase in the reduction potential of the CYP126A1 heme iron Fe\textsuperscript{III}/Fe\textsuperscript{II} couple, an effect likely dominated by the extensive conversion of the CYP126A1 ferric heme iron to the HS state\textsuperscript{38,39}. However, the significant structural arrangements accompanying the binding of \textbf{compound 1} (Figure 4.10) also alter the CYP126A1 heme environment and are likely to influence the heme redox potential.

CYP126A1 binds avidly to a number of azole drugs, but shows a clear preference for binding azole antifungal drugs with imidazole rather than triazole groups. Azole binding induces a red (type II) Soret band shift with a distinctive decreasing of Soret absorbance intensity (by $\sim$15-20%) as the titrations progress to completion (Figure 4.3). An interesting pattern of azole drug selectivity among \textit{Mtb} P450s emerges from data presented in Table 4.1. Firstly, it is clear that in the case of CYP126A1, CYP144A1 and CYP121A1 the affinity for the imidazole-containing azole drugs is markedly greater than those for the triazole-containing drugs tested. Indeed, no significant type II Soret shifts were observed for CYP126A1 with fluconazole, voriconazole or itraconazole. The CYP144A1 $K_d$ values for
econazole, clotrimazole and miconazole are sub-micromolar, while those for the triazoles fluconazole and voriconazole are >10 mM and 6.5 mM, respectively. CYP121A1 binds fluconazole and voriconazole more tightly (K_d values are 8.61 and 16.3 μM, respectively), but affinity remains ~120-680-fold weaker than for econazole and miconazole. However, there is much less discrimination between the imidazoles and triazoles in at least CYP51B1 and CYP125A1, although voriconazole does not bind the heme iron in CYP125A1. For CYP51B1, the binding of fluconazole and voriconazole is quite tight (K_d = 5.82 and 2.10 μM, respectively) and only 12-32-fold weaker than that for clotrimazole, the tightest binding imidazole drug. Ketoconazole displays similar affinity (K_d = 3.57 μM) to the triazole drugs in CYP51B1. It is known that that CYP51B1, CYP125A1 and CYP142A1 are all sterol oxidizing P450s in Mtb, whereas no such role has been assigned to CYP126A1 or the other Mtb P450s in Table 4.1. The binding of fluconazole and voriconazole to CYP126A1 is not detectable, whereas at least fluconazole binds each of three known sterol-metabolizing Mtb P450s with good to moderate affinity. It can thus be speculated that, despite apparent phylogenetic relationships between CYP126A1 and sterol-metabolizing Mtb P450s, CYP126A1 has taken a distinct evolutionary path and that its role in pathogenic and non-pathogenic mycobacteria (and other bacteria) may be unrelated to metabolism of host or environmental sterols. With regards to the effectiveness of the various azole drugs in inhibiting mycobacterial growth, preceding studies showed that the triazoles fluconazole and itraconazole had no significant effect on Mtb growth in vitro, whereas ketoconazole had an MIC of 8-16 μg ml^{-1} against the virulent Mtb H37Rv strain^{51}. Our own studies confirmed the greater potency of the imidazole-containing azole drugs, with fluconazole proving ineffective against Mtb H37Rv, while econazole (8 μg ml^{-1}), miconazole (8 μg ml^{-1}), clotrimazole (11 μg ml^{-1}), and ketoconazole (16 μg ml^{-1}) all showed good MIC values^{20}. Imidazole azole drug potency was even greater against the laboratory strain M. smegmatis mc^{2}155 in most cases,
with MIC values of <0.1 μg ml\(^{-1}\) for econazole, 0.1 μg ml\(^{-1}\) for clotrimazole and 1.25 μg ml\(^{-1}\) for miconazole, compared to >100 μg ml\(^{-1}\) for the triazole fluconazole\(^{19}\). The enhanced azole potency against \(M.\) \textit{smegmatis} mc\(^{2}\)155 may reflect greater drug permeability as a consequence of altered glycolipid composition in the cell envelope\(^{52}\).

In work to identify novel type I and type II compounds binding to CYP126A1, we undertook a large (ca 20,000) compound screen, and identified several molecules that induce CYP126A1 heme Soret shifts consistent with substrate-like (HS heme iron accumulation with Soret blue shift) or heme coordinating (Soret red shift, similar to that seen for azole drug binding). The compound hits selected for analysis were all extended molecules containing between 2-6 aromatic or non-aromatic ring structures (Figure 4.4). While these bulky compounds do not provide immediate clues to the identity of physiologically relevant substrates, type I CYP126A1 spectral shifts were obtained in several cases, and evidence of P450-mediated substrate oxidation of HTS type I compounds was obtained in three cases, with the clearest results obtained with compound 4.

Crystal structures were determined for CYP126A1 in absence of added ligand, and for CYP126A1 in complex with ketoconazole, and with HTS molecules compound 1 (type I hit) and compound 7 (type II hit). The imidazole group of compound 7 clearly ligates the heme iron, whereas the nitrobenzene moiety of compound 1 is placed directly above the heme plane and displaces the distal water, inducing adoption of the HS ferric state in CYP126A1. No evidence of oxidation of either compound 1 or compound 2 (which both possess a terminal 4-nitrophenyl group) could be obtained in turnover studies, and it thus appears that these molecules may “moonlight” as substrates, but are not readily oxidized and instead act as P450 inhibitors. A small number of nitroaromatic compounds were identified as inhibitors of the P450-like, cysteine thiolate-coordinated nitric oxide synthase (NOS) enzymes\(^{53}\). However, there is no available evidence for their binding in a similar mode to that shown...
here structurally for **compound 1**. The likely similar binding modes for both **compounds 1 and 2** to CYP126A1 suggests that development of related nitroaromatic compounds could lead to more potent inhibitors of CYP126A1, and possibly other P450 enzymes.

The CYP126A1 structures reveal a dynamic molecule, with the BC and FG regions clearly affected by ligand binding. The latter are also involved in the formation of the dimeric states of the protein, establishing a link between ligand binding and protein oligomerization. The malleable nature of CYP126A1 (as demonstrated by its ability to bind to a range of structurally diverse inhibitors and substrate-like molecules) suggests that the nature of the physiological substrate does not necessarily reflect the shape of the ligand-free CYP126A1 structure. However, the nature of the residues that line the various active site regions does suggest that the physiological substrate contains polar moieties that might establish interactions with residues such as Thr83, Asn96, Ser300, Lys303, Arg400, and Arg401. The propensity for CYP126A1 to dimerize was seen in our nanoESI mass spectrometry studies (Figure 4.1B) and this may be an important factor in underpinning crystallographic dimer formation. Ketoconazole binding reorients the BC-loop region of CYP126A1 and disrupts the dimer interface, resulting in a binding mode for ketoconazole that is approximately orthogonal to those for HTS **compounds 1 and 7**, which extend upwards from the heme, close to the path followed by the I-helix (Figure 4.10). The binding modes of these molecules are compatible with CYP126A1 dimer formation in the crystal, although differences in the dimer interface regions are observed. The ability of CYP126A1 to form crystallographic dimers is similar to properties reported for the *Mtb* CYP130A1 P450, with similar regions of the P450s forming the dimer interfaces. However, while CYP130A1 crystallizes as a monomer in the “open” conformation when ligand-free, the econazole-bound form has a “closed” conformation in a dimeric state. In contrast, the ketoconazole-bound
CYP126A1 is a crystallographic monomer, while the ligand-free form of CYP126A1 is a dimer containing both “open” and “closed” monomers.

In conclusion, we present the first biochemical and structural studies of the Mtb P450 CYP126A1, revealing novel substrates and inhibitors for the enzyme, defining its relatively polar active site and its ability to adapt structurally to facilitate the binding of bulky ligands. Future work is directed at identification of the natural substrate(s) for this widely conserved P450 enzyme.

4.5 Experimental Procedures

4.5.1 Cloning and Production of a CYP126A1 Expression Plasmid Construct

The gene encoding CYP126A1 (Rv0778) from the Mtb H37Rv strain was amplified by PCR from a Mtb H37Rv chromosomal cosmid DNA library (supplied by Dr. Roland Brosch, Institut Pasteur, Paris). The CYP126A1–containing BAC clones were prepared using standard PCR amplification protocols with Pfu Turbo DNA polymerase (Agilent, Cheadle UK) used to amplify the gene. The Mtb H37Rv genome sequence was used to design primers to generate the Rv0778 gene for insertion into the pET15b plasmid vector using the forward primer:

5’-GCGGCAGCCATATGACTACCAGCG-3’ and the reverse primer:

5’-GCGCGAGGCTAGGATCCGGTGC-3’ (Merck-Millipore, Watford UK). The underlined letters in the forward and reverse primers indicate engineered NdeI and BamHI restriction endonuclease 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 polymerization step of 72 °C for 10 min. The PCR product was then digested using NdeI/BamHI and cloned into pET15b using a Quick Ligation Kit (NEB, Hitchin UK) to generate the pCYP126A1 construct.
4.5.2 Expression and purification of CYP126A1

The production of CYP126A1 protein was achieved by transforming E. coli strains BL21 Gold (DE3) (Agilent) or C41 (DE3) (Merck-Millipore, Watford UK) with the pCYP126A1 plasmid construct. Gene expression used an isopropyl β-D-thiogalactopyranoside (IPTG)-inducible T7 RNA polymerase/promoter system. IPTG-dependent expression of the T7 RNA polymerase from a chromosomally integrated gene copy in C41 (DE3) resulted in T7 RNA polymerase-dependent transcription of the CYP126A1 gene in pET15b. Production of CYP126A1 protein was typically done in ~15 litre cultures of 2xYT growth medium (ForMedium, Hunstanton UK). The culture medium was distributed between 24 x 2 litre conical flasks, each containing ~600 ml of growth medium and ampicillin (50 g/ml) for plasmid selection. The medium was inoculated with 6 ml of E. coli transformant cells from an overnight culture of cells grown in the same medium. The cells were grown at 37 °C with agitation (200 rpm) until an OD\textsubscript{600} of 0.5 was reached, and then the growth temperature was decreased to 23 °C and bacterial cell growth continued until an OD\textsubscript{600} of 0.7 was reached. 100 μM IPTG was then added to induce CYP126A1 gene expression, along with 100 μM delta-aminolevulinic acid (ΔALA) to promote heme synthesis and its incorporation into the P450. The transformant cells were then grown for a further 36 h under the same conditions. The cells were harvested by centrifugation at 6000 g for 10 min at 4 °C using a JLA-8.100 rotor in an Avanti J-26 XP centrifuge, after which the supernatant was discarded and cell pellets were resuspended in ~300 ml of ice cold 50 mM potassium phosphate (KPi, pH 8.0) containing 250 mM NaCl and 10% glycerol (buffer A). The protease inhibitors phenylmethanesulfonyl fluoride (PMSF, 1 mM), benzamidine hydrochloride (1 mM) and six cOmplete EDTA-free tablets (Roche Diagnostics Ltd, West Sussex UK) were added to inhibit proteases. The cells were lysed on ice by ultrasonication (Bandelin Sonopuls sonicator) with 6 cycles of 30 s on
and 60 s rest periods at 45% amplitude. The cell lysate was then centrifuged at 40,000 g for 45 min at 4 °C and the supernatant collected.

After centrifugation, the supernatant was loaded immediately onto a Ni-IDA column (Generon, Maidenhead UK) pre-equilibrated with cold buffer A plus the aforementioned protease inhibitors, using a peristaltic pump (GE Healthcare, Little Chalfont UK). The column was washed with ~100 ml of buffer A and the flow-through discarded. Proteins were eluted from the column by washing consecutively with increasing concentrations of imidazole [30 mM (250 ml), 60 mM (150 ml), 120 mM (100 ml) and 180 mM (100 ml) in buffer A]. Each eluted fraction was analyzed both by UV-visible spectroscopy (250-800 nm), and by SDS-PAGE. Fractions containing relatively pure CYP126A1 (mainly the 80 mM, 120 mM and 180 mM samples) were pooled and concentrated to ~100 ml using ultrafiltration with Amicon concentrators (Merck-Millipore) at 4 °C. Thereafter, concentrated protein was dialyzed into 50 mM Tris HCl (pH 7.2, dialysis buffer) containing 50 mM KCl and 1 mM EDTA to remove excess imidazole. The dialyzed protein was then loaded onto a Q-Sepharose column (10 cm x 4 cm) pre-equilibrated with the dialysis buffer, and the column was then washed and the protein eluted with a linear gradient of KCl (50-500 mM) in dialysis buffer using an AKTA purification system (GE Healthcare). Fractions were analysed by UV-vis spectroscopy and by SDS-PAGE as before. Samples with high A_{418}/A_{280} (or Reinheitszahl, Rz) ratios (≥1) were pooled and concentrated to ~200 μl by ultrafiltration using a Centriprep 30 concentrator (Merck-Millipore) at 1500 g. The concentrated protein was further dialysed into 10 mM Tris HCl (pH 7.5) containing 150 mM NaCl. The protein was then subjected to a final purification step using Sephacryl S-200 size exclusion chromatography (SEC) on an AKTA purifier. Fractions were again analysed both by UV-vis spectroscopy and by SDS-PAGE. Fractions with Rz values of ≥1.5 were seen to be pure by SDS-PAGE and these were
pooled and concentrated as before, and dialyzed into 50 mM Tris HCl, pH 7.5 containing 50 mM NaCl and 20% glycerol, and stored at -80 °C until use.

4.5.3 Native Mass and NanoESI Mass Spectrometry

Protein stock solutions (~40 μM) were prepared by dilution of purified CYP126A1 protein (~500 μM) in 200 mM ammonium acetate buffer, pH 7.0. Samples were buffer exchanged by size exclusion chromatography using Micro Biospin 6 columns, molecular weight cut-off 6 kDa (BioRad, Hemel Hempstead, UK). Ligands were prepared as stock solutions in d<sub>6</sub>-DMSO at 0.2-1 mM concentrations. Ligand-protein samples were prepared by diluting protein (10 μl) and ligand stocks (1 μl) with ammonium acetate buffer (9 μl) to give final concentrations of 10 μM CYP126A1, 10-50 μM ligand and 5% v/v d<sub>6</sub>-DMSO. Mass spectra were recorded on a Synapt HDMS instrument (Waters UK Ltd., Manchester UK). Capillaries for nano-electrospray ionization mass spectrometry (nano-ESI MS) were purchased from ThermoFisher (Hemel Hempstead, UK). Capillary tips were cut under a stereo microscope to give inner diameters of 1−5 μm and then loaded with 2.5 μl of sample solutions. Given below are the general instrumental conditions used to acquire the reported spectra. However, parameters were recorded and varied over the course of each experiment to observe the strength of protein-ligand complexes under different ionizing strengths. All measurements were carried out in a positive ion mode with ion source temperature of 20 °C. A capillary voltage of 1.5 kV, a cone voltage of 40 V and an extraction cone voltage of 4.8 V were applied to perform nanoESI MS. All reported spectra were collected with a trap collision energy 30-60 V, transfer collision energy 12-30 V, IMS pressure 5.02 × 10<sup>-1</sup> mbar, TOF analyser pressure 1.17 × 10<sup>-1</sup> mbar. External calibration of the spectra was achieved using cesium iodide at 100 mg ml<sup>-1</sup> in water. Data acquisition and processing were performed using Micromass MassLynx v4.1. Mass differences resulting from ligand binding were calculated from the unbound protein peak internal to each spectrum. The unbound protein peak was
compared to the relevant 5% v/v $d_6$-DMSO control spectrum for consistency. Mass differences were divided by the molecular weight of the ligand to calculate binding stoichiometry.

### 4.5.4 UV-Visible Absorbance Titrations

Optical titrations to determine $K_d$ values were carried out on a Cary 60 UV-visible spectrophotometer (Varian, UK) according to previously described methods$^{29,54,55}$. All titrations were performed using 1 cm pathlength quartz cuvettes. For binding studies, the following compounds were used: (i) azole antifungal drugs; (ii) compound hits identified from a high throughput screen. These compounds were identified on the basis of CYP126A1 heme spectral perturbation using a library of approximately 20,000 organic compounds; and (iii) a group of small compounds able to bind CYP126A1 heme iron through amine, pyridine or pyrazole nitrogens.

High throughput compound screening was performed at the Screening Unit at the Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin, using a 20,000 compound library of organic compounds. In brief, compound stock solutions were prepared in DMSO at 10 mM. Medium throughput pipetting robots (Sciclone 3000, Caliper Life Sciences, Hopkinton, USA) were used to aliquot the library (0.4 μl) into 384 well microtiter plates, and the compounds were diluted in 50 mM Tris-HCl, 50 mM pH 7.5 buffer (40 μl) containing 10% glycerol to a concentration of 100 μM. The plates were then incubated at 37°C for 15 min, followed by 5 min of sonication to ensure solubilization of the library compounds, and then centrifugation for 10 min to remove any particulate material. Absorbance readings of the plates were done using a plate reader (Safire, Tecan, Reading, UK) as a background value as well as for starting reference spectra of the ligand-free CYP126A1 protein. CYP126A1 (10 μl of 10 μM solution) was then pipetted into the plates and absorbance spectra were recorded between
350-460 nm to analyze compound-induced changes in the CYP126A1 heme spectrum. Control data were also generated during the screening process, including following addition of miconazole and clotrimazole as verified type II ligands for CYP126A1. Background spectra of the compounds and reference spectra of ligand-free CYP126A1 were subtracted from the spectra generated following additions of the compound library molecules in order to generate difference spectra. Measurements at each wavelength were averaged and assigned scores based upon the extent of changes in the difference spectra in order to determine hit compounds. 15 type I and 15 type II compounds were identified as ‘hits’ and the top ranking compounds were purchased, where available, from ChemDiv (San Diego, USA) or Vitas M (Apeldoorn, The Netherlands).

Compounds were prepared as stock solutions (0.1-100 mM) in DMSO and titrated (in 0.1-0.2 μl aliquots) into 1 ml cuvettes containing either a solution of CYP126A1 (typically 4-6 μM) in 100 mM KPi (pH 7.0) plus 10 mM KCl (buffer B), or buffer alone. DMSO concentrations did not exceed 1% of the final assay volume and the CYP126A1 absorbance spectrum was not affected by DMSO within this range. Spectra were recorded continuously between 800-250 nm at 25 °C. Spectra collected from the buffer control cuvette were subtracted from protein spectra to remove any optical interference from small molecule absorbance. Difference spectra were generated by subtraction of the initial ligand-free protein spectrum from each successive ligand-bound protein spectrum. The maximum change in absorbance for each difference spectrum (ΔA_{peak} minus ΔA_{trough}) was then plotted against ligand concentration. Data were fitted using a standard hyperbolic (Michaelis-Menten) function, or the Morrison equation for tight-binding ligands in cases where the K_d for the ligand was ≤ 5x the CYP126A1 concentration used in the binding assay. The Hill function was used when plots of induced absorbance change versus ligand concentration were clearly sigmoidal in character, as described in previous studies\textsuperscript{29,54,55}.
4.5.5 EPR and MCD Spectroscopy

X-band Electron Paramagnetic Resonance (EPR) spectra of ligand-free CYP126A1 (200 μM) as well as CYP126A1 (200 μM) bound to various ligands (typically 500 μM) were recorded using a Bruker ER-300D series electromagnet and a microwave source interfaced with a Bruker EMX control unit, which was fitted with an ESR-9 liquid helium flow cryostat (Oxford Instruments), and a dual-mode microwave cavity from Bruker (ER-4116DM). Spectra were recorded at a temperature of 10 K, a microwave power of 2.08 mW, and an amplitude of 1 mT. Samples were prepared in buffer B, with ligands added from concentrated stock solutions in DMSO. Magnetic Circular Dichroism (MCD) spectra were collected for CYP126A1 in the near UV-visible and near-infrared (near-IR) regions, using Jasco J-810 and J-730 circular dichrographs, respectively, and using an Oxford Instruments superconducting solenoid with a 25 mm ambient bore to generate a magnetic field of 6 tesla. A 0.1 cm pathlength quartz cuvette was used to record near-IR spectra at a sample concentration of 200 μM. UV-visible MCD spectra for CYP126A1 (also at 200 μM) were recorded using a 0.2 cm pathlength quartz cuvette in 50 mM potassium phosphate in 2H2O (pH*, 7.5) as buffer (where pH* is the apparent pH measured in 2H2O using a standard glass pH electrode). Data analysis was done using Origin Software (OriginLab, Northampton MA).

4.5.6 Redox Potentiometry

The midpoint potential for the CYP126A1 heme FeIII/FeII couple was determined for the substrate-free CYP126A1 and for the P450 bound to a molecule identified through high throughput screening to induce a substantial substrate-like shift in CYP126A1 FeIII heme iron equilibrium towards HS (compound 1 – (N-isopropyl-3-(4-nitrophenyl)-N-((3-N-(p-tolyl)-1,2,4-oxadiazol-5-hexyl)methyl)prop-1-en-2-amine)). Spectroelectrochemical titrations were performed in an anaerobic glove-box (Belle Technology) under a nitrogen gas atmosphere. All solutions were degassed under vacuum with nitrogen prior to use in the glove-box.
Oxygen levels were maintained at less than 5 ppm. The concentrated CYP126A1 protein samples were passed through a Sephadex G25 column (1 x 20 cm) (10DG column, Bio-Rad) immediately on admission to the glove-box to remove traces of oxygen. This column was pre-equilibrated and proteins were buffer-exchanged into anaerobic 100 mM potassium phosphate, pH 7.0, plus 10% glycerol to stabilize the protein from aggregation during the titration. The CYP126A1 solutions (~10-15 μM enzyme in 5 ml buffer), were titrated electrochemically according to the method of Dutton\textsuperscript{56,57}, using sodium dithionite as reductant and potassium ferricyanide as oxidant. For Compound 1-bound CYP126A1, the compound was added until no further induction of HS heme iron accumulation was observed. Mediators were added to expedite electronic equilibration in the system (2 μM phenazine methosulfate [PMS], 5 μM; 2-hydroxy-1,4-naphthoquinone [HNQ], 5 μM; methyl viologen [MV], 1 μM; and benzyl viologen [BV], 1 μM) and to mediate in the range between approximately +100 mV and -480 mV versus the normal hydrogen electrode (NHE)\textsuperscript{57}. 10-15 minutes were allowed to elapse between each addition of reductant/oxidant to ensure equilibration and stabilization of the electrode reading. Spectra (250-750 nm) were recorded using a Cary UV-50 Bio UV-Visible scanning spectrophotometer coupled to a fibre optic probe immersed in the CYP126A1 solution. The potential was measured using a SevenEasy S20-K meter (Mettler Toledo, Leicester UK) coupled to a Calomel electrode (ThermoRussell, Cupar UK) at 25 °C. The calibration of the electrode was done by using the Fe\textsuperscript{III}/Fe\textsuperscript{II} EDTA couple as a standard (+108 mV). The electrode reading was corrected by +244 mV relative to the NHE. Absorbance data at wavelengths reporting on the transition of the heme Soret band between oxidized (Fe\textsuperscript{III}) and reduced (Fe\textsuperscript{II}) forms of the CYP126A1 were plotted \textit{versus} applied potential, and the data fitted using the Nernst function in Origin 8.0 (OriginLab, Northampton MA).
4.5.7 Analysis of CYP126A1-dependent oxidation of substrate-like compounds from compound screening

CYP126A1 activity assays with type I molecules from compound screening were carried out with 0.5-1 μM CYP126A1, 10-20 μM spinach ferredoxin or *E. coli* flavodoxin (FLD), 2.5 μM *E. coli* flavodoxin reductase (FLDR) and 10 μM screening compound. Turnover reactions were performed in total volumes of 1 ml (or 5 ml for MS-fragmentation analysis) in 50 mM KPi, 50 mM KCl, pH 7.0. Reactions were initiated by the addition of 1 mM NADPH with a regenerating system (10 mM glucose 6-phosphate and 2 units glucose-6-phosphate dehydrogenase) and incubated at 30 °C for 30 minutes. Reactions were terminated by the addition of 2-5 ml of dichloromethane. The organic phase was extracted twice following centrifugation, evaporated, and dissolved in 50:50 acetonitrile/methanol for LC-MS studies. Samples were analyzed on an Agilent 6550 iFunnel Q-TOF LC–MS with a 1290 Infinity LC system. A ZORBAX Eclipse Plus C18 (2.1×50 mm; 1.8 μm) Rapid Resolution HT column (Agilent) was used with gradients of 0.1% formic acid to acetonitrile or methanol to resolve products. Fragmentation data analysis was performed with the MassHunter MSC (Molecular Structure Correlator) program (Agilent).

4.5.8 Crystallography and CYP126A1 Structure Determination

Crystals of native, ligand-free CYP126A1 were obtained by mixing 1 μl of 20 mg ml⁻¹ P450 with 1 μl of a reservoir solution containing (v/v) 38% saturated soldium sulfate, 10% 1 M Tris HCl (pH 6.5) and 52% H₂O, and by incubating at 4 °C for 24-48 h. Tetragonal crystals were formed with dimensions approximately 0.9 x 0.3 x 0.2 mm. Prior to data collection, the CYP126A1 crystals were soaked for 5-10 s in a cryoprotectant solution consisting of mother liquor supplemented with 10% polyethylene glycol 200. The crystal was then attached to a CryoLoop and placed in a 100 K nitrogen stream for data collection. Data were scaled and integrated, and the ligand-free structure was solved to a resolution of 1.7 Å
by molecular replacement using Phaser\textsuperscript{58}. Refinement and model building were carried out using ArpWarp, Phenix and COOT in conjunction with Molprobity to validate the structure\textsuperscript{59}. Detailed data and final refinement statistics are given in Table 4.3. The crystals were found in space group P212121, with unit cell parameters $a = 58.9$, $b = 69.9$, $c = 233.4$ Å.

The crystal structure of the CYP126A1 complex with the azole drug ketoconazole was obtained by co-crystallizing the P450 with ketoconazole. CYP126A1 (20 mg/ml) was mixed with 2 μl of a saturated ketoconazole solution. Sitting drops were prepared by mixing 1 μl of ketoconazole-bound CYP126A1 with 1 μl mother liquor (40% sodium sulfate, 0.1 M Tris, pH 7.0). These were incubated at 4 °C and formed small needle shaped crystals. Streak-seeding was then used to generate better quality, rod-shaped crystals in 38% sodium acetate, 0.1 M sodium cacodylate, pH 6.5). Seeded crystals were soaked in cryoprotectant, as above, and flash-cooled in liquid nitrogen for X-ray diffraction experiments.

Structures of CYP126A1 were also determined in complex with two molecules identified by high-throughput compound screening to induce substrate-like (type I) and inhibitor-like (type II) spectral shifts of the CYP126A1 Soret spectrum. Compound 1 produces a type I CYP126A1 (blue) Soret spectral shift, while compound 7 (1-(3-(1H-imidazol-1-yl)propyl)-3-((3s,5s,7s)-adamantan-1-yl)urea) produces a type II (red) shift. Crystals of these CYP126A1 complexes were again obtained by co-crystallization following mixing CYP126A1 with these compounds, as described previously for ketoconazole. The CYP126A1-compound 1 complex formed crystals in 0.1 M carboxylic acids, 0.1 M imidazole-MES pH 6.5 and 30% ethylene glycol-PEG 8K. The CYP126A1-compound 7 complex formed crystals in 0.2 M sodium sulfate, 0.1 M pH 7.5 and 20% PEG 3350. CYP126A1-compound 7 crystals formed small needles, and streak-seeding into the sodium acetate conditions (as done for the ketoconazole complex) resulted in improved CYP126A1-compound 7 crystals within 3 days of incubation at 4 °C. The crystals of both CYP126A1-
compound 7 and -compound 1 complexes appeared were tetragonal, similar to those for the ligand-free CYP126A1 form.

Table 4.3. Data reduction and final structural refinement statistics for the *M. tuberculosis* H37Rv CYP126A1 and drug complexes.

Data are presented for CYP126A1 in its ligand-free form and for CYP126A1 in complex with the azole drug ketoconazole, the type I compound screening hit compound 1 [N-isopropyl-N-((3-(4-methoxyphenyl)-1,2,4-oxadiazol-5-yl)methyl)-2-(4-nitrophenyl)acetamide], and the type II compound screening hit compound 7 [1-(3-((1H-imidazol-1-yl)propyl)-3-((3s,5s,7s)-adamantan-1-yl)urea]. (PDB codes pending).

<table>
<thead>
<tr>
<th>CYP126A1</th>
<th>CYP126A1 ketoconazole</th>
<th>CYP126A1 compound 1</th>
<th>CYP126A1 compound 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁,2₁,2₁</td>
<td>C2</td>
<td>P₂₁</td>
</tr>
<tr>
<td>Cell dimensions a,b,c (Å)</td>
<td>58.93, 70.05, 233.66</td>
<td>118.23, 59.66, 69.84</td>
<td>69.62, 59.66, 118.47</td>
</tr>
<tr>
<td>α,β,γ (°)</td>
<td>90, 90, 120</td>
<td>90, 93.2, 90</td>
<td>90, 98.03, 90</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>0.034 (0.423)</td>
<td>0.048 (0.666)</td>
<td>0.031 (0.842)</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>13.4 (1.7)</td>
<td>19.1 (2.2)</td>
<td>11.9 (1.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (99.0)</td>
<td>92.47 (59.1)</td>
<td>99.5 (96.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.7 (3.9)</td>
<td>3.9 (3.3)</td>
<td>4.0 (3.4)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>46.99-1.80</td>
<td>43.86-1.83</td>
<td>32.14-1.95</td>
</tr>
<tr>
<td>No. reflections</td>
<td>90448 (8427)</td>
<td>37754 (1747)</td>
<td>66993 (4759)</td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td>15.94/19.77 (22.7/30.6)</td>
<td>18.88/23.87 (29.8/37.6)</td>
<td>0.169/20.34 (24.7/29.7)</td>
</tr>
<tr>
<td>No. non-hydrogen atoms</td>
<td>7079</td>
<td>3613</td>
<td>6918</td>
</tr>
<tr>
<td>Mean B factor (Å²)</td>
<td>19.54</td>
<td>19.4</td>
<td>23.46</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td>0.016</td>
<td>0.022</td>
<td>0.018</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>1.444</td>
<td>1.946</td>
<td>1.807</td>
</tr>
</tbody>
</table>

4.5.9 Materials

Azole anti-fungal drugs were purchased from MP Biomedicals Inc. Novel compounds (compounds 1-9) were from the compound library of the Screening Facility at the Department of Medicinal Chemistry, Leibniz Institute of Molecular Pharmacology (Berlin, Germany), or sourced from ChemDiv (San Diego, USA) or Vitas M (Apeldoorn, The Netherlands). Bacterial growth media (2xYT) were from Melford Laboratories (Ipswich, Suffolk UK). Unless otherwise mentioned, all other reagents were purchased from Sigma Aldrich, and were of the highest grade available.
References


52. Etienne, G. et al. The cell envelope structure and properties of Mycobacterium smegmatis mc2(155): is there a clue for the unique transformability of the strain? *Microbiology* 151, 2075-86 (2005).


5 Chapter Five – Summary, Conclusions & Future Prospects

*Mycobacterium tuberculosis* (*Mt*) persists as a major threat to human health even though several medications have been developed to target various physiological pathways of the bacterium. In 1998, investigations into sequencing the entire genome of *Mt* revealed an intriguingly high density of genes coding for twenty P450 enzymes\(^1\). This unusually large number of P450s in a pathogenic bacterium suggested important catalytic roles and hinted at them being utilized for important functions in bacterial metabolism. Various methods have previously been explored to investigate gene expression and regulation in the *Mt* genome. Genes up- or down-regulated under various growth conditions have been identified, which shed light on their functionality and where the resulting enzymes fit into metabolic pathways of *Mt*\(^2-5\). Furthermore, studies on gene disruption have aided the identification of not only important genes that are absolutely essential for bacterial growth and viability, but also other genes whose disruption is detrimental to the growth or to *Mt* survival under adverse conditions, or which are important in pathogenic processes. A notable example involves genes required when *Mt* is stressed after being engulfed by macrophages during attack by the host immune system in the formation of a granuloma\(^6-9\). Some of the *Mt* P450s were also identified as playing important roles from studies involving up/down-regulation of their gene expression under various conditions\(^10,11\).

The first *Mt* P450s to be investigated were CYP51B1 and CYP121A1\(^12-14\). These studies were carried out sixteen years ago (1999) and showed that CYP51B1 was a sterol demethylase similar to its counterparts from other organisms, although of uncertain function in *Mt* given the absence of other enzymes of a sterol biosynthetic pathway\(^12\). In the case of CYP121A1, the exact function was not proven until ten years later (2009) when it was shown
to catalyse the formation of a novel aromatic ring coupled product (mycocyclosin) from the cyclic dipeptide cyclo-L-tyrosine-L-tyrosine (cYY)\textsuperscript{15}. The crystal structures of both CYP51B1 and CYP121A1 have been solved and both P450s were shown to bind azole antifungals with high affinity\textsuperscript{16,17}. These studies paved the way for other \textit{Mtb} P450s to be investigated and provided supporting evidence that these enzymes may be involved in important metabolic pathways towards \textit{Mtb} growth and virulence, and that they could be novel drug targets for therapeutic investigations. Various \textit{Mtb} P450s have since then been added to the list of possible drug target enzymes in the bacterium. Among these are CYP124A1 (a cholesterol and methyl branched chain lipid hydroxylase), CYP125A1 (a cholesterol hydroxylase) and CYP130A1 (of uncertain function), which have been extensively characterized and for with their crystal structures show characteristic inhibitory azole binding modes in line with CYP51B1 and CYP121A1 data\textsuperscript{18-20}. Another \textit{Mtb} P450, the prospective menaquinone hydroxylase CYP128A1, was also shown to be required for optimal growth of \textit{Mtb}, but difficulties in expressing this enzyme in a soluble form have hindered efforts to characterize this enzyme structurally and biochemically\textsuperscript{2,21}. Nevertheless, these data further highlighted \textit{Mtb} P450s as potential drug targets.

Other P450s have also emerged from the list as possible drug targets, including CYP144A1 and CYP126A1, which are the enzymes for which structural and enzymatic data are presented in this thesis. My investigations into CYP144A1 and CYP126A1 have highlighted important structural features of these \textit{Mtb} P450s and how there are evidently alternative transcription mechanisms for producing different variants of the same P450 from the same gene, as well as further highlighting that these P450s are good drug target enzymes in the bacterium.
CYP144A1 emerged as a potential drug target in *Mtb* after it was proposed to be upregulated under macrophage-like conditions\(^{22}\), suggesting its involvement in the latent phase of *Mtb* growth, which is a stage not targeted by current medications. Furthermore, our bioinformatics analysis into the phylogeny of CYP144A1 ancestry revealed this enzyme to be exclusively found in several pathogenic mycobacteria, which suggest an important role in the pathogenicity of *Mtb*. We then proceeded to *in vitro* production of this enzyme for further studies. Initially, the expression of CYP144A1 was challenging and often resulted in very low yields (0.2 mL of 20 mg/mL from 20 L growth) compared to other recombinant *Mtb* P450 systems (1 mL of 50 mg/ml for CYP126A1). However, CYP144A1 was purified to homogeneity and proven to be a cytosolic enzyme. Previous studies on CYP144A1 reported this enzyme to display spectroscopic properties similar to those of other reported P450s. The heme iron of the purified CYP144A1 is in the ferric form in its resting state, and a cysteinate- ligated hemoprotein was demonstrated using RR and EPR spectroscopy. Redox potentiometric studies showed CYP144A1 to have a midpoint reduction potential for the heme iron Fe\(^{III}/Fe^{II}\) couple at -355 mV (versus NHE) in the substrate-free form\(^{23}\), which is slightly positive for a ligand-free P450, but similar to the potentials reported for other previous important *Mtb* P450s\(^{14,20,24}\). In addition, gene knock-out studies using the virulent *Mtb* H37Rv strain showed retarded growth of the \(ΔCYP144A1\) variant, as well as demonstrating significantly poorer growth than the WT strain in the presence of the azole drug clotrimazole\(^{23}\). This further increased our interest in this enzyme as a possible drug target and thus I pursued further investigations in order to characterize this P450. In chapter one, I presented data on the structural characterization of this enzyme. Attempts to crystalize intact CYP144A1 were largely unsuccessful, with crystals diffracting poorly to ~3.5 Å. Another problem associated with CYP144A1 purification was an observable clipping of the enzyme, which was evident on SDS-PAGE gels. This raised questions as to the stability of
CYP144A1 and whether the proteolyzed form might actually be a physiologically relevant form of the P450. In support of this hypothesis, structural alignments of this enzyme with CYP144 orthologues in other mycobacteria always produced the best alignment with a second potential start codon 30 amino acids downstream from the site annotated in the Tuberculosis database, hinting at the possibility of a second, alternative, transcriptional start site for CYP144A1. Further supporting these data, our results from Mtb H37Rv transcriptomic analysis revealed two alternative versions of CYP144A1 – a full length form and a 30 amino acid “truncated” form, which are referred to in this thesis as the FLV and TRV forms of CYP144A1. Alternative transcriptional start sites and leaderless transcripts play important roles in diversifying protein production in microbial genomes. Transcriptomic studies on the M. tuberculosis H37Rv genome have also shown that about 25% of transcripts from this organism are leaderless.

The CYP144A1 TRV) version of the enzyme was produced from the appropriate second methionine codon and both the FLV- and TRV-CYP144A1 proteins were expressed and characterized in detail to compare their biophysical, spectroscopic and ligand-binding properties. Our data indicated that both CYP144A1 versions had similar biophysical properties. Both CYP144A1–FLV and –TRV displayed thiol-coordinated P420 species when bound to carbon monoxide in their reduced (ferrous) forms. No evidence for any significant amount of a thiolate-coordinated P450 species was obtained. The exact reason for the cysteinate heme protonation from ferric thiolate to the ferrous thiol form in the Fe^{II}-CO complex is not yet known. Further investigations of this CYP144A1 transition process should be done in various buffer systems as well as by analysing the effects of buffer pH in efforts to find conditions conducive to retention of the P450 state. Other factors, such as the nature of residues in the environment surrounding the active site of CYP144A1, may also affect the protonation state of the cysteine ligand in the reduced form. I further probed the effect of
P450 truncation on the heme environment of the CYP144A1-TRV protein using EPR. In previous studies with the full length CYP144A1, we demonstrated a characteristic low spin X-band EPR rhombic spectrum for the ligand-free form, confirming that it exists in a ferric state and has cysteine thiolate coordination in this form\textsuperscript{23}. Our comparative EPR analysis of both CYP144A1 versions showed that they exhibit similar electronic properties by EPR, indicating that the truncation had no significant effect on the heme coordination state of this P450 enzyme. Thermal stability studies with DSC on both ligand-free and azole-bound forms also showed that both the CYP144A1-TRV and -FLV enzymes are more stable when bound to an azole inhibitor. Our bioinformatics analysis of both protein forms, however, indicated a significant level of structural disorder residues within the initial 30 amino acids in the CYP144A1-FLV, thus confirming that the CYP144A1-TRV enzyme should generate a more stable P450 form. This was confirmed during crystallographic studies of both CYP144A1 forms, with the TRV form being more readily crystallizable than the FLV form. It seems likely that the N-terminal 30 amino acids in the CYP144A1-FLV form are largely unstructured and detrimental to the crystallizability of the protein. At this stage it is difficult to completely rule out the importance of these N-terminal 30 amino acids, as the physiological substrate of CYP144A1 has not yet been identified and further interrogation of the interactions of a \textit{bona fide} substrate with both the CYP144A1-FLV and -TRV enzymes would be highly informative as to influence of the N-terminal “extension” on the enzymes’ affinity and activity profiles. The first structure of a CYP144A1 enzyme was thus solved by X-ray crystallography using molecular replacement with the \textit{Mtb} cholesterol oxidase CYP142A1. The structure revealed features similar to other P450s. CYP144A1 was shown to have a relatively large active site cavity primed for accommodating large hydrophobic substrates, based on its predominantly hydrophobic active site residue composition.
The determination of the crystal structure of CYP144A1-TRV paves the way for rational drug design for this enzyme. In chapter two of this thesis I presented detailed chemoproteomic profiling along with biophysical characterizations of candidate ligands for CYP144A1. Structural comparisons of CYP144A1 with other *Mtb* P450s had shown that this enzyme is structurally similar to CYP121A1, based on active site cavity size and amino acids similarity score. Thus, ligand profiling of CYP144A1, in part, involved imidazole-containing polyphenol ligands (Compounds 1, 2 & 3) that had previously been developed for CYP121A1. These ligands were probed for binding to CYP144A1 using a combination of spectroscopic and biophysical techniques, and demonstrated various binding modes to the P450. A variety of azole antifungals were shown to have high affinity for CYP144A1, with the highest affinity being for clotrimazole and econazole, while there was lower affinity for the triazoles (itraconazole and voriconazole). These data also point to the shallow but wide active site cavity of this enzyme, where tighter binding is achieved with more compact and bulky compounds and where there is lower affinity for narrow elongated compounds. The CYP121A1 polyphenol compounds demonstrated inhibitor-like binding similar to those of the azole antifungals. However, the binding of one of these compounds (Compound 1) to CYP144A1-TRV resulted in a substrate-like heme perturbation, with a type-I spectral shift typical of a P450 substrate-bound complex. Although this seemed promising in terms of identifying a novel substrate for this enzyme, turnover experiments indicated no significant oxidative transformation of Compound 1 by CYP144A1. This could be due to the particular binding mode of this molecule, which perhaps places the chemical too far from a reactive CYP144A1 compound I species for efficient turnover. Other fragment molecule scaffolds were also probed for binding to CYP144A1-TRV, but only a few positive hits emerged, with relatively weak binding affinities when compared to the fragment screening results on other *Mtb* P450s (unpublished data). This further highlights the particular specificity of this
enzyme as dictated by the structure of its active site. Further biophysical studies were done to investigate ligand-binding stoichiometry using NanoESI-MS. The results indicated the active site of CYP144A1 to be large enough to accommodate two molecules of some of the CYP121A1-binding compounds and some azoles, albeit at higher ligand concentrations in the latter case. However, these studies served to demonstrate the large size of the active site pocket size of CYP144A1 and its likely evolutionary design for large hydrophobic substrates.

Further work on CYP144A1 should most definitely start with identifying a physiological substrate. This could be achieved through a number of approaches. Compound library screening may identify a possible substrate based on type I heme spectral shift data. While this may not be a physiological substrate, the data should definitely hint towards a true substrate’s general chemical structure. Upon identification of such substrates from compound screening, further structural studies can be performed to gain insight into key residues that are important for substrate binding and catalysis. This approach has been utilized to identify heterocyclic arylamines as a new group of inhibitors for CYP130A1\textsuperscript{26}. Other approaches could be to test a variety of \textit{Mtb} metabolites using a combination of UV/Vis and LCMS techniques to check for type I binding shifts and for evidence of substrate turnover, respectively.

Investigations into the \textit{Mtb} P450 CYP126A1 as a potential drug target were initiated due to its localization in a gene cluster containing essential \textit{Mtb} genes encoding enzymes for the \textit{de novo} anabolism of purines. CYP126A1 is positioned within an operon that contains a confirmed adenylosuccinate lyase enzyme (PurB), which raised anticipation that CYP126A1 may also play a crucial role in \textit{Mtb} purine metabolism\textsuperscript{27}. Bioinformatics analysis also showed that CYP126A1 shares considerable sequence homology (35\% identity) to the \textit{Mtb} cholesterol hydroxylases CYP125A1 and CYP124A1, and that it is also widely conserved.
across strains of both pathogenic and non-pathogenic actinobacteria. This unquestionably raises confidence in its probable involvement in important biochemical functions. Previous studies that aimed to predict the biological function of CYP126A1 based on its productive binding to compounds from a virtual TB metabolome (substrate-like, type I spectrum-inducing biofragments) indicated that this P450 showed binding preference for aromatic molecules, notably those with chlorophenol scaffolds. These data may suggest CYP126A1’s potential involvement in oxidative dehalogenation of such substrates.

In chapter four, I present a detailed biophysical and structural characterization of azole compound binding and fragment screening studies with CYP126A1. These data also indicated that this enzyme is highly soluble and located in the cytosol of the E. coli expression cells. After purification, our initial UV-Vis spectroscopic analysis indicated that various type I (substrate-like) and type II (inhibitor) compounds from our HTS and fragment screening studies bound to CYP126A1 with a wide range of affinities, suggesting that CYP126A1 may have capability to bind and possibly oxidize a wide variety of unnatural substrates. In comparison, CYP144A1 has limited ligand binding variety and high specificity for particular compounds. X-band EPR spectroscopic analysis of CYP126A1 in ligand-free and ligand-bound forms produced typical low spin spectra when bound to inhibitors, and evidence for retention of high spin spectral features when bound to substrate-like compounds, despite the very low temperatures used for hemoprotein EPR (10 K). These properties are consistent with those reported for other characterized Mtb P450s. CYP126A1 also demonstrated some capability to bind two molecules of one of the HTS ligands (Compound 1) simultaneously in its active site pocket, though this occurred at a higher ligand-to-protein ratio than was used for CYP144A1, further supporting the conclusion that the CYP144A1 enzyme may be “primed” for catalysis with multiple substrates of various sizes.
In order to further understand the active site and general structural morphology of CYP126A1, the P450 was crystallized in both ligand-free and ligand-bound forms with various azoles, compounds from high-throughput screening. CYP126A1 was crystallized as a dimer in an open/closed state (with respect to active site entry conformation) as was previously observed for the heterocyclic arylamine-binding CYP130A1\(^{26}\). The ligand-bound forms of CYP126A1 revealed various residues involved in substrate and inhibitor interactions. Gly254, Ala253, Asn96 and Thr257 are involved in hydrophobic interaction with one of the substrate-like HTS hits (\textbf{Compound 1}). Ser300 and Asn96 showed hydrophobic interactions with the azole inhibitor ketoconazole and with \textbf{Compound 7}. There is apparent retention of Asn96 Thr83, Ser300, Lys303, Arg400, and Arg401, in interactions with each of these ligands. These amino acids are thus potential key residues in CYP126A1 that play crucial roles in modulating substrate and inhibitor interactions.

These investigations into CYP144A1 and CYP126A1 have significantly advanced our understanding of the properties of the \textit{Mtb} P450s and have contributed to our general understanding of P450s as a whole. This is true particularly with respect to identifying the ligand specificity for these P450s, including their interactions with azole drugs and compounds identified from fragment screening. Other important insights have come from X-ray crystallography, as well as from the application of spectroscopic and biophysical methods including EPR and mass spectrometry. Future studies of \textit{Mtb} P450s will undoubtedly uncover novel properties of these enzymes, and in so doing should focus efforts on those isoforms that are most attractive as drug targets. Our ongoing studies on fragment screening have already identified several molecules that bind tightly to one or more of the \textit{Mtb} P450s. This provides us with an important lead in this area towards identification of substrate-like molecules and to advance the development of more effective therapeutics against \textit{Mtb}.
References


