Cytochrome P450 monooxygenase analysis in the genus *Mycobacterium*: Special focus on CYP123

By

Mohammad Parvez

Dissertation submitted in fulfilment of the requirements of the degree

Doctor of Health Sciences in Biomedical Technology

Department of Health Sciences

Central University of Technology, Free State

15 May 2018

Supervisor: Prof Khajamohiddin Syed

Co-supervisor: Prof Samson Sitheni Mashele

DECLARATION

I, **MOHAMMAD PARVEZ** (INDIAN PASSPORT NUMBER), hereby certify that the dissertation submitted by me for the degree DOCTOR OF HEALTH SCIENCES IN BIOMEDICAL TECHNOLOGY, is my own independent work; and complies with the Code of Academic Integrity, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology (Free State). I hereby declare, that this research project has not been previously submitted before to any university or faculty for the attainment of any qualification. I further waive copyright of the dissertation in favour of the Central University of Technology (Free State).

I also state that expression vector modified/generated in this study is in collaboration with my colleagues, **Mr HANS DENIS BAMAL** (student number:) and **Ms IPELENG KOPANO ROSINAH KGOSIEMANG** (student number:).

MOHAMMAD PARVEZ **DATE**

This thesis is dedicated to my

Father the late

ACKNOWLEDGEMENTS

This thesis would have remained a dream, had it not been for my supervisor, Prof Khajamohiddin Syed, who continuously supported me during my doctoral study. I would like to extend my gratitude to him, for his patience, motivation, enthusiasm, and immense knowledge in P450 research. I would also like to thank him for giving me the opportunity to attend national and international scientific conferences. He has shown great faith and trust in me throughout my studies and without such attributes I may not be where I am now. I would also like to thank Prof Samson Sitheni Mashele for his support, and always listening to our concerns.

Thanks to the National Research Foundation, South Africa and the Research Innovation Fund, Central University of Technology, Free State for funding my studies and enabling dreams to be realised.

Thanks also to my fellow lab mates for the stimulating discussions, for the hard work we put in towards our projects, and for all the fun we have had in the last three years.

Special thanks to my mother, SHAMEEM, who always believed in me, who always supported me through difficult times, and for her sympathetic ears. Thank you again, if it wasn't for you none of this would have been possible. To my sisters, Tarrnum, thank you for your prayers and moral support throughout these years.

In addition, would like to thank my late father ABDUL KAREEM, for raising me, giving me the love, and also guiding my every step. Thanks to my mother shameem, for always having his home open for me and my laptop, and also encouraging me to pursue my dreams.

I would also like to thank, ARIF MIADAD for providing me with unfailing support, continuous encouragement, and his patience throughout my years of study and through the process of research and writing this thesis.

Above all, I owe it to the Almighty ALLAH for granting me the wisdom, health and strength to undertake this research task and enabling me to its completion.

“When ALLAH pushes you to the edge of difficulty, trust Him fully because two things happen either He will catch you when you fall or He will teach you how to fly.”

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

±	Plus –minus
≥	Greater or equal to
°C	Degree Celsius
3D	Three-dimensional
Å	Angstrom
AIDS	Acquired immunodeficiency syndrome
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
C	Carbon
C-C	Carbon-carbon bond
cDNA	Complementary DNA
ClustalW2	Multiple sequence alignment program
CO	Carbon monoxide
C-O	Carbon-oxygen bond
CPR	Cytochrome P450 reductase
CYP or P450	CytochromeP450
cYY	Cyclo-L-Tyr-LTyr
dDFIRE	Updated energy function of DFIRE
DFIRE	Distance-scaled, finite ideal-gas reference
DFIRE2	Updated energy function of DFIRE

DNA	Deoxyribonucleic acid
FAD	Flavin adenine dinucleotide
FASTA	File format for DNA and protein sequences
FDR	Ferrodoxin reductase
FdR	Ferredoxin reductase
FDX	Ferrodoxin
Fdx	Ferredoxin
Fe^{2+}	Iron (II) cation
Fe-S	Iron-sulphur
FMN	Flavin mononucleotide
HEM	Heme group
HIV	Human immunodeficiency virus
HMMER	biosequence analysis using profile hidden Markov models
ID	Identity
JGI	Joint Genome Institute
kDa	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
MAC	*M. avium* complex
MCAC	*M. chelonae-abscessus* complex
MCL	Mycobacteria causing leprosy

MCS	Multiple cloning site
MDR-TB	Multi-drug resistant TB
MEGA	Molecular evolutionary genetic analysis
MIT	Massachusetts Institute of Technology
MOE	Molecular Operating Environment
MTBC	*Mycobacterium tuberculosis* complex
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
nm	Nanometre
NMR	Nuclear magnetic resonance
NO	Nitric oxide
NS	New subfamily
N-terminal	Amino terminal end
NTM	Nontuberculous mycobacteria
O-	Substitution of oxygen atom
PDB	Protein Data Bank
pDRAW	DNA analysis software
PROMALS3D	PROfile Multiple Alignment with predicted Local Structures and 3D constraints
RH	Substrate

RMSD	Root mean square derivative
R-OH	Hydroxylated product
S-	Substitution of sulphur atom
SAP	Saprophytes
TB	Tuberculosis
TDR-TB	Totally-drug resistant TB
WHO	World Health Organization
XDR-TB	Extensively-drug resistant TB
µ	Micro

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABSTRACT

One-third of the human population is infected by latent tuberculosis (TB). Active TB can be cured with drugs unless it is drug-resistant. The study showed that P450 CYP123A1 is highly expressed during the latent phase of *M. tuberculosis*, the infectious agent that causes TB. Based on meta-analysis of the expression data, CYP123A1 is selected as the best drug candidate against the dormant phase of *M. tuberculosis*. It is noteworthy that CYP123A1 is present only in TB-causing bacteria, suggesting its essential role.

Despite the grate importance of CYP123A1 as drug target for both latent TB and active TB, to date this P450s has not been characterised for its function or for its structure. This study is the first of its kind on *M. tuberculosis* P450 CYP123 characterisation. Furthermore, the distribution of P450s in the genus *Mycobacterium* is unknown. Thus, in this study, the molecular evolutionary dynamics of P450s were examined with special focus on mycobacterial P450s.

In this study, 17 598 P450s belonging to 113 P450 families (bacteria − 42; fungi − 19; plant − 28; animal − 22; plant and animal − 1 and common P450 family − 1) were analysed and highly conserved and rapidly evolving P450 families was identified. The results suggested that bacterial P450s, particularly P450s belonging to mycobacteria, are highly conserved both at protein and DNA levels. Mycobacteria possess the highest P450 diversity percentage compared to other microbes and have high coverage of P450s (≥1%) in their genomes, as found in fungi and plants. Phylogenetic and functional analyses revealed the functional conservation of P450s despite belonging to different biological kingdoms, suggesting the adherence of P450s to their innate function, such as their involvement in either generation or oxidation of steroids and structurally related molecules, fatty acids and

terpenoids. This study's results offer new understanding of the dynamic structural nature of P450s.

In silico structural analysis of CYP123A1 is carried out where a 3D homology model of CYP123A1 is constructed using the template 3A4G. The 3D model of CYP123A1 was found to be of high quality, based on validation programmes. The docking of model CYP123A1 with the ligands clotrimazole, econazole, fluconazole, itraconazole, ketoconazole, miconazole, posaconazole and voriconazole was performed in Autodock vina. The molecular docking studies showed that ketoconazole forms a much better complex than other azole drugs, with the best interaction rate and lowest energy of -9.0 kcal/mol. ketoconazole forms a tight hydrogen bond with residue Lys246 of CYP123A1. Clotrimazole did not show interaction with the 3D model of CYP123A1. Lys246 forms a hydrogen bond with fluconazole and heme forms two Pi-Pi stacking with fluconazole. Voriconazole is involved in one Pi-Cation interaction and hydrogen bond with the CYP123A1 residues Arg82.

In order to validate the *in silico* studies, *CYP123A1* was cloned and recombinant cells carrying *CYP123A1* in a novel expression vector was generated. In conclusion, the multiple cloning site of the pINK-A vector was successfully modified by adding more suitable restriction enzymes and the revised vector was named pINK-d. Sixteen restriction enzymes were added in such a way that they did not cause any shift or change in the reading frame of the vector. The new vector can be used in future for the cloning of other P450s. CYP123A1 cDNA and the modified expression vector were successfully synthesised by GenScript. Synthesised *CYP123A1* was further cloned into expression vector pINK-d. The construct containing the CYP123A1 gene and the empty pINK-d vector were transformed into *E. coli* cells and recombinant cells were selected on Luria-Bertani ampicillin plates. Plasmids were isolated and the presence of the correct size insert (*CYP123A1*) was verified by restriction

enzyme digestion analysis. Recombinant *E. coli* cells carrying *CYP123A1* and an expression vector were stored at -80^{o}C for future expression and functional analysis of CYP123A1.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1. *Mycobacterium tuberculosis*

The actinomycete *Mycobacterium tuberculosis* causes tuberculosis (TB), a chronic lung disease, in humans and remains one of the greatest threats to mankind. *M. tuberculosis* was discovered by Robert Koch and described in a well-known lecture on 24 March 1882 (Koch, 1882). *M. tuberculosis* is a rod-shaped bacterium and grows as sticky, waxy lumps on the petri dish (Figure 1.1).

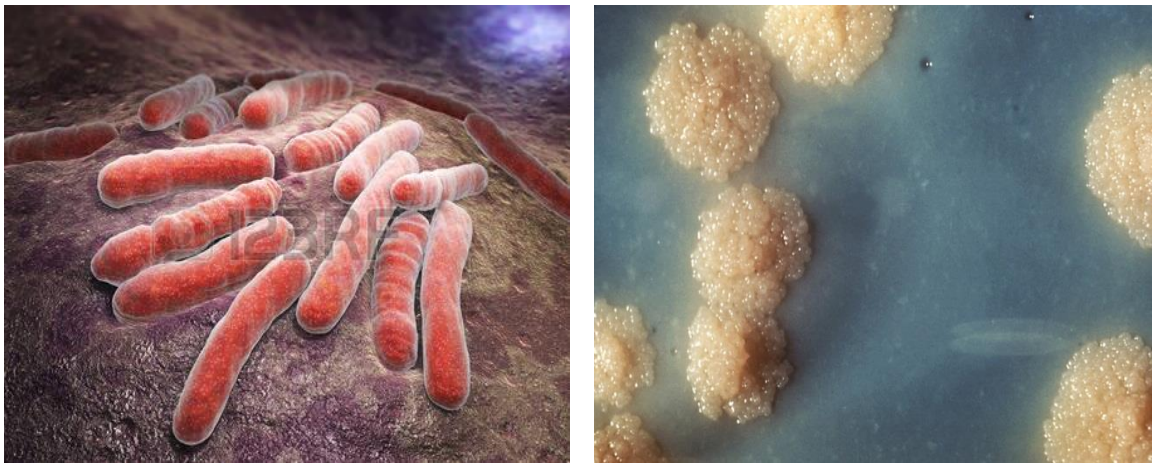

Figure 1.1. *M. tuberculosis* appearance as rod-shaped bacterium (left panel) and sticky growth on plates (right panel) (pictures taken from http://atccmicrobiology.blogspot.co.za/2014/04/the-need-for-tuberculosis-reference.html).

The unusual waxy coating of the *M. tuberculosis* cell surface is primarily due to the presence of mycolic acids in its membrane. A study by Bansal-Mutalik and Nikaido (2014) sheds light on the membrane composition of *M. tuberculosis*. The authors discovered that the outer leaflet of the outer membrane contains a similar number of hydrocarbon chains as the

inner leaflet composed of mycolic acids covalently linked to cell-wall arabinogalactan (Figure 1.2). The inner membrane contains one-half of the hydrocarbon chains contributed by an unusual lipid, diacyl phosphatidylinositol dimannoside. The inner leaflet of this membrane is probably composed almost entirely of this lipid. The proposed model of the *M. tuberculosis* cell envelope is presented in Figure. 1.2.

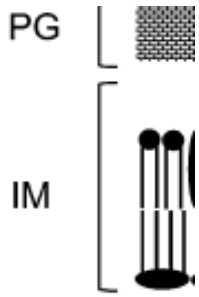

Figure 1.2. A probable model of a mycobacterial cell envelope (taken from Bansal-Mutalik and Nikaido, 2014). Based on lipid composition data obtained by the authors, a model of the outer membrane (OM) and inner membrane (IM) is proposed.

Based on the membrane structure and lipid composition of the outer membrane and inner membrane, the authors proposed that the bilayer environment of *M. tuberculosis* creates unusually low fluidity and may slow the influx of drugs, contributing to the general drug resistance phenotype of mycobacteria (Bansal-Mutalik and Nikaido, 2014).

1.2. *M. tuberculosis*: mode of infection

M. tuberculosis is carried in airborne particles, called droplet nuclei, of 1– 5 microns in diameter. Infectious droplet nuclei are generated when people who have pulmonary or laryngeal TB disease cough, sneeze, shout or sing. Depending on the environment, these tiny particles can remain suspended in the air for several hours. *M. tuberculosis* is transmitted through the air, not by surface contact. Transmission occurs when a person inhales droplet nuclei containing *M. tuberculosis*, and the droplet nuclei traverse the mouth or nasal passages, upper respiratory tract and bronchi to reach the alveoli of the lungs (Figure 1.3).

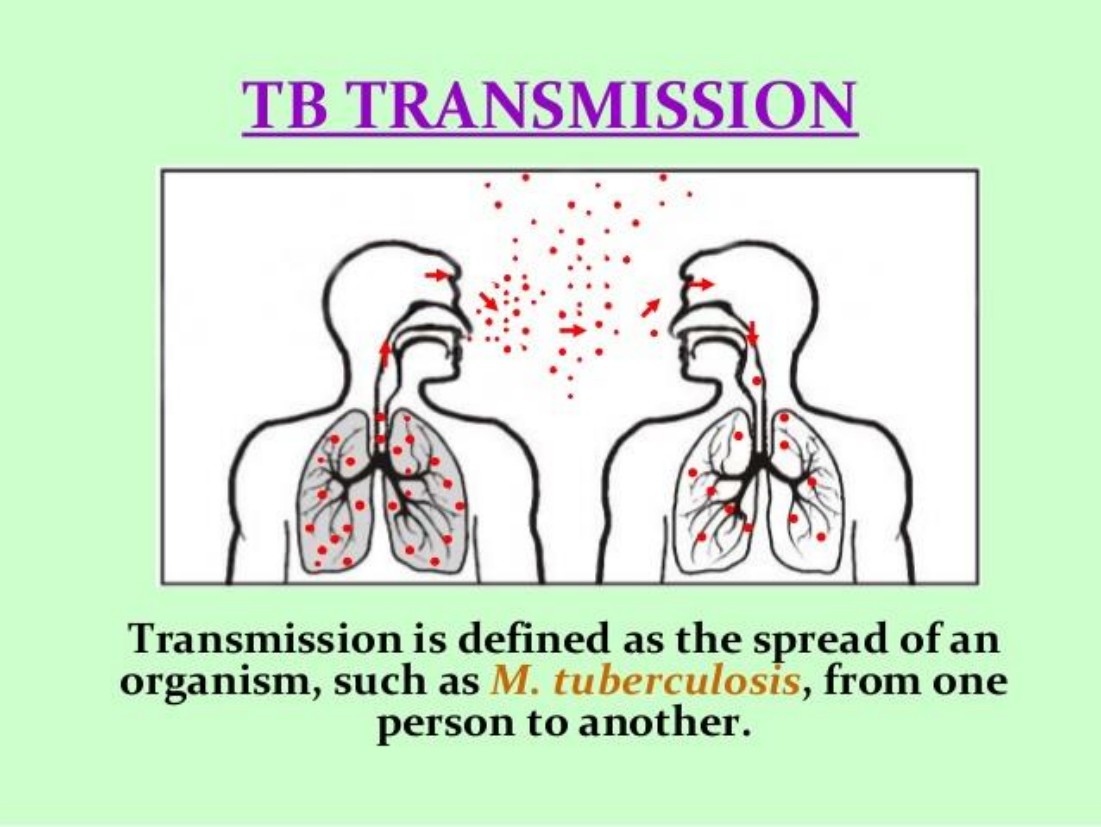

Figure 1.3. Transmission of *M. tuberculosis* (taken from Bates, 1980).

1.3. *M. tuberculosis*: life cycle

Exposure to *M. tuberculosis* can result in the elimination of the pathogen, either because of innate immune responses or because of acquired T cell immunity (Pai *et al*., 2016) (Figure 1.4) or depending on changes in host immunity and comorbidities, *M. tuberculosis* can be in active form (active TB) that may lead to TB disease or stay in the latent form (latent TB) (Figure. 1.4). It may also advance or reverse between the two stages, i.e. active TB and latent TB (Pai *et al*., 2016).

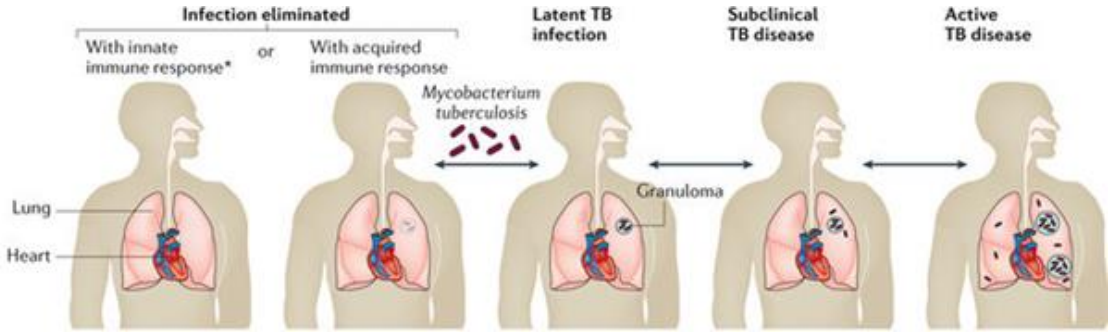

Figure 1.4. The spectrum of TB (taken from Pai *et al*., 2016).

1.4 TB: global pandemic

TB, a pre-historic disease (Bunyan, 1988; Zink *et al*., 2003), remains a leading infectious cause of death worldwide, despite global efforts in disease-control programmes during the past 20 years (WHO, 2016; Quan *et al*., 2017; TB Alliance). TB is a global disease, found in every country in the world (Figure 1.5.). The World Health Organisation (WHO) estimates that two billion people, one third of the world's population, are infected with *M. tuberculosis*. Each year, 10.4 million fall ill from TB and 1.8 million die (WHO, 2016; Quan *et al*., 2017; TB Alliance). TB is a leading killer of people living with HIV, causing one quarter of all deaths. Co-occurrence of TB with HIV/AIDS has further worsened the treatment of TB or

HIV/AIDS and many patients are dying because of TB rather than the initial viral infection (Gandhi *et al*., 2006; WHO, 2014) (Figure 1.6). Death rates due to TB in sub-Saharan Africa, especially in South Africa, are alarming and need immediate attention. The prevalence of HIV/AIDS in the region worsens the situation (Gandhi *et al*., 2006; WHO, 2016; Quan *et al*., 2017; TB Alliance) (Figure 1.7).

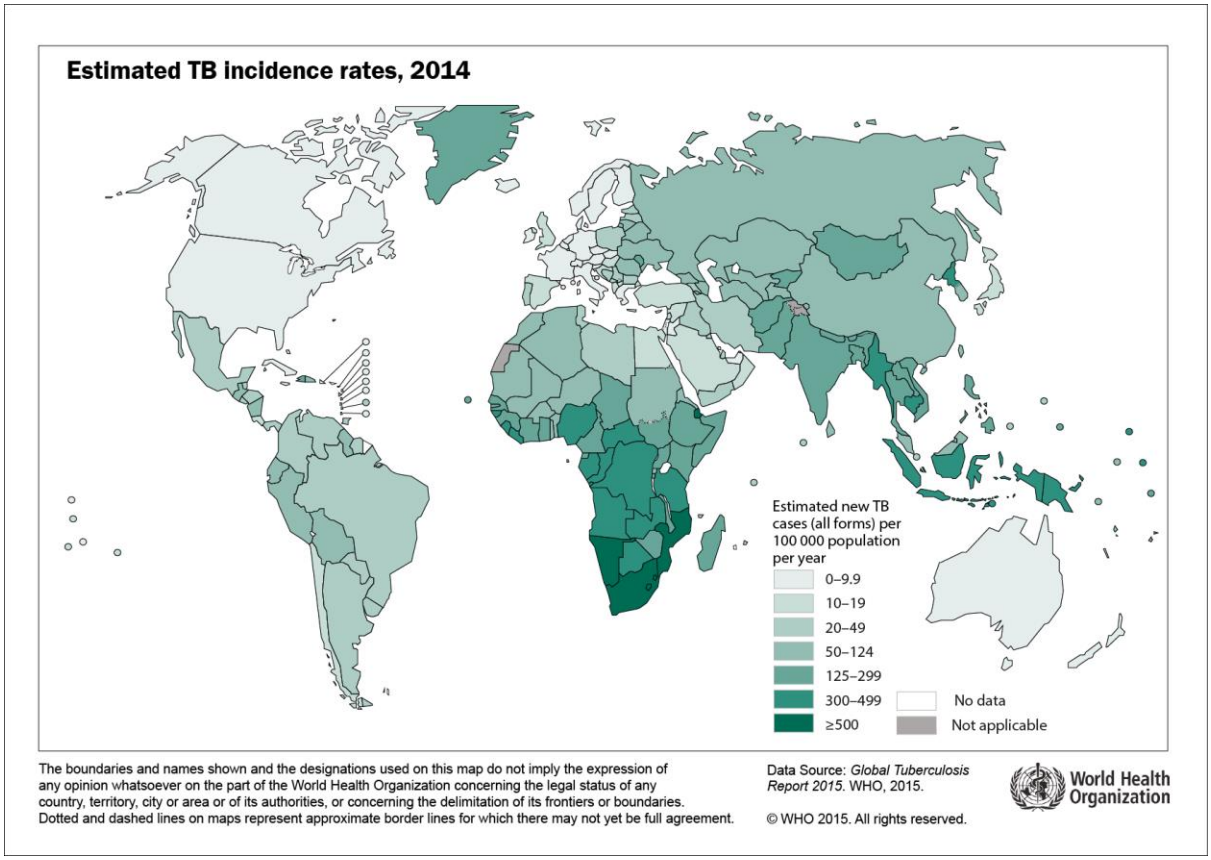

Figure 1.5. World-wide distribution of new TB cases (WHO, 2015).

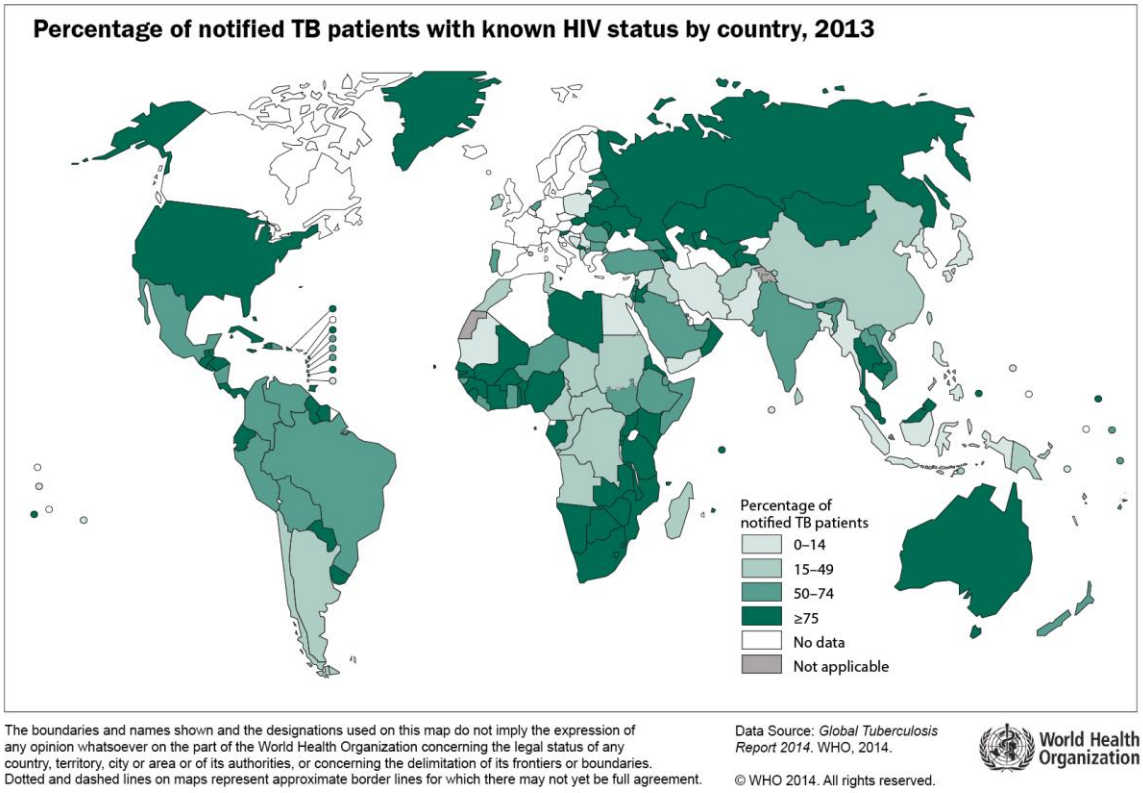

Figure 1.6. World-wide distribution of TB patients with HIV (taken from WHO, 2014).

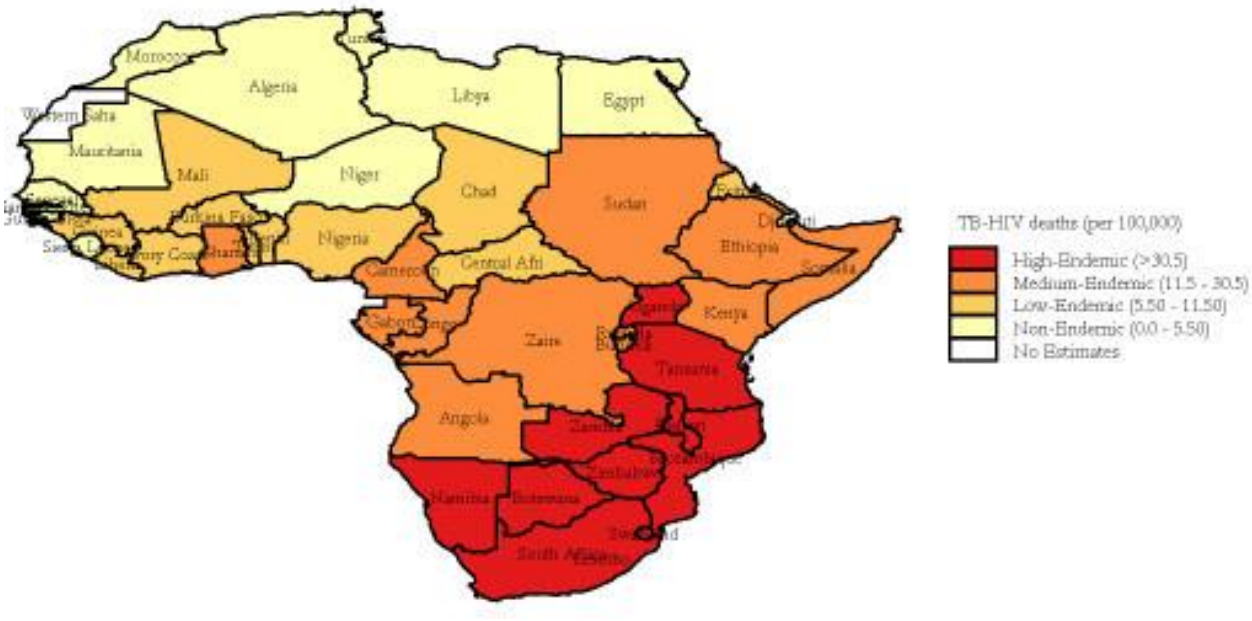

Figure 1.7. Tuberculosis in Africa (taken from

https://sites.google.com/a/bluedemons.org/tuberculosis-in-africa/).

1.5. Effect of TB in South Africa

According to WHO, 8.8 million new cases of TB worldwide are reported in a year (Floyd *et al*., 2011). South Africa is one of the countries reporting a high number of cases, with an incidence of 971 per 100 000. This means that one percent of the population develops TB each year. Fifteen percent are children and two thirds will be co-infected with HIV. The presence of drug-resistant bacteria in South Africa, including multidrug-resistant (MDR) strains, defined as resistant to rifampicin and isoniazid, has been reported (Jackson *et al*., 2013). According to the WHO, in 2009 about 600 cases of extensively drug-resistant (XDR) TB were diagnosed. Most cases of MDR-TB and XDR-TB in South Africa have been detected in the Western Cape, the Eastern Cape and KwaZulu-Natal (Table 1.1).

Table 1.1. The incidence of multidrug-resistant and extensively drug-resistant TB by province in 2009 (taken from Jackson *et al*., 2013).

Province	Number of cases of multidrug-resistant tuberculosis	Number of cases of extensively drug-resistant tuberculosis
Eastern Cape	1 858	123
Free State	123	3
Gauteng	1 307	65
KwaZulu-Natal	773	254
Limpopo	204	6
Mpumalanga	446	18

North West	520	13
Northern Cape	631	40
Western Cape	2 078	72

1.6. TB: treatment and problems

The currently available anti-TB drugs (Table 1.2) were developed over 40 years ago and they have become ineffective against drug-resistant TB. Long-term TB treatment (6-12 months) causes serious side effects and promotes the development of drug-resistant TB (WHO, 2011; Abubakar *et al*., 2013) (Figures 1.8 and 1.9).

Table 1.2. Currently available anti-TB drugs (Rendon *et al*., 2016)

Grouping of drugs	**Drugs line category**
First-line anti-TB drugs	
Group 1	Isoniazid, rifampicin, ethambutol, pyrazinamide
Second-line anti-TB drugs	
Group 2	Moxifloxacin, high dose levofloxacin (fluoroquinolones)
Group 3	Linezolid, delamanid, bedaquiline (newer drugs with increased evidence)
Group 4	Amikacin, capreomycin, kanamycin (injectables)
Group 5	Clofazimine, ethionamide/prothionamide, carbapenems (imipenem, meropenem, ertapenem)
Group 6	Cycloserine, para-amino salicylic acid, amoxicillin/clavulanate

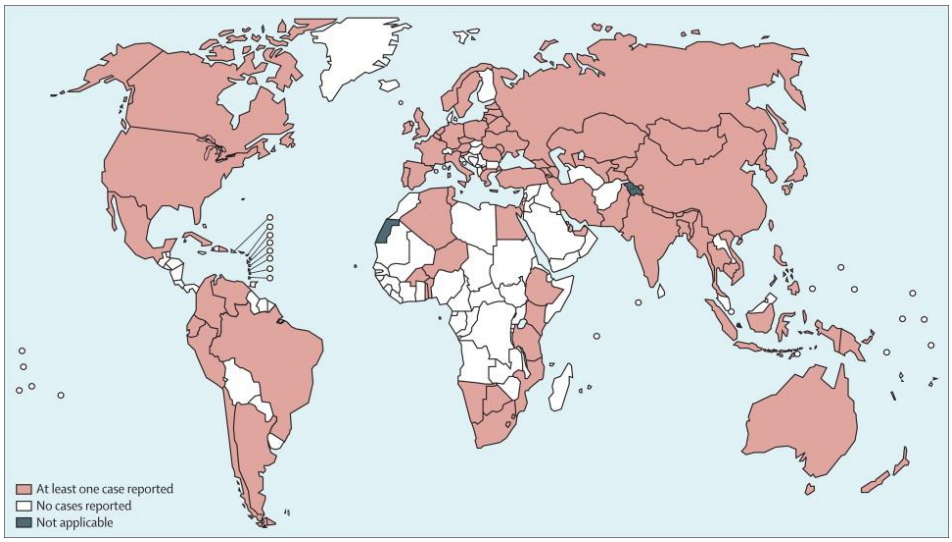

Figure 1.8. Countries with at least one case of extensively drug-resistant TB (taken from Abubakar *et al*., 2013).

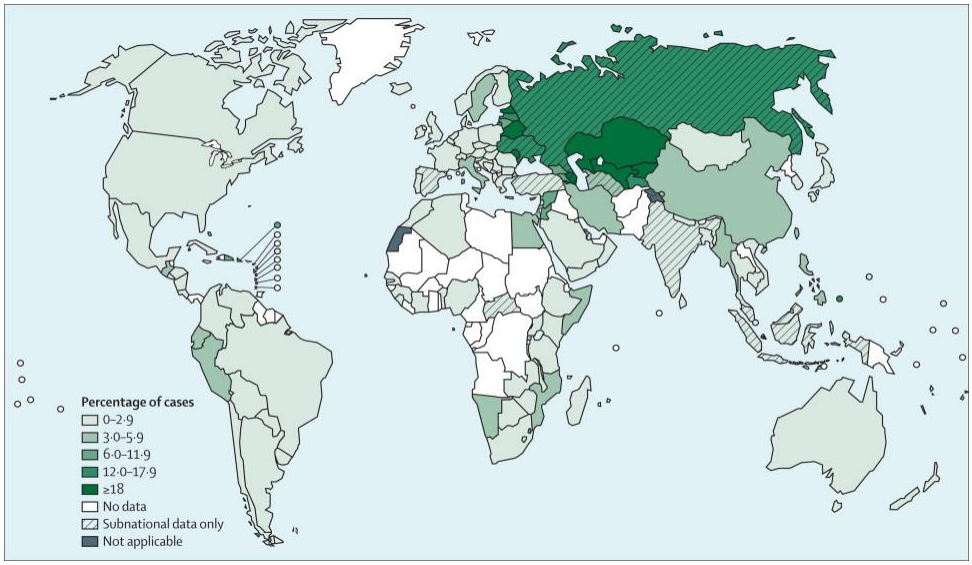

Figure 1.9. Percentage of new TB cases with multidrug-resistant TB (taken from Abubakar *et al*., 2013).

Since 1990, about 10 new drugs have been in clinical trials (Table 1.3). These drugs are relatively safe, have potent bactericidal effectiveness and early bactericidal activity, resembling that of isoniazid.

Table 1.3. New anti-TB drugs in development (Neurmberger *et al.*, 2010; Field *et al.*, 2012).

Drug	**Class**	**Mechanism of action**	**Target**
Bedaquiline	Diarylquinoline	Inhibits ATP synthesis	Subunit C in F_o proton ATP synthesis
PA-824	Nitroimidazopyran	Inhibits protein synthesis, inhibits cell wall lipid synthesis and releases nitric oxide	The production of mycolic acid, protein and intracellular NO
Delaminid (OPC-67683)	Nitroimidazol oxazole	Inhibits lipid synthesis, cell wall synthesis, inhibits protein synthesis and releases nitric oxide	The production of mycolic acid, protein and intracelluar NO
SQ-109	Diethylamine	Inhibits cell wall synthesis	Unknown
LL-3858	Pyrrole	Unknown	Unknown

PNU-100480 Linezolid, AZD5847	Oxazolidinone	Inhibits protein synthesis	Ribosomal initiation complex
BTZO43	Benzothrazinone	Interferes with production of polysaccharides in the cell wall	Inhibits epimerase
moxifloxation	Fluoroquinolone	Inhibits DNA synthesis	DNA gyrase
gatifloxacin	Fluoroquinolone	Inhibits DNA synthesis	DNA gyrase

1.7. TB: urgency of basic research

Despite people living in the most advanced medicinal era, TB remains a major threat to human health. TB is one of mankind's oldest and worst enemies owing to its widespread nature across the world and the development of resistance to known and available drugs. The development of MDR-TB, XDR-TB and the recent occurrence of totally drug-resistant *M. tuberculosis* (TDR-TB) strains (Migliori *et al*., 2012), together with the paucity of new drug targets currently being explored, suggest that new basic research is required to delineate novel potential targets.

TB is a deadly threat, especially in areas with high HIV rates. Co-occurrence of TB with HIV/AIDS has further worsened the treatment of TB or HIV/AIDS and many patients are dying because of TB rather than the initial viral infection (Gandhi *et al*., 2006). The alarming death rates due to TB in sub-Saharan Africa, especially in South Africa, necessitate more basic research to find new drugs that act with novel mechanisms, leading to a shorter treatment period, fewer side-effects and compatibility with HIV/AIDS treatment.

1.8. *M. tuberculosis* and cytochrome P450 monooxygenases

Cytochrome P450 monooxygenases, also known as CYPs/P450s, are heme-thiolate enzymes playing key roles in nature, particularly in the evolution of organisms, including the dawn of multicellular life (Nelson, 2013; Parvez *et al*., 2016). P450s are well known for their stereo- and regio-specific oxidation of substrates, which makes these enzymes essential in the primary and secondary metabolism of organisms. Apart from their well-known oxygenation reaction, P450s perform different reactions, as listed in Figure 1.10.

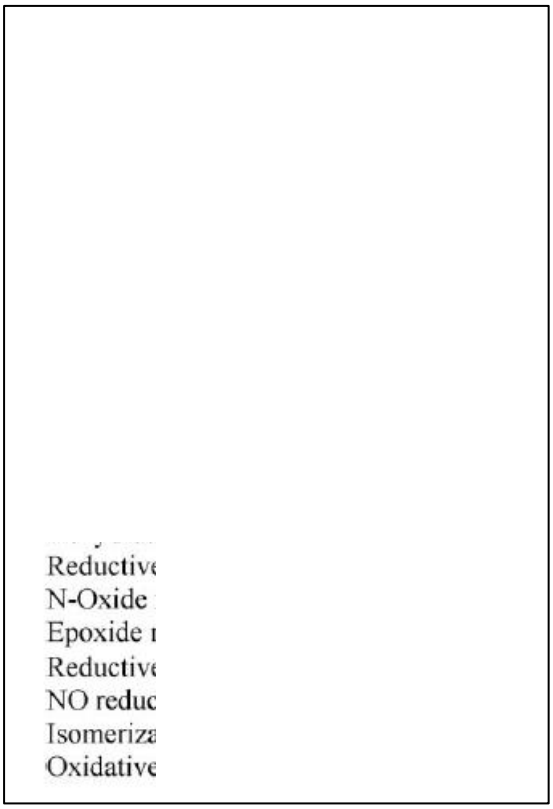

Figure 1.10. Reactions catalysed by cytochrome P450 (taken from Bernhardt, 2006).

Genome sequencing of *M. tuberculosis* revealed the presence of 20 P450s in its genome (Cole *et al*., 1998). The preponderance of P450 genes in the *M. tuberculosis* genome is an idiosyncratic feature for a prokaryote and it might underlie the possible importance of the P450 gene-family in the life history of *M. tuberculosis*.

Table 1.4. *M. tuberculosis* P450s and their respective gene IDs.

P450	Gene ID
CYP51B1	*Rv0764c*
CYP132A1	*Rv1394c*
CYP139A1	*Rv1666c*
CYP136A1	*Rv3059c*
CYP135A1	*Rv0327c*
CYP135B1	*Rv0568*
CYP137A1	*Rv3685c*
CYP138A1	*Rv0136*
CYP143A1	*Rv1785c*
CYP121A1	*Rv2276*
CYP141A1	*Rv3121*
CYP140A1	*Rv1880c*
CYP123A1	*Rv0766c*
CYP130A1	*Rv1256c*
CYP128A1	*Rv2268c*
CYP144A1	*Rv1777*

CYP124A1	*Rv2266*
CYP125A1	*Rv3545c*
CYP126A1	*Rv0778*
CYP142A1	*Rv3518c*

1.9. P450s' importance in *M. tuberculosis* physiology

Among 20 *M. tuberculosis* P450s, three P450s, CYP121A1 (McLean *et al*. 2008), CYP125A1 (Sassetti and Rubin, 2003) and CYP128A1 (Sassetti *et al*., 2003), were found to be essential for survival of *M. tuberculosis*. Further, gene knockout studies showed that CYP121A1 is essential for *in vitro M. tuberculosis* growth (McLean *et al*., 2008) and CYP125A1 is essential for infection in mice (Sassetti and Rubin, 2003) and survival in macrophages (Chang *et al*., 2007). Since the CYP128A1 mutant could not initially be obtained during *in vitro* growth, it could only be presumed that it played an important role during infection (Ouellet *et al*., 2010). However, a recent study has shown that the CYP128A1 mutant is viable and has proven to be hyper-virulent in nature (Sogi *et al*., 2016). *In vitro M. tuberculosis* latency model studies, including a carbon starvation model (Betts *et al*., 2002) and hypoxia model (Rustad *et al*., 2008), showed up-regulation of three *M. tuberculosis* P450s, CYP128A1 and CYP135A1 and CYP123A1, suggesting their potential role during *M. tuberculosis*. Based on meta-analysis of expression data, CYP123A1 is selected as the best drug candidate against the dormant phase of *M. tuberculosis* (Murphy and Brown, 2007). It is noteworthy that CYP123A1 and CYP135A1 are present only in TB-causing bacteria, suggesting their essential role.

1.10. Functional analysis of *M. tuberculosis* P450s

Despite the greater importance of *M. tuberculosis* P450s as novel drug targets (Ouellet *et al*., 2010; Hudson *et al*., 2012), only four *M. tuberculosis* P450s, apart from highly conserved CYP51, have been functionally characterised for their *in vivo* role (Hudson *et al*., 2012) (Figure 1.11). To date, physiological roles have been proposed for CYP121 for the formation of an intramolecular C-C bond between 2 tyrosyl carbon atoms of cYY (cyclic dipeptide cyclo-L-Tyr-L-Tyr) (Belin *et al*., 2009); CYP125 and CYP142 in cholesterol catabolism (McLean *et al*., 2009; Capyk *et al*., 2009; Van der Geize, 2007; Driscoll *et al*., 2010); CYP124 in oxidation of branched chain fatty acids (Johnston *et al*., 2009) and for CYP51, a conserved P450 across the phyla, in sterol demethylation (Bellamine, 1999). Based on the genetic location of CYP128A1 and *in vivo* studies, it is predicted to be involved in oxidation of menaquinone (Holsclaw *et al*., 2008; Sogi *et al*., 2016).

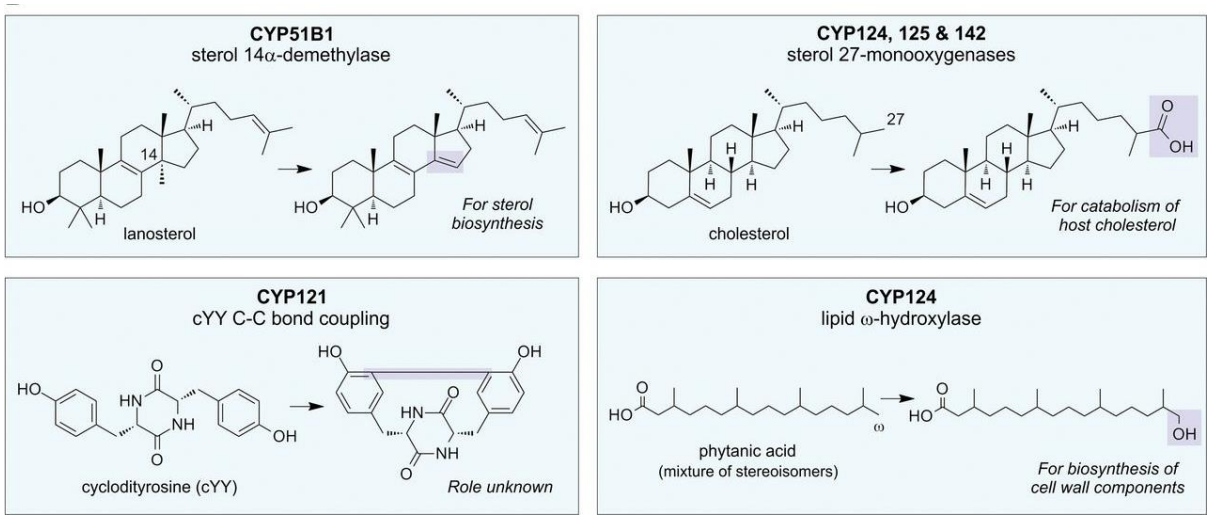

Figure 1.11. Catalytic reactions of characterised *M. tuberculosis* P450s (taken from Ouellet *et al*., 2010).

1.11. *M. tuberculosis* P450s' orphan nature and problems

Despite the most advanced methods available for gene expression and protein characterisation, to date 15 *M. tuberculosis* P450s remain orphans. The major challenges in *M. tuberculosis* P450 research are (i) expression of *M. tuberculosis* P450s and (ii) finding the physiological substrate(s). For example, even leading *M. tuberculosis* P450 research groups have been unable to express CYP128A1 (Ouellet *et al*., 2010; Driscoll, 2011) or to identify the substrate for CYP130 (Ouellet *et al*., 2008) and CYP144 (Driscoll *et al*., 2011).

1.12. Rationale and aims of the study

One-third of the human population is infected by latent TB. Active TB can be cured with drugs, unless it's drug-resistant, but not latent TB. A study showed that P450 CYP123A1 of *M. tuberculosis* is highly expressed during the latent phase of *M. tuberculosis* and also in macrophages (Rustad *et al*., 2008). Based on meta-analysis of expression data, CYP123A1 is selected as the best drug candidate against the dormant phase of *M. tuberculosis* (Murphy and Brown, 2007). It is noteworthy that CYP123A1 is present only in TB-causing bacteria, suggesting their essential role (Ouellet *et al*., 2010).

Despite the great importance of CYP123A1 as drug target for both latent TB and active TB, to date this P450s has not been characterised for its function or for its structure. This study is the first of its kind on *M. tuberculosis* P450s. The study paves the way to develop CYP123A1-based anti-TB drugs.

Furthermore, to date, the distribution of P450s in the genus *Mycobacterium* is unknown. Thus, in this study, molecular evolutionary dynamics of P450s were carried out with special focus on mycobacterial P450s.

1.13. References

- Abubakar, I., Zignol, M., Falzon, D., Raviglione, M., Ditiu, L., Masham, S., Adetifa, I., Ford, N., Cox, H., Lawn, S.D. and Marais, B.J., 2013. Drug-resistant tuberculosis: time for visionary political leadership. *The lancet infectious diseases*, *13*(6), pp. 529-539.
- Bansal-Mutalik, R. and Nikaido, H., 2014. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proceedings of the National Academy of Sciences*, *111*(13), pp. 4958-4963.
- Bates, J.H., 1980. Transmission and pathogenesis of tuberculosis. *Clinics in chest medicine*, 1(2), p. 167.
- Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C. and Dugave, C., 2009. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *106*(18), pp. 7426-7431.
- Bellamine, A., Mangla, A.T., Nes, W.D. and Waterman, M.R., 1999. Characterization and catalytic properties of the sterol 14α-demethylase from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *96*(16), pp. 8937-8942.
- Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology*, *124*(1), pp. 128-145.
- Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A. and Duncan, K., 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence

by gene and protein expression profiling. *Molecular microbiology*, *43*(3), pp. 717-731.

- Bunyan, J. and Stachniewski, J., 1998. *Grace abounding with other spiritual autobiographies*. Oxford University Press, USA.
- Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W.R., Eltis, L.D. and Mohn, W.W., 2009. Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27 steroids. *Journal of biological chemistry*, *284*(51), pp. 35534-35542.
- Chang, J.C., Harik, N.S., Liao, R.P. and Sherman, D.R., 2007. Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *The journal of infectious diseases*, *196*(5), pp. 788-795.
- Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., II III, C.B. and Tekaia, F., 1998. Erratum: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *396*(6707), p. 190.
- Driscoll M (2011) PhD thesis. https://www.escholar.manchester.ac.uk/uk-ac-man-scw:122243.
- Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E., Leys, D. and Munro, A.W., 2010. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *Journal of biological chemistry*, *285*(49), pp. 38270-38282.
- Field, S.K., Fisher, D., Jarand, J.M. and Cowie, R.L., 2012. New treatment options for multidrug-resistant tuberculosis. *Therapeutic advances in respiratory disease*, *6*(5), pp. 255-268.

- Floyd, K., Baddeley, A. and Dias, H.M., 2011. The sixteenth global report of tuberculosis. *World Health Organization [homepage on the Internet]*, p. c2012.
- Gandhi, N.R., Moll, A., Sturm, A.W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K., Andrews, J. and Friedland, G., 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *The lancet*, *368*(9547), pp. 1575-1580.
- Holsclaw, C.M., Sogi, K.M., Gilmore, S.A., Schelle, M.W., Leavell, M.D., Bertozzi, C.R. and Leary, J.A., 2008. Structural characterization of a novel sulfated menaquinone produced by stf3 from *Mycobacterium tuberculosis*. *ACS chemical biology*, *3*(10), pp. 619-624.
- Hudson, S.A., McLean, K.J., Munro, A.W. and Abell, C., 2012. *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets. Biochem Soc Trans 40(3), pp. 573-9.
- Jackson, T.A. and Thomas, J.M., 2013. Tuberculosis: the implications for anaesthesia. *Southern African journal of anaesthesia and analgesia*, *19*(6), pp. 301-305.
- Johnston, J.B., Kells, P.M., Podust, L.M. and de Montellano, P.R.O., 2009. Biochemical and structural characterization of CYP124: A methyl-branched lipid ω-hydroxylase from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *106*(49), pp. 20687-20692.
- Koch R., 1882. The etiology of tuberculosis. *Berlin clinical weekly*, 15, pp. 221–230.
- McLean, K.J., Carroll, P., Lewis, D.G., Dunford, A.J., Seward, H.E., Neeli, R., Cheesman, M.R., Marsollier, L., Douglas, P., Smith, W.E. and Rosenkrands, I., 2008. Characterization of active site structure in CYP121 a cytochrome P450 essential for

viability of *Mycobacterium tuberculosis* H37Rv. *Journal of biological chemistry*, *283*(48), pp. 33406-33416.

- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and Munro, A.W., 2009. The structure of *Mycobacterium tuberculosis* CYP125 molecular basis for cholesterol binding in a P450 needed for host infection. *Journal of biological chemistry*, *284*(51), pp. 35524-35533.
- Migliori, G.B., Centis, R., D'Ambrosio, L., Spanevello, A., Borroni, E., Cirillo, D.M. and Sotgiu, G., 2012. Totally drug-resistant and extremely drug-resistant tuberculosis: the same disease? *Clinical infectious diseases*, *54*(9), pp. 1379-1380.
- Murphy, D.J. and Brown, J.R., 2007. Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. *BMC infectious diseases*, *7*(1), p. 84.
- Nelson, D.R., 2013. A world of cytochrome P450s. *Phil. Trans. R. Soc. B. Biol. Sci*, 368(1612), p. 20120430.
- Nuermberger, E.L., Spigelman, M.K. and Yew, W.W., 2010. Current development and future prospects in chemotherapy of tuberculosis. *Respirology*, *15*(5), pp. 764-778.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2010. The *Mycobacterium tuberculosis* cytochrome P450 system. *Archives of biochemistry and biophysics*, *493*(1), pp. 82-95.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2010. The *Mycobacterium tuberculosis* cytochrome P450 system. *Archives of biochemistry and biophysics*, *493*(1), pp. 82-95.
- Ouellet, H., Podust, L.M. and de Montellano, P.R.O., 2008. *Mycobacterium tuberculosis* CYP130 crystal structure, biophysical characterization, and interactions with antifungal azole drugs. *Journal of biological chemistry*, *283*(8), pp. 5069-5080.

- Pai, M., Behr, M.A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C.C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D. and Raviglione, M. 2016. Tuberculosis. *Nature reviews disease primers 2*, Article number: 16076.
- Parvez, M., Qhanya, L.B., Mthakathi, N.T., Kgosiemang, I.K.R., Bamal, H.D., Pagadala, N.S., Xie, T., Yang, H., Chen, H., Theron, C.W. and Monyaki, R., 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: special focus on mycobacterial P450s. *Scientific reports*, *6*, p. 33099.
- Quan, D., Nagalingam, G., Payne, R. and Triccas, J.A., 2017. New tuberculosis drug leads from naturally occurring compounds. *International Journal of Infectious Diseases*, *56*, pp. 212-220.
- Rendon, A., Tiberi, S., Scardigli, A., D'Ambrosio, L., Centis, R., Caminero, J.A. and Migliori, G.B., 2016. Classification of drugs to treat multidrug-resistant tuberculosis (MDR-TB): evidence and perspectives. *Journal of thoracic disease*, *8*(10), p. 2666.
- Rustad, T.R., Harrell, M.I., Liao, R. and Sherman, D.R., 2008. The enduring hypoxic response of *Mycobacterium tuberculosis*. *PloS one*, *3*(1), p. e1502.
- Sassetti, C.M. and Rubin, E.J., 2003. Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences*, *100*(22), pp. 12989-12994.
- Sogi, K.M., Holsclaw, C.M., Fragiadakis, G.K., Nomura, D.K., Leary, J.A. and Bertozzi, C.R., 2016. Biosynthesis and regulation of sulfomenaquinone, a metabolite associated with virulence in *Mycobacterium tuberculosis*. *ACS infectious diseases*, *2*(11), pp. 800-806.
- TB Alliance. https://www.tballiance.org/why-new-tb-drugs/global-pandemic.
- Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W. and Eltis, L.D., 2007. A gene

cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences*, *104*(6), pp. 1947-1952.

- WHO 2014. Data sources: Global tuberculosis report. http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf
- WHO 2015. Data sources: Global tuberculosis report. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf
- WHO Global tuberculosis report 2016. http://www.who.int/tb/publications/global_report/en/.
- World Health Organization, 2011. Global tuberculosis control: WHO report 2011. Geneva: WHO; 2011. *Report No.: WHO/HTM/TB*.
- Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H. and Nerlich, A.G., 2003. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *Journal of clinical microbiology*, *41*(1), pp. 359-367.

CHAPTER 2

MOLECULAR EVOLUTIONARY DYNAMICS OF CYTOCHROME P450 MONOOXYGENASES ACROSS KINGDOMS: SPECIAL FOCUS ON MYCOBACTERIAL P450s

2.1. INTRODUCTION

Cytochrome P450 monooxygenases, also known as CYPs/P450s, are heme-thiolate enzymes playing key roles in nature, particularly in the evolution of organisms, including the dawn of multicellular life (Nelson *et al.*, 2013). P450s are well known for their capability for stereo- and regio-specific oxidation of substrates, which makes these enzymes essential in organisms' primary and secondary metabolisms. Since their identification five decades ago, quite a large number of P450s have been identified in species across biological kingdoms, primarily because of the current genome sequencing rush (Nelson *et al.*, 2009). Studies on P450 enzymes have been reported from animals owing to their role in drug metabolism (particularly from mammals) (Guengerich *et al.*,2015) or analysis of diversity (Nelson, 2013; Sezutsu *et al.*, 2013), from fungi owing to their role as drug target (Kelly and Kelly, 2013; Jawallapersand *et al.*, 2014) and for evolutionary analysis (Moktali *et al.*, 2012; Chen *et al.*, 2014; Syed *et al.*, 2014) from bacteria owing to P450 structure-functional analysis (Poulos *et al.*, 2015) and generation of products valuable to humans (McLean *et al.*, 2015); and from plants owing to their role in key cellular processes and defense mechanisms (Mizutan and Ohta, 2010; Hamberger and Bak, 2013; Schuler, 2015). Irrespective of their origins, P450s from all organisms have been exploited for their biotechnological potential (Girhard *et al.*, 2015). Although quite a large number of P450s

have become known to date (Guengerich *et al.*, 2015), genome annotation of P450s in recently available organism genomes has led to the discovery of a novel P450 fused family (Sello *et al.*, 2015) , suggesting that much remains to be explored and understood about the evolution of these enzymes.

The origin of P450s in organisms was ascribed to CYP51 (Nelson, 1999; Yoshida *et al.*, 2000). CYP51 is regarded as an ancient P450 despite confusion on its origin, either in prokaryotes or in eukaryotes (Nelson, 1999; Yoshida *et al.*, 2000). It has been postulated that the evolution of present-day P450s is due to the divergence and duplication of this ancient P450 (Nelson, 1999; Yoshida *et al.*, 2000). The proposed hypothesis of "descendance of P450s from CYP51" is strongly supported by the fact that CYP51 is the only P450 that is conserved across species belonging to different biological kingdoms (Lepesheva and Waterman, 2007). Recent studies showing the conservation of amino acid patterns in EXXR and CXG motifs in CYP51 P450s collected from species across biological kingdoms (Syed and Mashele, 2014) further supported the idea that the origin of CYP51 predates the divergence of biological kingdoms, as mentioned in earlier studies (Nelson and Strobel, 1987).

Since CYP51 P450 was established to be the origin of all P450s, further studies have focused on unravelling the P450 complement from different organisms, analysis of P450 diversity and their evolution. To date, a large number of P450 families have been reported in species from different biological kingdoms, with fungi reported to have more diverse P450 families than plants, animals or bacteria (Nelson, 2011). Thorough studies on the origins of P450 diversity have been conducted using animal and plant P450s as model P450s (Sezutsu *et al.*, 2013). Eukaryotic microbes such as fungi and oomycete P450s have also been extensively studied with respect to their diversity and duplications (Moktali *et al.*, 2012; Chen *et al.*, 2014;

Syed *et al.*, 2014; Sello *et al.*, 2015; Qhanya *et al*., 2015). However, bacterial P450s have been virtually unexplored in the above context, considering the general assumption that bacterial species have a low number of P450s, with many bacteria lacking P450s.

Although a plethora of information on P450s is available, to date two key research gaps have not been addressed in P450 research. The first gap is that, since the origin and divergence of P450s from CYP51 into different biological kingdoms (Nelson and Strobel, 1987), P450s' molecular dynamics in terms of their primary structure both at protein and DNA level have not been studied. Studies on understanding the molecular dynamics of P450s' primary structure are limited to the P450 family CYP51 (Yoshida *et al.*, 1997) (in 1997) and the biosynthetic-type (710 P450s) and detoxification-type P450s (543 P450s) from vertebrate (14) and invertebrate (6) genomes (Kawashima and Satta, 2014). Moreover, comprehensive analysis of P450 families from different biological kingdoms and their dynamics in terms of evolutionary rate both at protein and DNA levels have not been reported. Identification of P450 families with the highest evolutionary rate will provide important answers, such as P450 families that are possibly adapting to different substrates or emerging into new P450s, thus creating new P450 families. Quite the opposite is true for the P450s with a lower evolutionary rate, indicating their conservation by the organism owing to their critical role in organisms' physiology and therefore possibly strict substrate specificity. The second gap is that comparative genomic analyses of P450s in bacteria have not been reported despite the comprehensive bacterial P450 functional data (McLean *et al.*, 2015). As mentioned above, compared to fungal, animal and plant P450s, bacterial P450s have not been explored in terms of their diversity and evolutionary analysis.

In this study, we address these research gaps *per se* by performing genome-wide identification, annotation and phylogenetic analysis of P450s in 60 mycobacterial species

belonging to six different categories. The species belonging to the genus *Mycobacterium* are selected because of the public availability of quite a large number of mycobacterial species genomes and the presence of diverse species (Ventura *et al.,* 2007), thus possibly representing diverse P450 complements. Furthermore, the highest G+C content and high genome level conservation among mycobacterial species make them interesting candidates for observation of the distribution, diversity and evolution of P450s in these organisms. The results from the mycobacterial P450 analysis allowed us to carry out comprehensive evolutionary analysis (both at protein and DNA level) and phylogenetic analysis of 17 598 P450s belonging to 113 P450s families from species across biological domains: bacteria, fungi, plants and animals.

2.2. METHODS

2.2.1. Mycobacterial species

A total of 60 mycobacterial species belonging to six different categories were used in this study (Table 2.1). The six categories include *Mycobacterium tuberculosis* complex (MTBC) (27 species), *M. chelonae-abscessus* complex (MCAC) (6 species), *M. avium* complex (MAC) (8 species), Mycobacteria causing leprosy (MCL) (2 species), nontuberculous mycobacteria (NTM) (6 species) and Saprophytes (SAP) (11 species). The criteria for separation of mycobacterial species into six different groups is based on their characteristic features, including ecological niches, nature of infection and site of infection as described elsewhere (Ventura *et al.,* 2007). Also taxonomical grouping of mycobacterial species is taken into consideration as described elsewhere (Tortoli *et al.,* 2012). Detailed information on species and category is listed in Table 2.1.

Table 2.1. List of mycobacterial species used in the study. As listed in the table, 60 mycobacterial species were grouped under six different categories based on their characteristic features, including ecological niches, nature of infection and site of infection as described elsewhere (Reddy *et al*., 2009; Galagan *et al*., 2010; Kanehisa and Goto, 2000; Kanehisa *et al*., 2016; The UniProt Consortium, 2015). Also taxonomical grouping of mycobacterial is taken into consideration as described elsewhere (Tortoli *et al.,* 2012).

***Mycobacterium tuberculosis* complex (MTBC)**	
1	*Mycobacterium africanum* GM041182
2	*Mycobacterium tuberculosis* C
3	*Mycobacterium tuberculosis* F11
4	*Mycobacterium tuberculosis* H37Ra
5	*Mycobacterium tuberculosis* H37Rv
6	*Mycobacterium tuberculosis* Haarlem
7	*Mycobacterium tuberculosis* KZN 1435
8	*Mycobacterium tuberculosis* KZN 605
9	*Mycobacterium tuberculosis* KZN 4207
10	*Mycobacterium tuberculosis* RGTB327
11	*Mycobacterium tuberculosis* CDC1551
12	*Mycobacterium tuberculosis* strains CCDC5079
13	*Mycobacterium tuberculosis* 7199-99
14	*Mycobacterium tuberculosis* Beijing/NITR203
15	*Mycobacterium tuberculosis* CAS/NITR204
16	*Mycobacterium tuberculosis* EAI5
17	*Mycobacterium tuberculosis* EAI5/NITR206
18	*Mycobacterium tuberculosis* Erdman= ATCC 35801
19	*Mycobacterium tuberculosis* UT205
20	*Mycobacterium canetii* CIPT 140010059
21	*Mycobacterium canetii* CIPT 140060008
22	*Mycobacterium canetii* CIPT 140710010
23	*Mycobacterium bovis* AF 2122/97
24	*Mycobacterium bovis* BCG Pasteur 1173P2
25	*Mycobacterium bovis* BCG Korea 1168P
26	*Mycobacterium bovis* BCG Mexico
27	*Mycobacterium bovis* BCG Tokyo 172
***Mycobacterium chelonae-abscessus* complex (MCAC)**	
28	*Mycobacterium abscessus* ATCC 19977
29	*Mycobacterium abscessus* subsp. *bolletii* 50594
30	*Mycobacterium abscessus* 47J26
31	*Mycobacterium abscessus*103

32	*Mycobacterium abscessus* subsp. *bolletii* MA 1948
33	*Mycobacterium abscessus* VO6705
***Mycobacterium avium* complex (MAC)**	
34	*Mycobacterium Avium* 104
35	*Mycobacterium Avium* subsp. *paratuberculosis* K10
36	*Mycobacterium avium* subsp. *paratuberculosis* MAP4
37	*Mycobacterium intracellulare* ATCC 13950
38	*Mycobacterium intracellulare* MOTT-02
39	*Mycobacterium Intracellulare* MOTT-64
40	*Mycobacterium intracellulare* MOTT-36Y
41	*Mycobacterium Indicus pranii* MTCC 9506
Mycobacteria causing leprosy (MCL)	
42	*Mycobacterium leprae* Br4923
43	*Mycobacterium Leprae* TN
Nontuberculous mycobacteria (NTM)	
44	*Mycobacterium* sp. JDM601
45	*Mycobacterium liflandii* 128FXT
46	*Mycobacterium ulcerans* Agy99
47	*Mycobacterium Marinum*
48	*Mycobacterium massiliense*
49	*Mycobacterium kansassii* ATCC 12478
Saprophytes (SAP)	
50	*Mycobacterium* sp. JLS
51	*Mycobacterium* sp. KMS
52	*Mycobacterium* sp. MCS
53	*Mycobacterium vanbaalenii* PYR-1
54	*Mycobacterium smegmatis* MC2 155
55	*Mycobacterium chubuense* NBB4
56	*Mycobacterium gilvum* PYR-GCK
57	*Mycobacterium gilvum* Spyr1
58	*Mycobacterium smegmatis* JS623
59	*Mycobacterium rhodesiae* NBB3
60	*Mycobacterium neoaurum* VKM Ac-1815D

2.2.2. P450 mining in mycobacterial genomes

Mycobacterial genomes that are publicly available at different genome databases as listed in Table 2.2 were mined for P450s as described elsewhere (Sello *et al.*, 2015; Syed and Mashele, 2014; Qhanya *et al.*, 2015; Syed *et al.*, 2014). Briefly, the whole proteome of mycobacterial

species was downloaded from the respective data bases listed Table 2.2 and subjected to the NCBI *Batch* Web CD-Search Tool (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi).

Table 2.2. Information on mycobacterial species genome databases and their web links used in the study.

Database	Weblink	Reference
TB Database	http://genome.tbdb.org/tbdb_sysbio/GenomesIndex.html	Reddy *et al*., 2009; Galagan *et al*., 2010
KEGG	http://www.genome.jp/kegg-bin/show_organism?category=Mycobacterium	Kanehisa and Goto, 2000; Kanehisa *et al*., 2016
UniProt	http://www.uniprot.org/taxonomy/36809	The UniProt Consortium, 2015

Proteins that belong to a P450 superfamily were selected and further subjected to BLAST analysis against bacterial P450s at the Cytochrome P450 Homepage (Nelson, 2009). Based on the International P450 Nomenclature Committee rule, proteins with >40% identity and >55% identity were grouped under the same family and subfamily, respectively. For each species of P450s different databases were consulted; all databases gave the same results and hence P450s from one of the databases were selected. Some mycobacterial P450s were annotated and made available at the Cytochrome P450 Homepage (Nelson, 2009). In this case, the same nomenclature for P450s was continued. The same procedure was followed for mycobacterial P450s documented in the literature. P450s that showed less than 40% identity to known P450s at the Cytochrome P450 Homepage (Nelson, 2009) were assigned to new P450 families and subfamilies as per International P450 Nomenclature Committee rules.

2.2.3. Protein and cDNA sequence collection

Protein sequences and the respective cDNAs for 113 P450 family members (17 598 P450s) were collected from different databases and published resources. Some of the bacterial P450 family members' protein sequences and cDNAs, particularly P450 families present in mycobacterial species were collected from the KEGG database (http://www.genome.jp/kegg/catalog/org_list.html). These P450 families include: CYP108, CYP121, CYP123-CYP126, CYP128, CYP130, CYP132, CYP135-CYP144, CYP150, CYP164, CYP185, CYP187-CYP191, CYP268, CYP279 and CYP291. Identification of P450 family members was carried out using the methodology described elsewhere (Sello *et al.*, 2015; Syed and Mashele, 2014; Qhanya *et al.*, 2015; Syed *et al.*, 2014). Briefly, a representative P450 for each of the families was taken from mycobacterial species from the Cytochrome P450 Homepage (Nelson, 2009) and protein BLAST was performed at KEGG against prokaryote genomes. The resulting hit proteins were sorted into P450 families according to the International P450 Nomenclature Criteria i.e. >40% identity as a family. The hit proteins were subjected to BLAST analysis against named bacterial P450s at the Cytochrome P450 Homepage (Nelson, 2009). Based on the percentage identity to the homolog P450 (>40% identity), the hit proteins were sorted into different P450 families.

Fungal P450 sequences and corresponding cDNAs belonging to P450 families CYP61, CYP63, CYP512, CYP5035, CYP5037, CYP5136, CYP5139, CYP5141, CYP5144, CYP5150 and CYP5152 were retrieved from published work (Chen *et al*., 2014; Syed *et al*., 2014; Syed and Mashele, 2014) and their corresponding cDNAs were retrieved from each of the species' genome databases at MycoCosm (Grigoriev *et al.*,2014) . The remaining P450 family members' protein sequences, i.e. CYP1-9, CYP11, CYP12, CYP17, CYP19, CYP21, CYP24, CYP26-

CYP28, CYP33, CYP39, CYP46, CYP51-CYP53, CYP55, CYP58, CYP65, CYP71-CYP76, CYP78, CYP79, CYP81, CYP82, CYP84, CYP86, CYP87, CYP89, CYP90, CYP92-CYP94, CYP96-CYP98, CYP102, CYP105-CYP107, CYP110, CYP116, CYP147, CYP152, CYP153, CYP157, CYP195, CYP202, CYP325, CYP501, CYP505, CYP584, CYP620, CYP704-CYP707, CYP709, CYP714 and CYP716 were retrieved from the Cytochrome P450 Engineering Database (https://cyped.biocatnet.de/) (Sirim *et al.*, 2009) and their cDNAs were collected from NCBI using the respective P450 family member protein IDs. The P450 protein sequences (17 598 P450 sequences) belonging to 113 P450 families used can find out online version (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/) in this study are listed in Supplementary Dataset 1 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1).

2.2.4. Analysis of amino acid conservation

Analysis of the number of amino acids conserved across different mycobacterial P450 families was carried out using PROfile Multiple Alignment with predicted Local Structures and 3D constraints (PROMALS3D) (Pei *et al.*, 2008). PROMALS3D aligns multiple protein sequences based on their secondary structure prediction using the available homolog crystal structures. The output alignment assigns numbers as conservation index from 4 to 9, where number 9 is the invariantly conserved amino acid across the analyzed protein sequences (Pei *et al.*, 2008). The majority of mycobacterial P450 families contain fewer than 10 members. Hence in this study, mycobacterial P450 families with more than 15 members were selected for amino acid conservation analysis, considering that this number of member P450s will provide sufficient information on amino acid conservation. This cut-off value is also set taking into account that bacteria possess a low number of P450s in their genomes compared to species from other biological kingdoms (Nelson, 2011). Based on the cut-off value, a total of 32 mycobacterial

P450 families (CYP51, CYP105, CYP108, CYP121, CYP123, CYP124, CYP125, CYP126, CYP128, CYP130, CYP132, CYP135, CYP136, CYP137, CYP138, CYP139, CYP140, CYP141, CYP142, CYP143, CYP144, CYP150, CYP164, CYP185, CYP187, CYP188, CYP189, CYP190, CYP191, CYP268, CYP279 and CYP291) were qualified for this analysis. The P450 families from fungi, plants and animals were selected based on criteria described elsewhere (Syed and Mashele, 2014). The P450 families used from different biological kingdoms were listed in the Supplementary Dataset1 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1). CYP51 members from bacteria and fungi individually were included in the analysis to see their ranking of conservation compared to CYP51 members across biological kingdoms.

2.2.5. Construction of P450s phylogenetic trees

The phylogenetic analysis of P450s was conducted as described elsewhere (Chen *et al.*, 2014). Briefly, the P450 protein sequences were aligned by HMMER package 3.1 (http://hmmer.janelia.org/) through adjusting them to the P450 profile hidden Markov model PF00067 downloaded from the Pfam protein families database (http://pfam.xfam.org/) (Eddy, 2011; Finnet *et al.*, 2014). Then, the phylogenetic trees from alignments were inferred by Fast Tree version 2.1.4 with the maximum-likelihood method (http://www.microbesonline.org/fasttree/) (Price *et al.*, 2009). In this study, the phylogenetic tree of 1 772 mycobacterial P450s was viewed by iTOL (http://itol.embl.de/upload.cgi) (Letunic and Bork, 2007) and the phylogenetic tree of 17 598 P450s from 113 families was viewed by Hypertree 1.2.2 (Bingham and Sudarsanam, 2000).

2.2.6. Analysis of P450 diversity

The P450 diversity percentage is the percentage contribution of the number of P450 families in the total number of P450s in an organism (Sello *et al*., 2015). The P450 diversity percentage in mycobacterial species was calculated using the methodology described elsewhere (Sello *et al.*, 2015). For comparative analysis of the P450 diversity percentage between different microbial populations, the P450 diversity percentage data from lower eukaryote microbes such as fungi and oomycetes was retrieved from published literature (Nelson, 2009; Syed *et al.*, 2014; Sello *et al.*, 2015; Qhanya *et al.*, 2015; Syed *et al.*,2013; Kgosiemang *et al.*, 2014).

2.2.7. Evolutionary rate analysis

The above-mentioned cDNA sequences of 113 P450 family members were used for evaluating their evolutionary rates. Evolutionary rate for all P450s were calculated using full-length cDNA sequences. Firstly, the cDNA sequences of each P450 family member were aligned by Muscle in the codons module (Edgar, 2004). All positions containing gaps and missing data were eliminated. Then, the aligned sequences were subjected to examination in order to estimate their evolutionary rates under the Tamura-Nei model (Tamura and Nei, 1993). A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate of substitution for each site was drawn from a Gamma distribution with shape parameter α. If α was <1, the distribution implied that there was a relatively large amount of rate variation, with many sites evolving very slowly but some sites evolving at a high rate. For values of α >1, the shape of the distribution changed qualitatively, with less variation and most sites having roughly similar rates (Liò and Goldman, 1998). The evolutionary rate analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

2.3. RESULTS AND DISCUSSION

2.3.1. Mycobacterial P450s

Genome data-mining and annotation of P450s in 60 mycobacterial species revealed the presence of 1 784 P450s in their genome (Fig. 2.1 and Table 2.3 and can find out online version of Datasets 2 and 3: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1). Among the species, *M. leprae* species showed a single P450, whereas the highest number of P450s (70 P450s) was found in *M. rhodesiae* NBB3, followed by *M. indicus pranii* MTCC 9506 (61 P450s). An interesting pattern was observed when comparing the P450 pattern between mycobacterial species belonging to different categories (Fig. 2.1). Comparison of the number of P450s and average number of P450s among different categories revealed a gradual loss of P450s from SAP to MTBC species (Fig. 2.1 and Table 2.3).

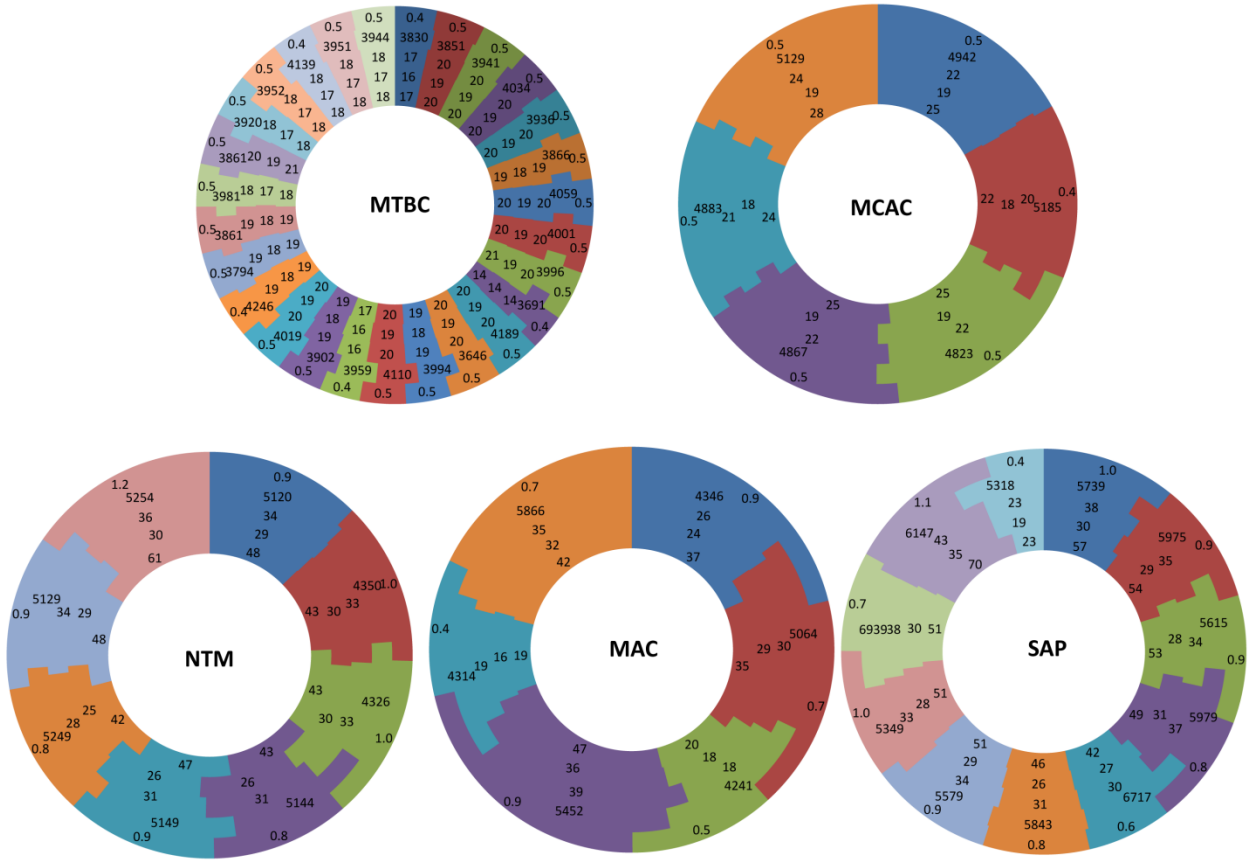

Figure 2.1. Comparative analysis of P450s in mycobacteria. Mycobacterial species were grouped into different groups such as MTBC (*Mycobacterium tuberculosis* complex), MCAC (*Mycobacterium chelonae-abscessus* complex), MAC (*Mycobacterium avium* complex), NTM (Nontuberculous mycobacteria) and SAP (Saprophytes). Each color in the circle represents a mycobacterial species in that group. The numerical order from the inside to the outside of the circle is as follows: number of P450s, number of P450 families, number of P450 subfamilies, number of ORFs in an organism and percentage of P450s compared to ORFs of an organism. Considering the presence of a single P450 in MCL (Mycobacteria causing leprosy) species, this group is omitted from comparative analysis. Details on species and their P450s were listed in table 2.3.

Table 2.3. Genome data-mining, identification and annotation of cytochrome P450 monooxygenases in 60 mycobacterial species. Open reading frames (ORFs) for each species were obtained from KEGG database (Kanehisa and Goto, 2000; Kanehisa *et al*., 2016).

Name of the species	P450 count	Families	Sub-families	ORFs	% of P450s
Mycobacterium tuberculosis complex (MTBC)					
Mycobacterium africanum GM041182	17	16	17	3830	0.4
Mycobacterium tuberculosis C	20	19	20	3851	0.5
Mycobacterium tuberculosis F11	20	19	20	3941	0.5
Mycobacterium tuberculosis H37Ra	20	19	20	4034	0.5
Mycobacterium tuberculosis H37Rv	20	19	20	3936	0.5
Mycobacterium tuberculosis Haarlem	19	18	19	3866	0.5
Mycobacterium tuberculosis KZN 1435	20	19	20	4059	0.5
Mycobacterium tuberculosis KZN 605	20	19	20	4001	0.5
Mycobacterium tuberculosis KZN 4207	21	19	20	3996	0.5
Mycobacterium tuberculosis RGTB327	14	14	14	3691	0.4
Mycobacterium tuberculosis CDC1551	20	19	20	4189	0.5
Mycobacterium tuberculosis strains CCDC5079	20	19	20	3646	0.5
Mycobacterium tuberculosis 7199-99	19	18	19	3994	0.5
Mycobacterium tuberculosis Beijing/NITR203	20	19	20	4110	0.5
Mycobacterium tuberculosis CAS/NITR204	17	16	16	3959	0.4
Mycobacterium tuberculosis EAI5	19	18	19	3902	0.5
Mycobacterium tuberculosis EAI5/NITR206	20	19	20	4019	0.5
Mycobacterium tuberculosis Erdman= ATCC 35801	19	18	19	4246	0.4
Mycobacterium tuberculosis UT205	19	18	19	3794	0.5
Mycobacterium canetii CIPT 140010059	19	18	19	3861	0.5
Mycobacterium canetii CIPT 140060008	18	17	18	3981	0.5
Mycobacterium canetii CIPT 140710010	21	19	20	3861	0.5
Mycobacterium bovis AF 2122/97	18	17	18	3920	0.5
Mycobacterium bovis BCG Pasteur 1173P2	18	17	18	3952	0.5
Mycobacterium bovis BCG Korea 1168P	18	17	18	4139	0.4
Mycobacterium bovis BCG Mexico	18	17	18	3951	0.5
Mycobacterium bovis BCG Tokyo 172	18	17	18	3944	0.5
Mycobacterium chelonae-abscessus complex (MCAC)					
Mycobacterium abscessus ATCC 19977	25	19	22	4942	0.5
Mycobacterium abscessus subsp. bolletii 50594	22	18	20	5185	0.4
Mycobacterium abscessus 47J26	25	19	22	4823	0.5

*Mycobacterium abscessus*103	25	19	22	4867	0.5
Mycobacterium abscessus subsp. bolletii MA 1948	24	18	21	4883	0.5
Mycobacterium abscessus VO6705	28	19	24	5129	0.5
Mycobacterium avium complex (MAC)					
Mycobacterium Avium 104	48	29	34	5120	0.9
Mycobacterium Avium subsp. paratuberculosis K10	43	30	33	4350	1.0
Mycobacterium avium subsp. paratuberculosis MAP4	43	30	33	4326	1.0
Mycobacterium intracellulare ATCC 13950	43	26	31	5144	0.8
Mycobacterium intracellulare MOTT-02	47	26	31	5149	0.9
Mycobacterium Intracellulare MOTT-64	42	25	28	5249	0.8
Mycobacterium intracellulare MOTT-36Y	48	29	34	5129	0.9
Mycobacterium Indicus pranii MTCC 9506	61	30	36	5254	1.2
Mycobacteria causing leprosy (MCL)					
Mycobacterium leprae Br4923	1	1	1	1604	0.1
Mycobacterium Leprae TN	1	1	1	1605	0.1
Nontuberculous mycobacteria (NTM)					
Mycobacterium sp. JDM601	37	24	26	4346	0.9
Mycobacterium liflandii 128FXT	35	29	30	5064	0.7
Mycobacterium ulcerans Agy99	20	18	18	4241	0.5
Mycobacterium Marinum	47	36	39	5452	0.9
Mycobacterium massiliense	19	16	19	4314	0.4
Mycobacterium kansassii ATCC 12478	42	32	35	5866	0.7
Saprophytes (SAP)					
Mycobacterium sp. JLS	56	30	38	5739	1.0
Mycobacterium sp. KMS	54	29	35	5975	0.9
Mycobacterium sp. MCS	53	28	34	5615	0.9
Mycobacterium vanbaalenii PYR-1	49	31	37	5979	0.8
Mycobacterium smegmatis MC2 155	42	27	30	6717	0.6
Mycobacterium chubuense NBB4	46	26	31	5843	0.8
Mycobacterium gilvum PYR-GCK	51	29	34	5579	0.9
Mycobacterium gilvum Spyr1	51	28	33	5349	1.0
Mycobacterium smegmatis JS623	51	30	38	6939	0.7
Mycobacterium rhodesiae NBB3	70	35	43	6147	1.1
Mycobacterium neoaurum VKM Ac-1815D	23	19	23	5318	0.4

The order is as follows, where the minimum and maximum number of P450s, and after the semicolon the average number of P450s, are shown in parenthesis: SAP (23-70:50)>MAC (42-61:47)>NTM (19-47:33)>MCAC (22-28:25)>MTBC (14-21:19) (Fig. 2.1 and Table 2.3).

This clearly indicates that the progression from soil mycobacteria (SAP) into human pathogens such as those living in human blood and then settling as a lung pathogen (MTBC) resulted in gradual loss of a considerable number of P450s. Furthermore, the reduction in P450 count followed the pattern of genome reduction as number of Open Reading Frames (ORFs) in organisms was reduced from SAP to MTBC (see ORF Table 2.3). The average number of P450s in mycobacterial species, especially belonging to NTM, MCAC and SAP, is higher than some human fungal pathogens (Syed *et al*., 2014; Kgosiemang, 2014). The percentage analysis of P450s in mycobacterial species genomes reflected the same pattern and some of the species belonging to MAC and SAP showed ≥1% of P450s in their genomes, indicating the important role of this large number of P450s in their physiology (Fig. 2.1 and Table 2.3). It is noteworthy that some fungi and plants also have ≥1% P450s in their genomes (Nelson, 2009), the same as SAP and MAC, suggesting P450s play and occupy an important role in both prokaryotic and eukaryotic organisms. Considering the presence of a single P450 in MCL species, this group is omitted from comparative analysis.

2.3.2. P450 family and subfamily analysis in mycobacteria

All 1 784 P450s were grouped into 77 P450 families and 132 subfamilies (See Table 32.4 and can Dataset 2: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1).

Table 2.4. Comparative analysis of P450 families between different mycobacterial categories. For each P450 family, the number of member P450s (P450 count) and their percentage in the total number of P450s (1 786) found in mycobacterial species were shown. This includes CYP164 P450s from the Mycobacterium causing leprosy (MCL). However, the MCL category is not presented in the table, considering the presence of a single P450 CYP164 in their genome. The number of species belonging to different categories (MTBC, MCAC, MAC, NTM and SAP) used in the analysis is shown in parenthesis next to the category name. The distribution of member P450s in the category is as follows: min-max number of P450s or number of member P450s in the category members/number of category members possessing the member P450. The empty boxes with green fluorescent color indicate the absence of P450 family members in the category. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *Mycobacterium chelonae-abscessus* complex; MAC, *Mycobacterium avium* complex; NTM, Nontuberculous mycobacteria; SAP, Saprophytes.

P450 family	P450 count	Percentage	MTBC (27)	MCAC (6)	MAC (8)	NTM (6)	SAP (11)
CYP51	54	3.0	1/24	1/6	1-2/8	1/4	1/11
CYP123	63	3.5	1-2/26	1/6	1-2/5	1/5	1-2/11
CYP125	114	6.4	1/25	3-4/6	1-3/8	1-5/5	1-4/10
CYP130	51	2.9	1/21	1/6	1/8	1/5	1/11
CYP135	81	4.5	1/27	1/6	1/4	1-3/4	1/10
CYP136	82	4.6	1/27	1/6	1/8	1-2/6	1-2/11
CYP138	72	4.0	1/26	1-2/6	1/8	1-3/5	1-2/11
CYP140	72	4.0	1/27	1/6	1/8	1/6	1/9
CYP144	58	3.2	1/27	1/6	1/6	1-2/6	1-2/11

CYP1128	10	0.6	1/1	1/6	1/1	1/1	1/1
CYP108	56	3.1		1/6	1-2/8	1/4	1-3/11
CYP164	29	1.6		1/6	1/8	1/3	1/10
CYP124	59	3.3	1/27		1-2/7	1/4	1-2/11
CYP126	50	2.8	1/26		1/8	1/5	1/11
CYP128	32	1.8	1-2/26		1/2	1/1	1/3
CYP142	59	3.3	1/26		1-2/8	1/5	1/11
CYP143	62	3.5	1/27		1-2/7	1-2/5	1-2/6
CYP105	45	2.5			1-2/5	1/5	1-2/11
CYP185	18	1.0			1/3	1/2	1-2/10
CYP187	54	3.0			2-4/8	1-2/5	1-4/10
CYP188	39	2.2			2/8	1/5	1-3/10
CYP189	86	4.8			3-5/8	1-2/5	1-6/10
CYP190	44	2.5			1-3/8	1/4	1-3/10
CYP191	21	1.2			1/8	1/4	1/10
CYP243	6	0.3			1/4	1/1	1/1
CYP268	28	1.6			1/8	1/4	1-3/10
CYP279	47	2.6			1-4/8	1/3	1-4/10
CYP150	84	4.7			2-6/8	1-5/5	3-5/10
CYP1027	6	0.3			1/4	1/1	1/1
CYP137	32	1.8	1-2/27			1/3	1/1
CYP139	33	1.8	1/26		1/3	1/4	
CYP1134	7	0.4		1/6	1/1	1/1	
CYP286	8	0.4		1/6		1/1	1/1

CYP153	11	0.6		1/4		1/1	1/6
CYP161	8	0.4		1/6		1/1	1/1
CYP132	27	1.5	1/26			1/1	
CYP102	5	0.3		1/1			1/4
CYP1110	5	0.3		1/4		1/1	
CYP1130	16	0.9		1/6		1/2	
CYP1132	7	0.4		1/6		1/1	
CYP272	4	0.2			1/3	1/1	
CYP291	15	0.8			1/8		1-2/5
CYP1034	3	0.2			1/1		1/1
CYP1124	6	0.3			1/4		1/2
CYP1129	6	0.3			1-2/4		1/1
CYP107	2	0.1				1/1	1/1
CYP147	5	0.3				1/3	1/1
CYP278	4	0.2				1/2	1/2
CYP121	23	1.3	1/23				
CYP141	19	1.1	1/19				
CYP1131	1	0.1		1/6			
CYP1016	2	0.1			1/2		
CYP1017	2	0.1			1/2		
CYP1018	6	0.3			1-2/2		
CYP1019	4	0.2			2/2		
CYP1067	1	0.1			1/1		
CYP183	2	0.1				1/2	

CYP226	2	0.1				1/2	
CYP269	2	0.1				1/2	
CYP271	1	0.1				1/1	
CYP274	2	0.1				1/2	
CYP276	2	0.1				1/2	
CYP1120	1	0.1				1/1	
CYP1123	1	0.1				1/1	
CYP1133	1	0.1				1/1	
CYP1127	1	0.1				1/1	
CYP110	3	0.2					1/3
CYP145	4	0.2					1/3
CYP151	2	0.1					1/2
CYP186	4	0.2					1/4
CYP289	2	0.1					1/2
CYP292	5	0.3					1/5
CYP1119	1	0.1					1/1
CYP1121	1	0.1					1/1
CYP1122	1	0.1					1/1
CYP1125	1	0.1					1/1
CYP1126	1	0.1					1/1

P450 family annotation revealed the presence of 17 new P450 families in mycobacterial species. The new P450 families included: CYP1067 and CYP1119- CYP1134. Among subfamilies 43 new subfamilies were identified in mycobacterial species used in this study. The

lowest number of P450 families was found in *M. tuberculosis* RGTB327 (14 P450 families) and the highest number of P450 families was found in *M. marinum* (36 P450 families), followed by *M. rhodesiae* NBB3 (35 P450 families) (Table 2.3).

Analysis of P450 families revealed that the CYP125 family is the dominant P450 family with 114 members (6.4% of total P450s found in 60 mycobacterial species), followed by CYP189 with 86 members (4.8%), CYP150 with 84 members (4.7%) and CYP136 with 82 members (4.6%) (Table 2.4). CYP150 and CYP189 P450 families were present with a high copy number in mycobacteria, with one to six copies, followed by CYP125 (one to five copies), CYP187 and CYP279 (one to four copies) (Table 2.4). The presence of a high number of member P450s in the above P450 families suggests possible blooming of these P450 families. This implies that the bloomed P450 families play a key role(s) in the physiology of mycobacteria, hence they are present in high copy numbers, as suggested in other organisms (Syed *et al.*, 2014). Ten P450 families, CYP51, CYP123, CYP125, CYP130, CYP135, CYP136, CYP138, CYP140, CYP144 and CYP1128, were conserved across the different mycobacterial categories, suggesting their important role in mycobacteria (see Table 2.4). The P450 families CYP124, CYP126, CYP128, CYP142 and CYP143 were missing only in MCAC (Table .2.2). Interestingly, two P450 families, CYP121 and CYP141, are only present in MTBC. Among 27 MTBC species, CYP121 is present in 23 species and CYP141 is present in 19 species. The presence of CYP121 and CYP141 only in MTBC species suggests that these P450 families can serve as diagnostic markers in the detection of MTBC species. The use of CYP141 as a diagnostic marker in the detection of *M. tuberculosis* was reported earlier (Darban-Sarokhalil *et al.*, 2011). The results from this study however suggest that the use of CYP141 as a diagnostic marker is not limited to the detection of *M. tuberculosis*, but can also be employed in detecting

other MTBC species. Furthermore, this study suggests that CYP121 can serve as a better diagnostic marker compared to CYP141, as 23 species out of 27 contain this P450 (see Dataset 2: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1). The retention/evolution of these two P450 families by MTBC species suggest the key/essential role of these P450 families and experimental data further strengthen this assumption that CYP121 is found to be essential for the survival of *M. tuberculosis* (McLeanet *et al.,* 2008) . The new P450 family CYP1131 is present only in MCAC species. All six MCAC species contain this P450, suggesting that CYP1131 can serve as a diagnostic marker in the detection of MCAC species. Five P450 families, CYP1016, CYP1017, CYP1018, CYP1019 and CYP1067 (new P450 family), are present only in MAC species. Ten P450 families (CYP183, CYP226, CYP269, CYP271, CYP274, CYP276, CYP1120, CYP1123, CYP1127 and CYP1133) were unique to NTM and 11 P450 families (CYP110, CYP145, CYP151, CYP186, CYP289, CYP292, CYP1119, CYP1121, CYP1122, CYP1125, and CYP1126) were present only in SAP species. Among these unique P450 families four and five P450 families were new P450 families that were present in NTM and SAP species, respectively. One of the most studied self-sufficient P450s, CYP102, involved in fatty acid hydroxylation (Li and Poulos, 1997), was found in both MCAC and SAPs. A detailed analysis on the comparison of different P450 families across different mycobacterial categories is presented in Table 2.4.

Despite different geographical locations, *M. bovis* species showed a truncated CYP142 in their genome (see Dataset 3: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1). This suggests that the truncation of the CYP142 P450 event occurred before speciation and thus resulted in a non-functional CYP142 in *M. bovis* species. Further evidence for the non-functionality of CYP142 P450 in these species can be obtained from earlier studies where

CYP125 gene-knockout strains of *M. bovis* and *M. bovis* BCG were unable to grow on cholesterol as a carbon source compared to *M. tuberculosis* H37Rv, where CYP142 is complementing the CYP125 (Driscoll *et al.*, 2010). The CYP51 P450 family that is highly conserved across different living organisms (Lepesheva and Waterman, 2007) is absent from *M. tuberculosis* UT205, *M. canetii* CIPT 140060008, *M. canetii* CIPT 140710010, *M. liflandii* 128FXT and *M. massiliense,* further strengthening the argument that CYP51 is not an essential gene in mycobacteria and it is therefore unlikely to be an azole drug target in mycobacteria, as identified earlier (Sassetti *et al.*, 2003). Interestingly, *M. intracellulare* MOTT-02 contains two copies of CYP51 in its genome (see Dataset 2: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1).

2.3.3. Phylogenetic analysis of mycobacterial P450s

Phylogenetic analysis of mycobacterial P450s revealed that P450s within the same family are clustered together (Fig. 2.2 and Table 2.5).

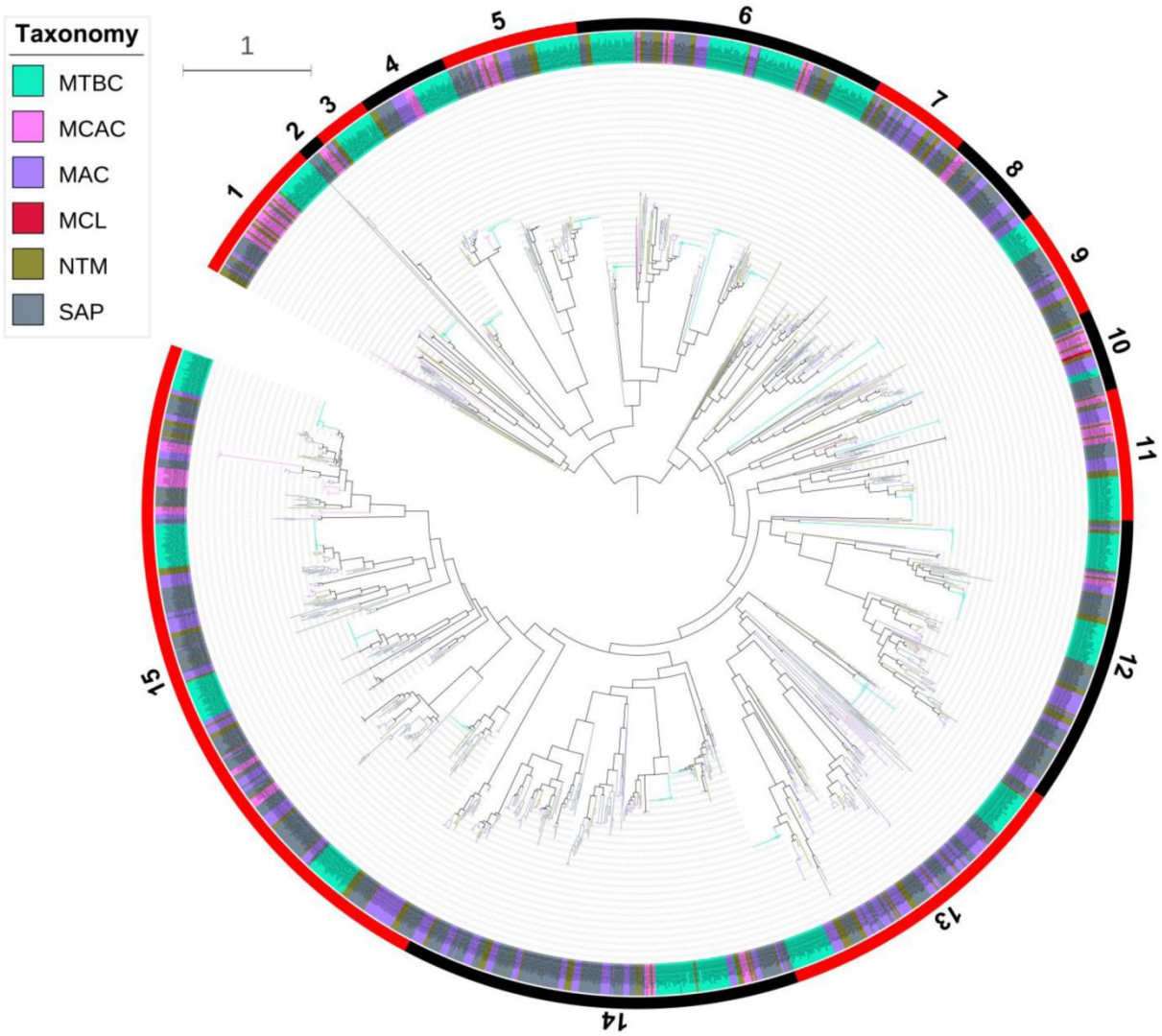

Figure 2.2. Phylogenetic analysis of P450s in mycobacteria. A phylogenetic tree was constructed with 1772 mycobacterial P450s. The inner circle is the phylogenetic tree based on the consensus sequences of the mycobacterial P450s against the Pfam seed PF00067. The branches with different colors show their taxonomic groups: MTBC (*Mycobacterium tuberculosis* complex), MCAC (*Mycobacterium chelonae-abscessus* complex), MAC (*Mycobacterium avium* complex), MCL (Mycobacteria causing leprosy), NTM (Nontuberculous mycobacteria) and SAP (Saprophytes). Ancestral branches with children that had identical colors

were assigned the same color as the children. The middle circle shows the corresponding CYPs, which are covered by different colors to show their taxonomic groups. Each taxon links the branch with a dotted line. The outermost numbers indicate the 15 clades based on this study, and their ranges are marked by alternating red and black. Distribution of P450s families into different clans was listed in Table 2.5.

Table 2.5. Clan-level classification of mycobacterial P450 families. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *Mycobacterium chelonae-abscessus* complex; MAC, *Mycobacterium avium* complex; NTM, Nontuberculous mycobacteria; SAP, Saprophytes.

Clan	CYP family	Group
1	CYP183, CYP274, CYP102, CYP185, CYP1131, CYP1132, CYP1130, CYP1121, CYP1133, CYP1134, CYP132	MTBC, MCAC, MAC, NTM, SAP
2	CYP1119, CYP186, CYP286	SAP, MCAC, NTM, SAP
3	CYP139	MTBC, MAC, NTM
4	CYP51	MTBC, MCAC, MAC, NTM, SAP
5	CYP136	MTBC, MCAC, MAC, NTM, SAP
6	CYP110, CYP137, CYP138, CYP135	MTBC, MCAC, MAC, NTM, SAP
7	CYP1125, CYP226, CYP1034, CYP187	MAC, SAP, NTM, MTBC
8	CYP276, CYP161, CYP151, CYP279	NTM, MCAC, MAC, SAP
9	CYP141, CYP1127, CYP105	MTBC, NTM, MAC, SAP
10	CYP147, CYP1128, CYP164	MTBC, MCAC, MAC, NTM, SAP
11	CYP1067, CYP1110, CYP140	MTBC, MCAC, MAC,

		NTM, SAP
12	CYP278, CYP1016, CYP269, CYP121, CYP1120, CYP1126, CYP144, CYP150	MTBC, MCAC, MAC, NTM, SAP
13	CYP1122, CYP1123, CYP128, CYP145, CYP289, CYP1018, CYP1019, CYP292, CYP1017, CYP1129, CYP243, CYP188, CYP271, CYP272, CYP1027, CYP143	MTBC, MCAC, MAC, NTM, SAP
14	CYP123, CYP130, CYP190, CYP189	MTBC, MCAC, MAC, NTM, SAP
15	CYP191, CYP142, CYP108, CYP291, CYP153, CYP125, CYP126, CYP268, CYP124, CYP1124	MTBC, MCAC, MAC, NTM, SAP

This suggests that the annotation (assigning the family and subfamilies) of P450s in mycobacterial species in this study is correct. Moreover, the phylogenetic relationship of the mycobacterial P450s is also related to species taxonomy. It is evident that the mycobacterial P450s from the same taxonomic group are generally clustered together in the tree. Particularly, P450s in the same family from the group MTBC are clustered together in the branches, suggesting their conserved evolution after speciation into different groups.

In order to understand the evolution of mycobacterial P450s and their relation to their groups, higher-level P450 classifications, i.e. clan level (Nelson, 1998), have been carried out. Clan-level classification of P450s will give an idea of P450 families that probably diverged from a single common ancestor irrespective of their host and possibly with overlapping functions. This type of classification has been reported for P450s belonging to different biological kingdoms (Chen *et al*., 2014; Sello, 2015; Nelson, 1998). In this study, based on their phylogenetic relationships, 77 mycobacterial P450 families were grouped into 15 clans (Fig. 2.2 and Table 2.5). Clan 13 has most P450 families (16 families), followed by clan 1 (11 families), clan 15 (10

families) and clan 12 (8 families). Clans 3-5 contain a single P450 family, suggesting the independent divergence of these P450 families. Analysis of P450 families with respect to mycobacterial groups suggested that most clans consist of members from wide mycobacterial groups, thus these P450 families probably diverged from a common ancestor. Grouping of CYP124, CYP125 and CYP142 P450 families in the same clan (clan 15) strongly authenticates our clan-level classification of mycobacterial P450s, being consistent with the argument that these P450 family members have overlapping functions, as it is evident from experimental data that these P450 families are shown to be involved in cholesterol catabolism in *M. tuberculosis* (Driscoll *et al*., 2010; McLean *et al*., 2009; Ouellet *et al*., 2010; Capyk *et al*., 2009; Frank *et al.*, 2014). In addition to cholesterol catabolism, CYP124 and CYP153 P450s are involved in oxidation of aliphatic hydrocarbons (Johnston *et al*., 2009; Funhoff *et al*., 2006) and were also grouped in the same clan, indicating divergence in function for CYP124 that is possibly acquired during the course of evolution. Considering that most mycobacterial P450s are orphans, future functional characterization of mycobacterial P450s will provide more information on clan-level grouping of P450 families with respect to their functional relationship.

2.3.4. Mycobacterial species have the highest P450 diversity

P450 diversity is a good indication of diverse P450 families that organisms harbor. Low P450 diversity suggests blooming of certain P450 families, possibly *via* duplication of the same P450 in an organism (Syed *et al*., 2014; Sello *et al*., 2015; Feyereisen, 2011). Analysis of P450s in microbes, particularly lower eukaryotes such as fungi and oomycetes, suggested that fungal species have the highest P450 diversity in their genomes (Sello *et al.*, 2015).

In this study, analysis of P450 diversity between different mycobacterial categories revealed that MTBC species have the highest P450 diversity percentage in their genomes (Fig. 2.3 and Table 2.6). The P450 diversity percentage between different mycobacterial groups is as follows: MTBC > NTM > MCAC > MAC > SAP (Fig. 2A). The low P450 diversity percentage observed in MAC and SAP group species is due to the fact that certain P450 family members are populated in their genomes (Table 2.6 and Dataset 2: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1). For example, five and six copies of CYP189 P450s are present in MAC and SAP; up to four copies of CYP125 and CYP187 are present in both groups; six and five copies of CYP150 P450s are present in MAC and SAP respectively (Table 2.4). Overall, 60 mycobacterial species showed an average P450 diversity percentage of 72%.

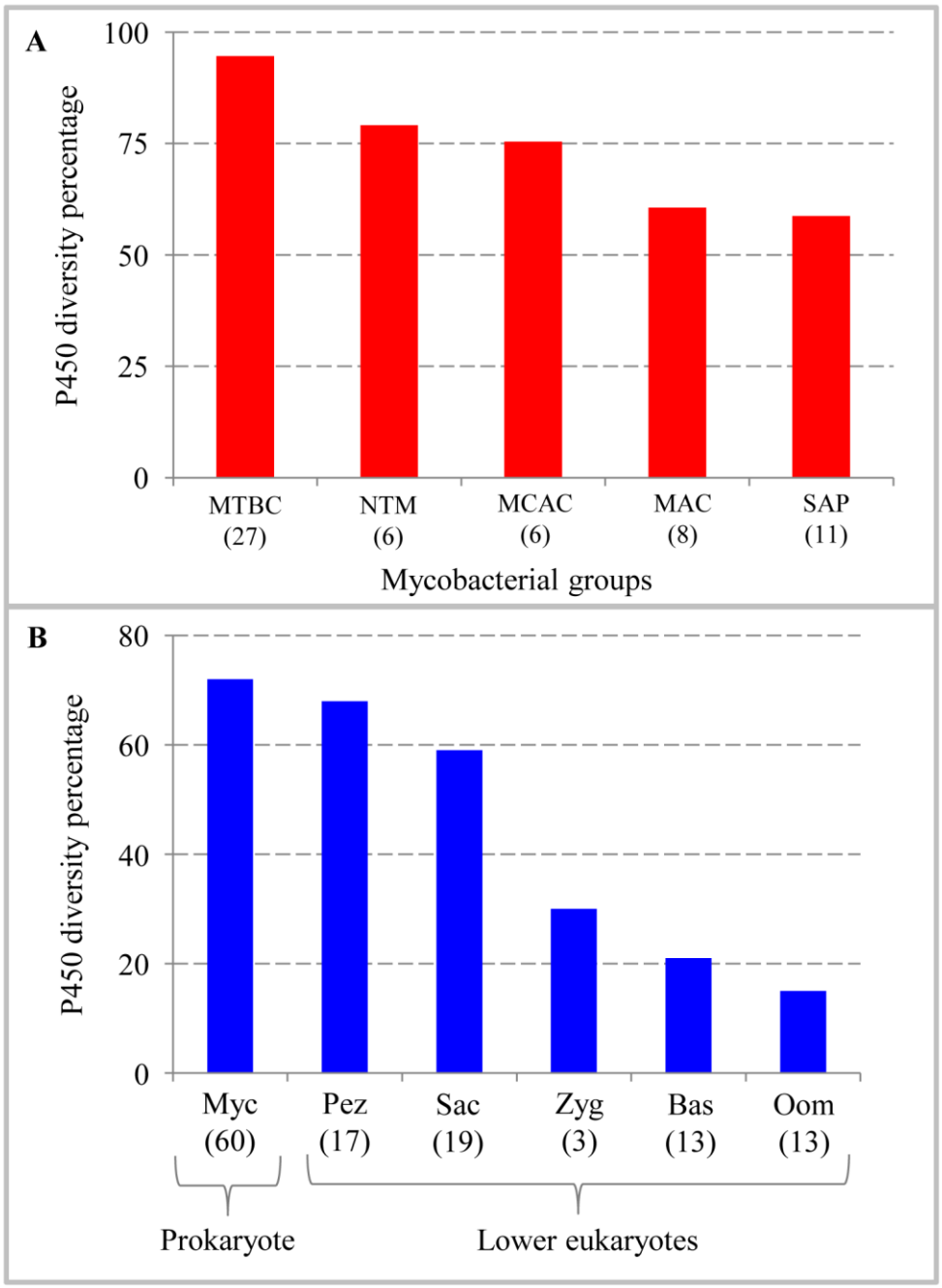

Figure 2.3. P450 diversity percentage analysis. (**A**) Comparative analysis of P450 diversity percentage between different mycobacterial categories. (**B**) Comparative analysis of P450 diversity percentage between microbes such as prokaryote mycobacterial species and lower eukaryote fungi and Oomycetes. The number of species in each of the groups (**A**) or microbe (**B**)

used for this analysis is shown in parenthesis. Detailed analysis of the P450 diversity percentage values were presented in Supplementary Table S6. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *Mycobacterium chelonae-abscessus* complex; MAC, *Mycobacterium avium* complex; MCL, Mycobacteria causing leprosy; NTM, nontuberculous mycobacteria; SAP, Saprophytes; Myc, Mycobacterial species; Sac, Saccharomycetes; Pez, Pezizomycetes; Bas, Basidiomycetes; Zyg, Zygomycetes and Oom, Oomycetes.

Table 2.6. P450 diversity percentage analysis in mycobacterial species. Comparative analysis of P450 diversity percentage between mycobacterial species with other lower eukaryote microbes such as fungi and oomycetes are also shown in the table. The P450 diversity percentage data for lower eukaryotes such as fungi and oomycetes were retrieved from published literature (Nelson, 2009; Syed *et al*., 2013; Kgosiemang *et al*., 2014; Syed *et al*., 2014; Sello *et al*., 2015; Chen *et al*., 2014).

Group	**Number of species**	**P450 diversity percentage**		
MTBC	27	95		
NTM	6	79		
MCAC	6	75		
MAC	8	61		
SAP	11	59		
	Number of species	**Average number of P450s**	**Average number of P450 families**	**Average P450 diversity percentage**
Mycobacterial species	60	29.75	21.5	72
Saccharomycetes	19	8.11	4.79	59
Pezizomycetes	17	87.35	59.71	68
Basidiomycetes	13	157.00	32.85	21
Zygomycetes	3	50.33	15.00	30
Oomycetes	13	27.38	4.00	15

Comparative analysis of the P450 diversity percentage between different microbial populations such as mycobacterial species and other lower eukaryote microbes (fungi and oomycetes) revealed that the mycobacterial species have the highest P450 diversity percentage in their genomes (Fig. 2.3B and Table 2.6). The P450 diversity percentage observed in mycobacterial species (average 72%) is slightly higher compared to Pezizomycetes (68%) and is clearly the highest compared to other fungal categories and oomycetes. Future analysis of P450s in a greater number of bacterial species will provide a clear answer on whether the observed highest P450 diversity percentage in mycobacterial species is a general characteristic of prokaryotes, and thus provide information of the P450 diversity percentage of prokaryotic vs eukaryotic microbes.

2.3.5. Some P450s show the highest amino acid conservation in Mycobacteria

During the annotation of mycobacterial P450s (this study) we observed that the mycobacterial P450s belonging to the same family showed the highest percentage identity, suggesting the highest conservation in member P450s' primary structure, i.e. amino acid level. This phenomenon, the highest number of amino acids conserved in the primary structure of P450, was recently reported for CYP53 family members in fungi (Jawallapersand *et al.*, 2014). In order to understand the nature and extent of conserved amino acids, mycobacterial P450 families were subjected to PROMALS3D analysis (Pei *et al.*, 2008). PROMALS3D analysis of 32 mycobacterial P450 families revealed the highest number of amino acids conserved in some of the P450 families (Fig. 2.4 and Table 2.7). As shown in Fig. 2.4, the CYP141 P450 family has the highest number of conserved amino acids (389 amino acids) in its members, followed by CYP121, CYP132, CYP137 and CYP51. It is noteworthy that CYP141 and CYP121 P450 families are present only in MTBC species and both families have the highest conservation in

their primary structure, indicating that these P450 families were protected from any changes in their primary structure. Changes in the primary structure might be deleterious, as CYP121 is known to be essential for *M. tuberculosis* survival (McLean *et al.,* 2008).

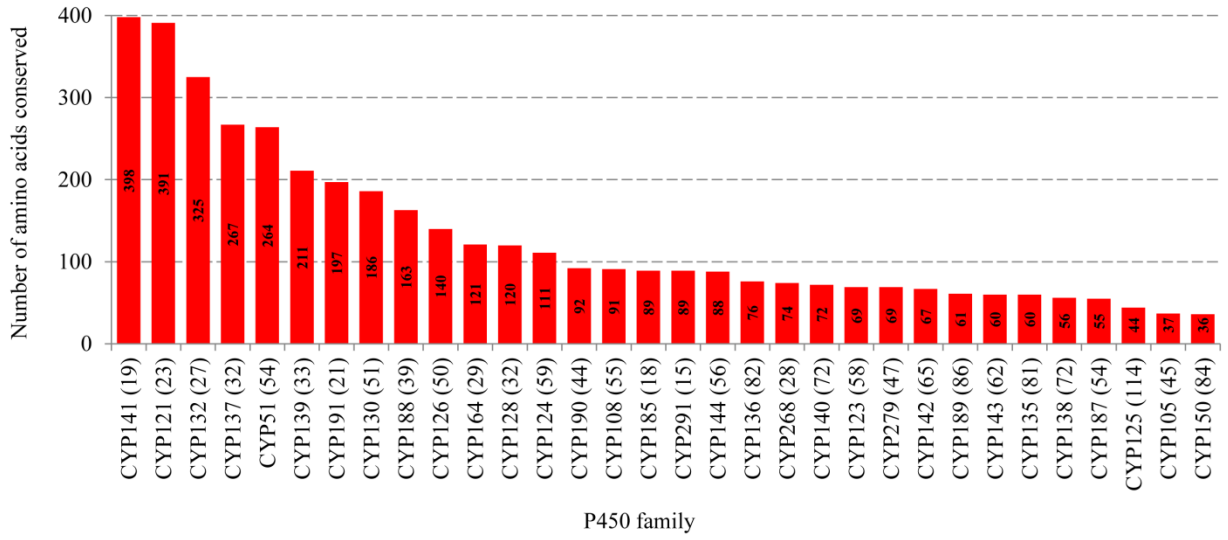

Figure 2.4. Conserved amino acid analysis in mycobacterial P450 families. Numbers of amino acids that are conserved in member P450s were determined using PROMALS3D31 and presented in the figure with conservation index where the number "9" indicates conserved amino acid in P450 family members. Number of member P450s analyzed for each P450 family is presented in parenthesis next to P450 family. A PROMALS3D analysis of member P450s and the conservation index scores for each mycobacterial P450 family were presented in Table 2.7.

Table 2.7. Conserved amino acid analysis in mycobacterial P450s. The conservation score (5-9) obtained *via* PROMALS3D is shown in the table, where the number "9" indicates conserved amino acid in P450 members.

P450 family	Number of member P450s	PROMALS3D conservation score				
		5	6	7	8	9
CYP141	19	0	0	0	0	398
CYP51	54	11	102	0	0	264
CYP137	32	146	0	0	0	267
CYP132	27	0	0	0	0	325
CYP121	23	0	0	0	0	391
CYP191	21	56	51	54	0	197
CYP140	72	65	24	57	0	72
CYP139	33	130	0	0	0	211
CYP124	59	60	23	86	0	111
CYP126	50	54	89	36	0	140
CYP123	58	86	9	76	0	69
CYP108	55	76	9	90	0	91
CYP130	51	48	35	66	0	186
CYP190	44	54	14	82	0	92
CYP144	56	33	7	88	0	88
CYP185	18	84	23	82	0	89
CYP291	15	0	128	0	0	89
CYP188	39	66	34	76	0	163
CYP136	82	61	20	76	0	76
CYP268	28	61	29	62	0	74
CYP142	65	99	55	48	0	67
CYP143	62	66	12	73	0	60
CYP135	81	51	30	73	0	60
CYP138	72	44	38	31	30	56
CYP128	32	0	147	0	0	120
CYP189	86	42	36	37	21	61
CYP164	29	58	30	60		121
CYP187	54	37	29	60		55
CYP150	84	43	50	23	30	36
CYP279	47	42	31	74	1	69
CYP105	45	41	18	34	23	37
CYP125	114	43	34	41	18	44

The highest conservation of amino acids in CYP121 and CYP141 is quite understandable, considering these families exist only in MTBC species. The lowest number of conserved amino acids was observed for P450 families CYP150 (36 amino acids), CYP105 (37 amino acids) and CYP125 (44 amino acids). This might suggest that P450 families that are populated across mycobacterial species (Table 2.4) are subjected to changes in their primary structure. This is however not true, as P450 families such as CYP51, CYP137, CYP132, CYP191 and CYP140 that are present across different mycobacterial categories (Table 2.4) showed the highest number of conserved amino acids (Fig. 2.4). This clearly suggests that irrespective of the widespread nature of a P450 family, some of the P450 families have been highly conserved despite their divergence into different mycobacterial species.

2.3.6. Bacterial P450s show highest amino acid conservation across biological kingdoms

The highest conservation of mycobacterial P450s at primary structure level (as shown above) led us to assess the amino acid conservation in the P450 families across biological kingdoms to see the conservation pattern with respect to biological kingdoms, if any, and also to assess where Bacterial P450s, particularly P450 families present in mycobacterial species, rank among other P450 families.

To this end, 17 598 P450 sequences belonging to 113 P450 families (Dataset 1: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018878/#S1) were collected, as described in the Methods. Among the P450 families, 42 belong to bacteria, 28 to plants, 22 to animals and 19 to fungi (excluding bacteria and fungi specific CYP51 members). CYP74 is common between plants and animals and CYP51 is common among all biological kingdoms. Analysis of conserved amino acids in 113 P450 families revealed that among the top 20 conservation ranked

P450 families, 18 P450 families belonged to bacteria, three belonged to animals and one family each belonged to fungi and plants (Fig. 2.5A and Table 2.8).

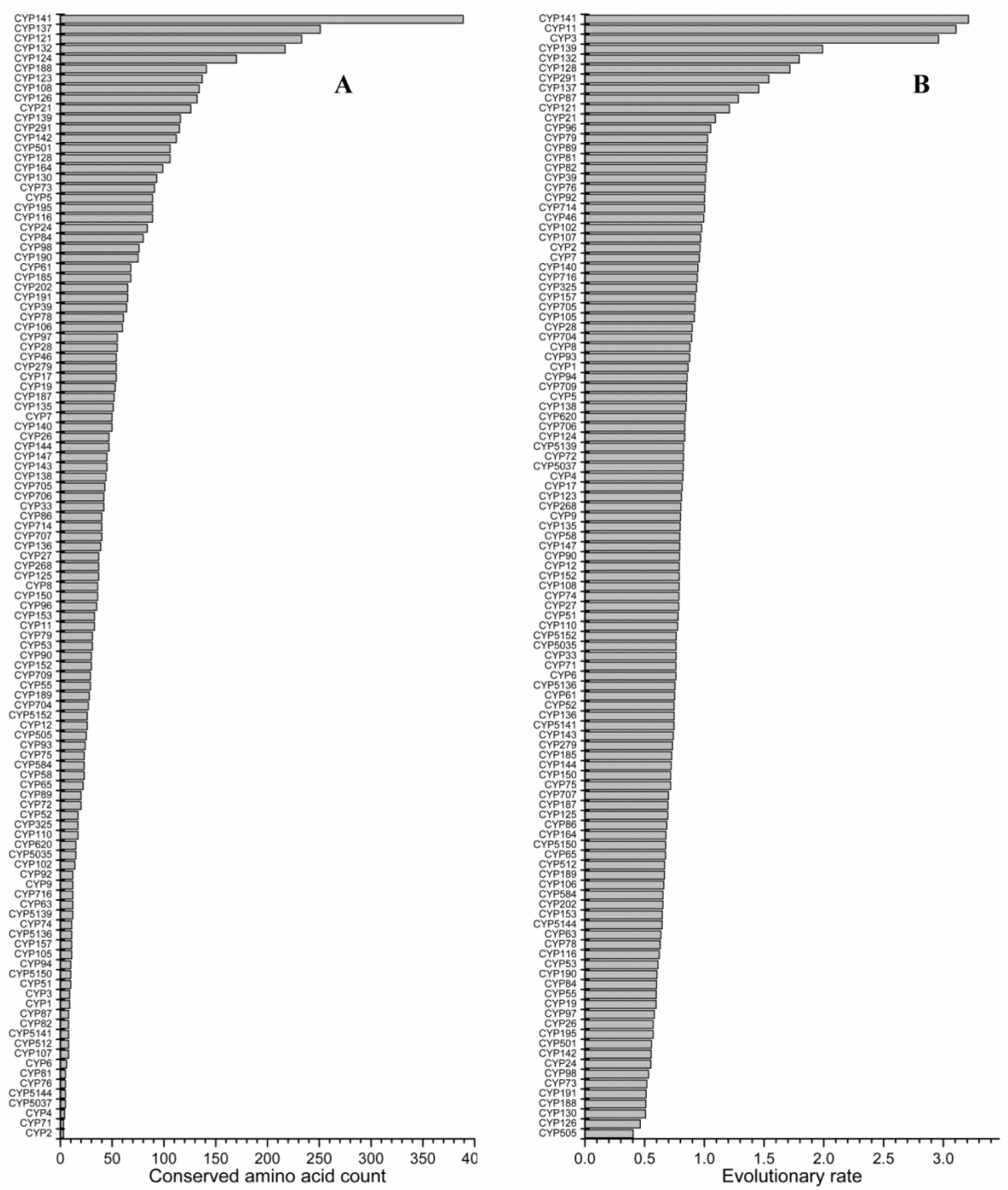

Figure 2.5. Protein-level (A) and DNA-level (B) P450s structural dynamic analysis. Structural dynamics for 17598 P450s belonging to 113 P450 families from different biological kingdoms were analyzed. (**A**) Protein-level structure dynamics were assessed based on the number of conserved amino acids present in each P450 family. The P450 families in the graph are presented such that the P450 family CYP141 that has the highest number of conserved amino acids is on top of the graph and the lowest number of conserved amino acids observed for the CYP2 family is on the bottom of the graph. A detailed analysis of the number of conserved amino acids and number of member P450s used and their hosts (biological kingdoms) and conservation ranking for each member of the family is presented in Table 2.8. (**B**) Evolutionary rate analysis of P450s. Evolutionary rates were estimated based on their cDNA sequences under the Tamura-Nei model40. A discrete Gamma distribution was used to model the evolutionary rate differences, as X axis of Fig. 2.5B presents, and more details are provided in Methods section. A discrete Gamma distribution was used to model the evolutionary rate differences. The P450 families in the graph are presented such that the P450 family CYP141 showing the lowest evolutionary rate is on top of the graph and the highest evolutionary rate observed for the CYP505 family is on the bottom of the graph. A detailed analysis of evolutionary rates for each of the P450 family and their ranking is presented in Table 2.9.

Table 2.8. Amino acid conservation analysis in P450 families from different biological kingdoms. The conservation score (5-9) obtained *via* PROMALS3D is shown in the table where the number "9" indicates conserved amino acid in P450 members. P450 families were arranged in order of the highest to the lowest number of amino acids conserved.

P450 family	Number of member P450s	Kingdom	PROMALS3D conservation Index					Rank (highest to lowest conservation)
			5	6	7	8	9	
CYP141	29	Bacteria	0	0	0	0	389	1
CYP51	50	Bacteria	11	102	0	0	264	2
CYP137	38	Bacteria	145	0	0	0	251	3
CYP121	34	Bacteria	0	0	0	0	233	4
CYP132	39	Bacteria	175	0	0	0	217	5
CYP124	71	Bacteria	52	35	59	0	170	6
CYP188	67	Bacteria	62	0	100	0	141	7
CYP123	74	Bacteria	62	0	82	0	137	8
CYP108	67	Bacteria	52	12	92	0	134	9
CYP126	78	Bacteria	65	16	98	0	132	10
CYP21	84	Animal	84	30	104	0	126	11
CYP139	54	Bacteria	126	0	0	0	116	12
CYP291	23	Bacteria	0	111	0	0	115	13
CYP142	90	Bacteria	60	6	83	0	112	14
CYP501	106	Fungi	44	60	113	0	106	15
CYP128	49	Bacteria	0	132	0	0	106	15
CYP164	50	Bacteria	49	9	94	0	99	16
CYP130	98	Bacteria	31	31	91	0	93	17
CYP73	155	Plant	52	69	57	71	91	18
CYP5	71	Animal	37	48	117	0	89	19
CYP116	93	Bacteria	98	92	39	55	89	19
CYP195	52	Bacteria	35	38	54	21	89	19
CYP24	65	Animal	87	11	96	0	84	20
CYP84	62	Plant	68	64	48	44	80	21
CYP98	77	Plant	68	83	44	48	76	22
CYP190	76	Bacteria	34	19	87	0	75	23
CYP185	42	Bacteria	65	20	72	0	68	24
CYP61	70	Fungi	54	63	38	25	68	24
CYP191	36	Bacteria	92	29	92	0	65	25

CYP202	84	Bacteria	34	34	27	23	65	25
CYP39	51	Animal	51	47	79	0	64	26
CYP78	112	Plant	40	51	36	40	61	27
CYP106	93	Bacteria	38	27	64	9	60	28
CYP28	51	Animal	84	59	44	28	55	29
CYP97	100	Plant	37	49	29	25	55	29
CYP17	99	Animal	56	68	44	29	54	30
CYP46	53	Animal	53	42	99	0	54	30
CYP279	71	Bacteria	38	49	26	37	54	30
CYP19	176	Animal	53	57	44	60	53	31
CYP187	103	Bacteria	35	33	42	13	52	32
CYP135	124	Bacteria	48	46	22	23	51	33
CYP140	113	Bacteria	38	32	21	26	50	34
CYP7	89	Animal	69	43	34	26	50	34
CYP144	107	Bacteria	45	33	34	35	47	35
CYP26	131	Animal	69	63	30	46	47	35
CYP143	103	Bacteria	43	43	28	30	45	35
CYP147	52	Bacteria	45	29	35	32	45	35
CYP138	114	Bacteria	42	44	24	32	44	36
CYP51	82	Fungi	64	42	39	33	43	37
CYP705	50	Plant	66	49	41	28	43	37
CYP33	67	Animal	61	49	34	40	42	38
CYP706	51	Plant	53	49	34	32	42	39
CYP86	139	Plant	64	39	34	29	40	40
CYP707	100	Plant	59	54	47	57	40	40
CYP714	58	Plant	58	43	25	26	40	40
CYP136	171	Bacteria	48	34	25	28	39	41
CYP268	73	Bacteria	43	43	27	33	37	42
CYP125	103	Bacteria	42	39	29	17	37	42
CYP27	116	Animal	66	46	25	25	37	42
CYP150	164	Bacteria	59	45	20	35	36	43
CYP8	91	Animal	65	34	36	37	36	43
CYP96	61	Plant	57	40	29	30	35	44
CYP11	171	Animal	46	47	27	29	33	45
CYP153	164	Bacteria	41	34	37	28	33	45
CYP53	102	Fungi	64	45	48	33	31	46
CYP79	102	Plant	67	55	30	31	31	46
CYP90	117	Plant	63	44	44	30	30	47
CYP152	90	Bacteria	46	38	19	25	30	47
CYP55	60	Fungi	34	56	51	21	29	48
CYP709	113	Plant	50	42	18	26	29	48

CYP189	188	Bacteria	54	37	17	27	28	49
CYP704	109	Plant	45	43	33	43	27	50
CYP5152	66	Fungi	46	28	25	15	26	51
CYP12	119	Animal	59	37	23	20	26	51
CYP505	165	Fungi	95	76	48	31	25	52
CYP93	151	Plant	68	42	26	20	24	53
CYP58	106	Fungi	52	43	29	21	23	54
CYP75	251	Plant	74	48	31	29	23	54
CYP584	96	Fungi	56	35	30	21	23	54
CYP65	203	Fungi	38	25	13	11	22	55
CYP72	208	Plant	57	27	28	27	20	56
CYP89	134	Plant	48	38	28	25	20	56
CYP52	161	Fungi	51	26	24	19	17	57
CYP110	113	Bacteria	50	27	17	18	17	57
CYP325	53	Animal	50	31	19	10	17	57
CYP5035	129	Fungi	37	30	17	9	15	58
CYP620	178	Fungi	33	21	12	17	15	58
CYP102	333	Bacteria	45	43	16	19	14	59
CYP63	133	Fungi	51	39	23	13	12	60
CYP5139	181	Fungi	35	12	3	9	12	60
CYP9	312	Animal	45	24	9	10	12	60
CYP92	167	Plant	51	32	33	17	12	60
CYP716	103	Plant	53	30	17	23	12	60
CYP5136	68	Fungi	44	42	20	15	11	61
CYP74	159	Plant/Animal	40	29	23	19	11	61
CYP105	329	Bacteria	28	14	19	6	11	61
CYP157	115	Bacteria	44	26	26	18	11	61
CYP5150	336	Fungi	48	26	8	7	10	62
CYP51	409	Bacteria/fungi /animal/plant	36	19	11	13	10	62
CYP94	170	Plant	39	44	34	26	10	62
CYP1	289	Animal	67	28	26	24	9	63
CYP3	248	Animal	59	44	27	20	9	63
CYP5141	86	Fungi	40	21	13	11	8	64
CYP512	247	Fungi	25	12	8	5	8	64
CYP82	174	Plant	68	31	20	21	8	64
CYP87	78	Plant	58	29	23	20	8	64
CYP107	217	Bacteria	40	25	8	12	8	64
CYP6	921	Animal	21	9	7	8	6	65
CYP5037	261	Fungi	19	6	7	5	5	65
CYP5144	514	Fungi	10	4	5	3	5	65

CYP76	206	Plant	56	31	21	10	5	65
CYP81	235	Plant	40	30	29	14	5	65
CYP4	1076	Animal	27	11	5	6	4	66
CYP2	857	Animal	48	18	10	9	3	67
CYP71	780	Plant	22	7	6	5	3	67

Table 2.9. Evolutionary rate analysis of P450 families. The maximum likelihood estimates for P450 evolutionary rates. Substitution pattern and rates were estimated under the Tamura-Nei model (+G) (Tamura and Nei, 1993). A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate of substitution for each site is drawn from a Gamma distribution with shape parameter α. If α is <1, the distribution implies that there is a relatively large amount of rate variation, with many sites evolving very slowly but some sites evolving at a high rate. For values of α >1, the shape of the distribution changes qualitatively, with less variation and most sites having roughly similar rates (Lio and Goldman, 1998).

CYP family	Biological kingdom	Shape parameter (α)	Ranking (highest to lowest rate of evolution)
CYP505	Fungi	0.4034	1
CYP126	Bacteria	0.4646	2
CYP130	Bacteria	0.5066	3
CYP188	Bacteria	0.5089	4
CYP191	Bacteria	0.5125	5
CYP73	Plant	0.5186	6
CYP98	Plant	0.5341	7
CYP24	Animal	0.5525	8
CYP142	Bacteria	0.555	9
CYP501	Fungi	0.557	10
CYP195	Bacteria	0.5709	11
CYP26	Animal	0.5711	12
CYP97	Plant	0.5816	13
CYP19	Animal	0.5959	14
CYP55	Fungi	0.5982	15
CYP84	Plant	0.6002	16
CYP190	Bacteria	0.6028	17
CYP53	Fungi	0.6119	18
CYP116	Bacteria	0.6236	19
CYP78	Plant	0.6308	20
CYP63	Fungi	0.6371	21
CYP5144	Fungi	0.647	22
CYP153	Bacteria	0.6485	23
CYP202	Bacteria	0.654	24
CYP584	Fungi	0.654	24

CYP106	Bacteria	0.6603	25
CYP189	Bacteria	0.6673	26
CYP512	Fungi	0.6681	27
CYP65	Fungi	0.6746	28
CYP5150	Fungi	0.6756	29
CYP164	Bacteria	0.6781	30
CYP86	Plant	0.6854	31
CYP125	Bacteria	0.6934	32
CYP187	Bacteria	0.695	33
CYP707	Plant	0.6989	34
CYP75	Plant	0.7185	35
CYP150	Bacteria	0.7188	36
CYP144	Bacteria	0.7228	37
CYP185	Bacteria	0.726	38
CYP279	Bacteria	0.7342	39
CYP143	Bacteria	0.7398	40
CYP5141	Fungi	0.7465	41
CYP136	Bacteria	0.7472	42
CYP52	Fungi	0.7476	43
CYP61	Fungi	0.7514	44
CYP5136	Fungi	0.7543	45
CYP6	Animal	0.7618	46
CYP71	Plant	0.7618	46
CYP33	Animal	0.763	47
CYP5035	Fungi	0.7634	48
CYP5152	Fungi	0.7641	49
CYP110	Bacteria	0.7748	50
CYP51	Animal/Plant/ Bacteria/Fungi	0.7794	51
CYP27	Animal	0.7858	52
CYP74	Plant/Animal	0.7879	53
CYP108	Bacteria	0.7896	54
CYP152	Bacteria	0.7904	55
CYP12	Animal	0.7908	56
CYP90	Plant	0.7922	57
CYP147	Bacteria	0.794	58
CYP58	Fungi	0.7948	59
CYP135	Bacteria	0.7983	60
CYP9	Animal	0.8016	61
CYP268	Bacteria	0.8037	62
CYP123	Bacteria	0.8084	63

CYP17	Animal	0.8148	64
CYP4	Animal	0.8196	65
CYP5037	Fungi	0.8245	66
CYP72	Plant	0.8262	67
CYP5139	Fungi	0.8267	68
CYP124	Bacteria	0.8361	69
CYP706	Plant	0.8369	70
CYP620	Fungi	0.8391	71
CYP138	Bacteria	0.8462	72
CYP5	Animal	0.8516	73
CYP709	Plant	0.8529	74
CYP94	Plant	0.8562	75
CYP1	Animal	0.8636	76
CYP93	Plant	0.8768	77
CYP8	Animal	0.8791	78
CYP704	Plant	0.8941	79
CYP28	Animal	0.8978	80
CYP105	Bacteria	0.915	81
CYP705	Plant	0.9213	82
CYP157	Bacteria	0.9259	83
CYP325	Animal	0.9344	84
CYP716	Plant	0.9416	85
CYP140	Bacteria	0.9464	86
CYP7	Animal	0.9584	87
CYP2	Animal	0.9639	88
CYP107	Bacteria	0.969	89
CYP102	Bacteria	0.9767	90
CYP46	Animal	0.995	91
CYP714	Plant	1.0019	92
CYP92	Plant	1.0027	93
CYP76	Plant	1.0065	94
CYP39	Animal	1.0091	95
CYP82	Plant	1.0158	96
CYP81	Plant	1.0219	97
CYP89	Plant	1.0247	98
CYP79	Plant	1.0276	99
CYP96	Plant	1.053	100
CYP21	Animal	1.0927	101
CYP121	Bacteria	1.2102	102
CYP87	Plant	1.2841	103
CYP137	Bacteria	1.4553	104

CYP291	Bacteria	1.5395	105
CYP128	Bacteria	1.7168	106
CYP132	Bacteria	1.7932	107
CYP139	Bacteria	1.9896	108
CYP3	Animal	2.9595	109
CYP11	Animal	3.1065	110
CYP141	Bacteria	3.2104	111

Among the 18 bacterial P450s, seven P450 families were exclusively present in mycobacterial species (CYP108, CYP121, CYP123, CYP124, CYP132, CYP137 and CYP141). The animal CYP21 P450 family was ranked 11 and CYP501 and CYP73 from fungi and plants were ranked 15 and 18, respectively. The animal CYP5 family ranked 19. Interestingly, bacterial CYP51 members ranked 2 compared to fungal CYP51 members (rank 37) and in general CYP51 members from all biological kingdoms (rank 62), suggesting that after speciation, CYP51 in each biological kingdom had been subjected to speciation/kingdom-specific primary structure conservation, as observed in a study described elsewhere (Chen *et al*., 2014).

The amino acid conservation ranking is independent of the number of member P450s, as some of the P450 families, although they contain a low number of members, for example CYP325 from animals (53 members), CYP706 and CYP87 from plants (51 and 78 members, respectively), members) and CYP147 from bacteria (52 members) were ranked 57, 39 and 64, and 35, respectively (Table 2.8). In contrast to the CYP5 P450 family (ranked 19), animal P450 families such as human P450 families CYP1 and 3, CYP2 and CYP4 ranked 63, 67 and 66, respectively. Interestingly, fungal P450 families CYP5141, CYP512, CYP5144 and CYP5037 that ranked above 63 are shown to have bloomed in fungi *via* duplication of their member P450s (Syed *et al.*, 2014), hence their variation in primary structure. Overall, based on the conserved amino acid analysis (see Table 2.8), it can be concluded that after the origin and divergence of

P450 families into different biological kingdoms, some P450 families ranked in the top 20 (see Table 2.8) have been conserved in organisms. The reason for the conservation of these P450 families is clearly that these families possibly play a key role in organisms' physiology. For example, two P450 families, CYP121 and CYP128, are known to be critical for the survival of *M. tuberculosis* (Sassetti *et al*., 2003; McLean *et al.*, 2008), suggesting that any changes in their primary structure might create a problem for the organism's survival. The P450 families that are highly ranked are the ones prone to diversifying their substrate range, thus possibly serving to generate new P450s. The broad substrate specificity of P450 families such as CYP1-3 (from animals) (Guengerich, 2015) and CYP5141, CYP512, CYP5144 and CYP5037 (from fungi) (Syed *et al.*, 2014) strongly supports this argument.

2.3.7. Evolutionary rate analysis of P450 families

In order to assess whether the conservation of amino acids observed for the top 20 ranked P450 families (as discussed in the above section) will align at the DNA level, we performed evolutionary rate analysis using the cDNAs of P450 family members (Fig. 2.5B and Table 2.9). Evolutionary rate analysis revealed that the CYP505 family showed the highest evolutionary rate at the DNA level. It is noteworthy that this family showed the lowest amino acid conservation, suggesting substitutions were nonsynonymous. Among the top 20 ranked P450 families for the highest evolutionary rate, eight P450 families belonged to bacteria, three P450 families to animals, four P450 families to fungi and six P450 families to plants (Fig. 2.5B and Table 2.9). Evolutionary rate analysis of P450 families revealed that among the top 20 conserved bacterial P450 families at the protein level only eight P450 families, the CYP21, CYP121, CYP128, CYP132, CYP137, CYP139, CYP141 and CYP291 P450s families, showed the lowest evolutionary rate at the DNA level. This indicates these P450 families are conserved both at the

DNA and the protein levels. In the P450 families such as CYP24, CYP73, CYP116, CYP126, CYP130, CYP142, CYP188, CYP195 and CYP501, despite showing the highest conservation at the protein level, DNA-level analysis revealed the highest evolutionary rate (ranking below 20) for these families, suggesting that nucleotide substitutions are synonymous (Fig. 2.5B and Table 2.9). This suggests that changes occurring at the DNA level in these P450 families are neutral or synonymous substitutions, since they do not alter the protein sequence. Some of the other top 20 conserved P450 families at protein level, such as CYP5, CYP108, CYP123, CYP124 and CYP164, also showed the highest evolutionary rate at the DNA level, suggesting most of the changes were synonymous. It is quite interesting to observe that the CYP1-3 families that showed the lowest protein-level conservation surprisingly showed the lowest evolutionary rate at the DNA-level, as CYP1-3 ranked 76, 88 and 109, respectively, suggesting the DNA-level changes were nonsynonymous substitutions and led to a change in the amino acid sequence, thus altering protein function. It is noteworthy that these P450s are known for their catalytic versatility and the single nucleotide polymorphism of CYP1-3 is well known to change the response to xenobiotics such as drugs etc. (Guengerich, 2015). The fungal P450 families with the lowest conserved amino acids, such as CYP5144, CYP512, and CYP5141, showed a moderate evolutionary rate as they ranked 22, 27 and 41, respectively, suggesting DNA-level changes were nonsynonymous substitutions.

2.3.8. P450 family dynamics of divergence

From the above results it is clear that irrespective of their parent organism (bacteria, fungi, plants and animals) P450s follow an independent evolutionary route either for their conservation or diversity at the protein or DNA level. This poses the question on how these P450 families diverged into different organisms since their origin.

In order to understand the dynamics of divergence of P450 families, we carried out evolutionary analysis of 17 598 P450 sequences from 113 P450 families covering the biological kingdoms animals, bacteria, fungi and plants. Based on their phylogenetic relationships, these 113 families were grouped into 15 clans (Fig. 2.6 and Table 2.10). As indicated earlier, the P450 families grouped into a clan possibly diverged from a common ancestor irrespective of their host (Nelson, 1999). P450 families grouped into clans 4 and 6 are distributed among all biological kingdoms, suggesting the ancestral P450 of these P450 families was evolved before the evolution of different biological kingdoms. Clan 11 and clan 15 contain P450 families that are distributed between animals and fungi and animals and bacteria, respectively. Clans 1 and 14 contain P450 families of plants and fungi, whereas clan 5 contains P450 families of plants and bacteria. The presence of P450 families belonging to two different biological kingdoms in the same clan suggests that these P450 families originated from a common ancestral P450 that evolved before the divergence of biological kingdoms. Clans 2, 3, 7 and 13 contain animal P450 families, suggesting an independent origin of these P450 families. The same applies to P450 families distributed among clans 8, 9 and 10 that contain exclusively bacterial P450s and clan 12 that contains plant P450s (Fig. 2.6 and Table 2.10). Considering the presence of these P450 families in a single biological kingdom, it appears that these P450 families originated after the divergence of biological kingdoms.

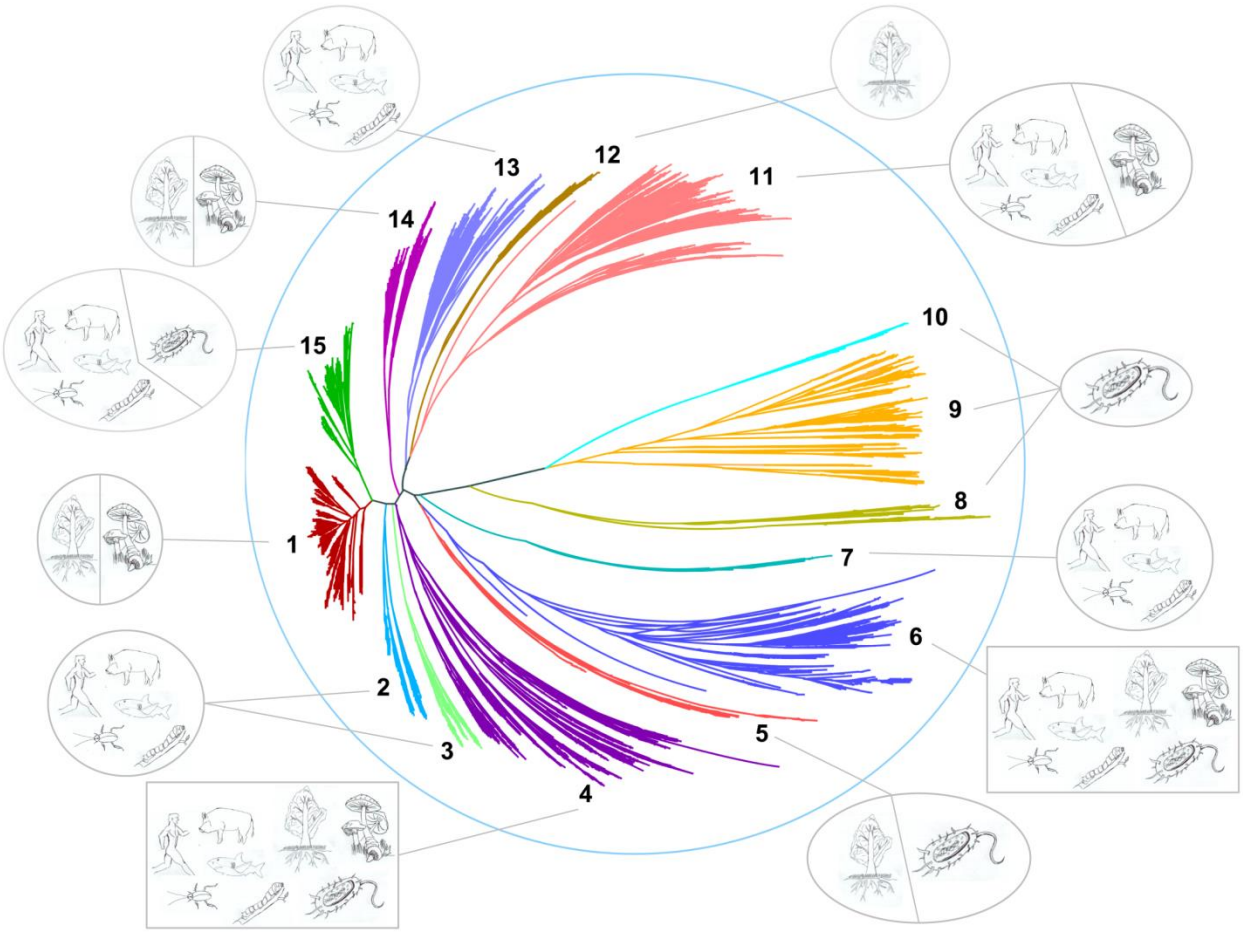

Figure 2.6. Phylogenetic analysis of 17 598 P450s belonging to 113 families from different biological kingdoms such as bacteria, fungi, animals and plants. All 113 P450 families were grouped under 15 clans based on their phylogenetic relationships and following the methods described elsewhere (Chen *et al*., 2014). Clans were presented in different colors and the host containing the respective P450 families that grouped under each clan is shown in schematic diagrams. Schematic diagrams are representative only and the P450 family is not necessarily confined to the depicted animals. Cartoon figures for plants, bacteria and fungi are shown as representative of their kingdoms. The tree was viewed by Hypertree in fisheye pattern. Detailed

information on clan-level grouping of 113 P450 families is presented in Table 2.10 and functional data on P450s at family level is presented in Table 2.11.

Table 2.10. Clan-level classification of 113 P450 families belonging to different biological kingdoms such as bacteria, fungi, animal and plant. A hundred and thirteen P450 families were classified into different clans based on their phylogenetic relationships, as shown in Fig. 2.6 and following the procedure described elsewhere (Chen *et al*., 2014).

Clan	CYP family	Biological kingdoms
1	CYP71, CYP73, CYP75, CYP76, CYP78, CYP79, CYP81, CYP82, CYP84, CYP89, CYP92, CYP93, CYP98, CYP620, CYP705, CYP706, CYP5037, CYP5144, CYP5152	Plant, Fungi
2	CYP7, CYP8, CYP19, CYP39	Animal
3	CYP11, CYP12, CYP24, CYP27	Animal
4	CYP26, CYP51, CYP53, CYP58, CYP61, CYP65, CYP87, CYP90, CYP110, CYP116, CYP135, CYP136, CYP137, CYP138, CYP501, CYP512, CYP707, CYP716, CYP5035	Animal, Fungi, Plant, Bacteria
5	CYP97, CYP132, CYP185	Plant, Bacteria
6	CYP55, CYP74, CYP102, CYP105, CYP106, CYP141, CYP153, CYP505	Fungi, Plant, Animal, Bacteria
7	CYP46	Animal
8	CYP157	Bacteria
9	CYP101, CYP107, CYP108, CYP121, CYP123, CYP124, CYP125, CYP126, CYP128, CYP130, CYP140, CYP142, CYP144, CYP147, CYP150,	Bacteria

	CYP164, CYP187, CYP188, CYP189, CYP190, CYP191, CYP195, CYP202, CYP268, CYP279, CYP291	
10	CYP143	Bacteria
11	CYP3, CYP5, CYP6, CYP9, CYP28, CYP5136, CYP5139, CYP5141, CYP5150	Animal, Fungi
12	CYP72, CYP709, CYP714	Plant
13	CYP4, CYP325	Animal
14	CYP52, CYP63, CYP86, CYP94, CYP96, CYP584, CYP704	Plant, Fungi
15	CYP1, CYP2, CYP17, CYP21, CYP33, CYP139, CYP152	Animal, Bacteria

2.3.9. Functional conservation of P450s

To assess the relationship, if any, at the clan-level grouping of P450 families and their functional conservation we assessed each of the P450 family functions at the family level (Table 2F.11). Overall analysis of P450s' function revealed that the majority of the P450s in all biological kingdoms are involved either in generation or oxidation steroids and structurally related molecules, fatty acids and terpenoids (Fig. 2.7 and Tables 2.11 & 2.12). The generation/oxidation of these molecules is critical in the generation of different molecules of biological significance, indicating that P450s are primarily evolved to serve organisms' physiological process. The best example of functional conservation of P450s can be obtained from Clan 1, where both plant P450s (CYP73, CYP75, CYP84, CYP98) and fungal P450s (CYP5037, CYP5144 and CYP5152) are involved in oxidation of different molecules involved either in biosynthesis of lignin (in plants) or oxidation of lignin-derived molecules generated

Table 2.11. Family-level functional analysis of P450s. P450s known for their major role (majority of member P450s function) is shown in the table.

Clan	CYP family	Broader biological/functional role	Biological kingdoms	References
1	CYP71	Synthesis of indole aldoxime (hormones) - terpenoid metabolism	Plant, Fungi	Nafisi *et al.*, 2007)
	CYP73	Biosynthesis of lignin (Cinnamic acid hydroxylation)		Hamberger and Bak, 2013; Teutsch, 1993; Pierrel *et al.*, 1994)
	CYP75	Biosynthesis of lignin (Flavonoid hydroxylation)		Kaltenbach *et al.*, 1999)
	CYP76	Terpenoid metabolism; biosynthesis of secoiridoids and terpene indole alkaloids; herbicide metabolism; 7-ethoxycoumarin O-de-ethylation (hydroxylation of geraniol and its derivatives, monoterpenes and diterpenes and phenylurea)		Höfer *et al.*, 2014; Batard *et al.*, 1998
	CYP78	Fatty acid hydroxylation; growth and gametophore formation		Katsumata *et al.*, 2011; Imaishi *et al.*, 2000; Kai *et al.*, 2009
	CYP79	Synthesis of indole aldoxime (hormones)		Sugawara *et al.*, 2009; Irmisch *et al.*, 2015
	CYP81	Indole glucosinolates hydroxylation		Pfalz *et al.*, 2011
	CYP82	Terpenoic metabolism		Pfalz *et al.*,2011
	CYP84	Biosynthesis of lignin (coniferaldehyde/coniferyl alcohol 5-hydroxylase)		Ehlting *et al.*, 2006; Mizutani and Ohta, 2010
	CYP89	Formation of major chlorophyll catabolites during leaf senescence		Christ *et al.*, 2013
	CYP92	Brassinosteroid synthesis		Kang *et al.*, 2011
	CYP93	Flavonoid biosynthesis		Du *et al.*, 2011
	CYP98	Biosynthesis of lignin (3'-hydroxylation of		Hamberger

		phenolic esters ρ-coumaroyltyramine meta-hydroxylation)		and Bak, 2013
	CYP620	Orphan		
	CYP705	Sterols desaturation and triterpenoid metabolism		(Hamberger and Bak, 2013; Mizutani and Ohta, 2010
	CYP706	Sesquiterpenoid synthesis (nootkatol and nootkatone)		Cankar *et al.*, 2013
	CYP5037	Plant defence chemicals and degradation of compounds derived from plant components (lignin, cellulose or hemicellulose; coumarin)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
	CYP5144	Polycyclic aromatic hydrocarbons and plant defence chemicals and degradation of compounds derived from plant components (lignin, cellulose or hemicellulose)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
	CYP5152	Plant defence chemicals and degradation of compounds derived from plant components (lignin, cellulose or hemicellulose)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
2	CYP7	Steroids hydroxylation (bile acid biosynthesis -cholesterol 7α-hydroxylation)	Animal	Cohen *et al.*, 1992; Guengerich , 2015;
	CYP8	Steroids hydroxylation (Prostacyclin and bile acid biosynthesis)		Guengerich , 2015; Yokoyana *et al.*, 1996
	CYP19	Steroids (aromatase: aromatization of androgens into estrogens)		Guengerich , 2015; Corbin *et al.*, 1988
	CYP39	Steroids hydroxylation (cholesterol hydroxylation; Steroid 7-alpha hydroxylase)		Guengerich , 2015
3	CYP11	Steroids hydroxylation (steroids biosynthesis)	Animal	Guengerich , 2015
	CYP12	Insecticide resistance		Feyereisen, 2006
	CYP24	Steroids hydroxylation (vitamin D)		Guengerich , 2015
	CYP27	Steroids hydroxylation (bile acid biosynthesis and viamin D)		Guengerich , 2015

4	CYP26	Retinoic acid hydroxylation (Vitamin A)	Animal, Fungi, Plant, Bacteria	Guengerich , 2015
	CYP51	Steroid metabolism (biosynthesis of membrane ergosterol)		Kelly and Kelly, 2013
	CYP53	Benzoate and its derivatives hydroxylation; plant material stilbene demethylation		Syed *et al*., 2014;Ide *et al*., 2012; Hirosue *et al*., 2011
	CYP58	Mycotoxin biosynthesis (Aflatoxin and trichothecene)		Ehrlich *et al*., 2004
	CYP61	Steroid metabolism (biosynthesis of membrane ergosterol)		Kelly *et al*., 1997
	CYP65	Mycotoxin biosynthesis (Trichothecene and fumonisin)		(Kimura *et al*., 2007; Proctor *et al*., 2006; Bojja *et al*., 2004)
	CYP87	Triterpoind biosynthesis (Maesasaponins - C-16α oxidations of β-amyrin)		Moses *et al*., 2015
	CYP90	Steroid hydroxylation (Brassinosteroids)		Szekeres,1996
	CYP110	Fatty acid ω-hydroxylase		Torres *et al*., 2005
	CYP116	Xenobiotics degradation		Nagy, 1995; Warman *et al*., 2012
	CYP135	Orphan		
	CYP136	Orphan (up-regulated by hydrocarbons)		
	CYP137	Orphan		
	CYP138	Orphan		
	CYP501	Orphan		
	CYP512	Steroids hydroxylation (progesterone and testosterone and dehydroabietic acid)		Syed *et al*., 2014; Ide *et al*., 2012; Hirosue *et al*., 2011
	CYP707	Hormones hydroxylaiton (Abscisic acid - plants)		Saito *et al*., 2004
	CYP716	Triterpoind biosynthesis (Maesasaponins - oxidation of b-amyrin to oleanolic acid)		Moses *et al*., 2015
	CYP5035	Plant chemicals, i.e. resin and flavonoids and the pharmaceutical chemical naproxen		Syed *et al*., 2014;Ide *et al*., 2012; Hirosue *et al*.,

				2011
5	CYP97	Tetraterpenoids hydroxylaiton (carotenoids)	Plant, Bacteria	Ruiz-Sola and Rodriguez-Concepcion, 2012
	CYP132	Orphan		
	CYP185	Orphan (up-regulated by hydrocarbons)		
6	CYP55	Denitrification	Fungi, Plant, Animal, Bacteria	Tomura *et al*., 1994
	CYP74	Conversions of fatty acid hydroperoxides to bioactive oxylipins		(Stumpe and Feussner, 2006; Grechkin, 2002
	CYP102	Fatty acid hydroxylation		Noble *et al*., 1999
	CYP105	Diterpenoids oxidation (abietic acid, dehydroabietic acid, isopimaric acid); Biosynthesis of drugs-antibiotics		Moody and Loveridge, 2014
	CYP106	Steroids, di- and triterpene hydroxylation (progesterone, 11-deoxycortisol; dehydroepiandrosterone; 11-keto-*β*-boswellic acid)		Kiss *et al*., 2015
	CYP141	Orphan		
	CYP153	Alkane hydroxylation		Maier *et al*., 2001
	CYP505	Fatty acid hydroxylation and mycotoxin biosynthesis (fumonisin)		Kitazume *et al*., 2000
7	CYP46	Steroids hydroxylation (cholesterol)	Animal	Guengerich, 2015; Lund *et al*., 1999
8	CYP157	Orphan (no EXXR)	Bacteria	Rupasinghe *et al*., 2006
9	CYP101	Terpenoid hydroxylaiton (camphor)	Bacteria	(Poulos, 1986)
	CYP107	Macrolide hydroxylation (antibiotics/drugs), fatty acids oxidative cleavage, steroid hydroxylation (vitamin D3)		Cupp–Vickery and Poulos 1995; . Sakaki *et al*., 2011; Cryle and De Voss, 2004
	CYP108	Terpineol oxidation		Fruetel *et al*., 1994
	CYP121	Mycocyclosin production via intramolecular C-C bond formation of cyclo (L-tyr-L-tyr)		McLean *et al*., 2008

		(cYY)-highly specific		
	CYP123	Orphan		
	CYP124	Steroids and fatty acids hydroxylation (Omega hydroxylation of methyl-branched fatty acids and cholesterol/cholest-4-en-3-one oxidation)		Johnston *et al*., 2009 & 2010
	CYP125	Steroids hydroxylation (cholesterol and 4-cholesten-3-one oxidation)		McLean *et al*., 2009
	CYP126	Orphan		
	CYP128	Menaquinone hydroxylaiton		Holsclaw *et al*., 2008; Sogi *et al*., 2016
	CYP130	Orphan		
	CYP140	Macrolide hydroxylation (mycolactone synthesis)		Mve-Obiang *et al*., 2005
	CYP142	Steroid hydroxylation (cholesterol and 4-cholesten-3-one oxidation)		Driscoll *et al*., 2010
	CYP144	Orphan		
	CYP147	Omega-fatty acid hydroxylation		Bhattarai *et al*., 2013
	CYP150	polycyclic aromatic hydrocarbon hydroxylation (pyrene, dibenzothiophene and 7-methylbenz(alpha)anthracene)		Brezna *et al*., 2006
	CYP164	Fatty acid hydroxylation		Agnew *et al*., 2012
	CYP187	Orphan		
	CYP188	Orphan		
	CYP189	Orphan		
	CYP190	Orphan		
	CYP191	Orphan		
	CYP195	Orphan		
	CYP202	Orphan		
	CYP268	Orphan		
	CYP279	Orphan		
	CYP291	Orphan		
10	CYP143	Orphan	Bacteria	
11	CYP3	Drug/Xenobiotic metabolism	Animal, Fungi	Guengerich, 2015
	CYP5	Eicosanoids hydroxylation (Thromboxane A2 synthesis)		Guengerich, 2015
	CYP6	Xenobiotic metabolism (fatty acids, insecticides), plant chemicals (furanocoumarins)		Schuler, 2015
	CYP9	Monoterpene conversion into pheromones, xenobiotics		Schuler, 2015

	CYP28	Orphan		
	CYP5136	Xenobiotics degradation (Polycyclic aromatic and alkylphenols hydroxylation)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
	CYP5139	Xenobioitics degradation (7-ethoxycoumarin; carbazole and phenanthrene)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
	CYP5141	Plant defense chemicals or degradation of compounds derived from plant components (lignin, cellulose or hemicellulose)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
	CYP5150	Xenobiotics degradation (Polycyclic aromatic compounds)		Syed *et al.*, 2014;Ide *et al.*, 2012; Hirosue *et al.*, 2011
12	CYP72	Steroid catabolism (Brassinosteroids)	Plant	Neff *et al.*, 1999
	CYP709	Fatty acid hydroxylation		Kandel *et al.*, 2005
	CYP714	Diterpene hydroxylation (Gibberellins)		Mizutani and Ohta, 2010
13	CYP4	Arachidonic acid and fatty acid hydroxylation; insecticide resistance	Animal	Guengerich, 2015
	CYP325	Insecticide resistance (based on up-regulation)		David *et al.*, 2005; Reddy *et al.*, 2012
14	CYP52	Alkanes and fatty acid hydroxylation	Plant, Fungi	Kim *et al.*, 2007; van Bogaert *et al.*, 2009
	CYP63	Xenobiotics degradation (alkanes, polycyclic aromatic compounds and alkylphenols)		Syed *et al.*, 2013
	CYP86	Fatty acid hydroxylation		Guengerich, 2015
	CYP94	Fatty acid hydroxylation		Guengerich, 2015
	CYP96	Alkane hydroxylation		Greer *et al.*, 2007
	CYP584	Orphan		
	CYP704	Fatty acid hydroxylation		Li *et al.*, 2010

15	CYP1	Drug/xenobiotic metabolism; Steroids hydroxylation	Animal, Bacteria	Guengerich, 2015; McKinnon *et al.*, 2008
	CYP2	Drug/xenobiotic and steroid metabolism; Fatty acids hydroxylation		Guengerich, 2015; McKinnon *et al.*, 2008
	CYP17	Steroid biosynthesis (testosterone and estrogen biosynthesis)		Guengerich, 2015; McKinnon *et al.*, 2008
	CYP21	Steroid biosynthesis		Guengerich, 2015; McKinnon *et al.*, 2008
	CYP33	Orphan		
	CYP139	Orphan		
	CYP152	Fatty acid hydroxylation		Rude *et al.*, 2011

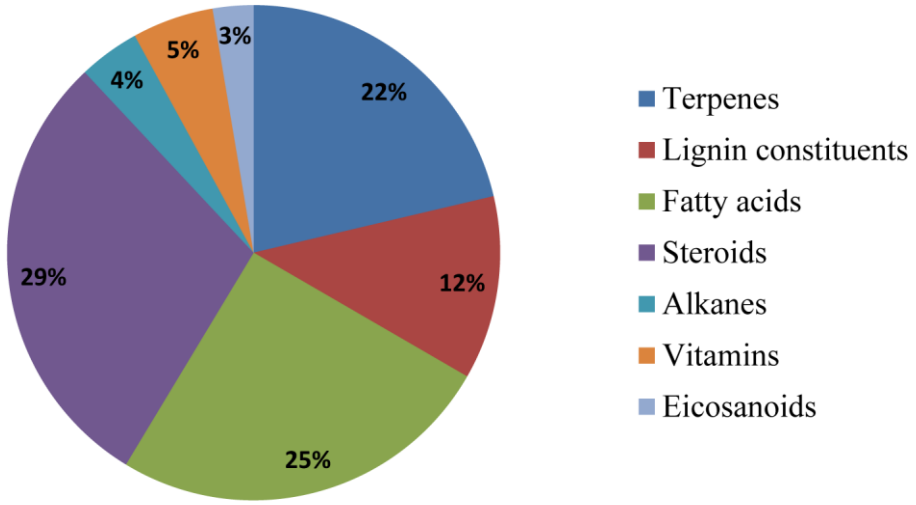

Figure 2.7. Classification of P450s based on their main substrate class. Percentage of P450s involved in oxidation of particular substrate class is shown in the figure. The functional classification is presented in a broader terms of substrates as described elsewhere (Guengerich *et al.*, 2015). Detailed information on classification of P450 family members into different substrate class is presented in Table 2.12.

during fungal-mediated degradation of wood, particularly the lignin component (see Tables 2.11 & 2.12). It is noteworthy that the molecules involved in biosynthesis of lignin were also found during the fungal-mediated degradation of lignin (Vanholme *et al*., 2010; Martinez *et al*., 2005). This indicates some overlap in substrate specificity in plant and fungal P450s grouped in this clan, except that plant P450s' reaction is directed towards lignin biosynthesis whereas that of fungal P450s is directed towards de-lignification. Oxidation of different plant chemicals by fungal P450s (Syed *et al*., 2014; Ide *et al*., 2012; Hirosue *et al*., 2011) that are part of the lignin biosynthesis pathway further strongly supports this hypothesis (Tables 2.11 & 2.12).

Furthermore, plant P450s (CYP86, CYP94, CYP96, CYP704) and fungal P450s (CYP52 and CYP63) grouped under Clan 14 show perfect functional conservation such that these P450s are involved in the oxidation of aliphatic hydrocarbons such as alkanes and fatty acids. Clan 4 harbors P450s from all biological kingdoms that possess oxidation activity towards steroids, terpenoids and other structurally related aromatic compounds (see Table 2.11). The same phenomenon can be found in Clan 6, where fungal (CYP505), bacterial (CYP102 and CYP153) and plant/animal (CYP74) P450s are involved in oxidation of aliphatic hydrocarbons, including fatty acids. Functional conservation can also be found between animal and fungal P450 families grouped under Clan 11 where these P450s perform diverse catalytic reactions, including oxidation of different xenobiotics (carcinogenic and endocrine-disrupting chemicals) and plant-related compounds (Table 2.11). Functional conservation of different P450s belonging to different biological kingdoms grouped under the same clans indicate that these P450s potentially have a common parental ancestor and our analysis of grouping different P450 families into clans makes sense.

Table 2.12. Classification of P450s based on their main substrate class as described elsewhere (Guengerich, 2015). Percent of P450s is calculated considering 75 P450s that are functionally characterized as 100% for each of the substrate class. Considering study is aimed to find common functions of all P450s and many P450s are involved in oxidation of different class of xenobiotics, xenobiotics as substrate class data is not represented. This table is derived from the functional data presented in Table 2.11.

Substrate class	P450s	Number of P450s	Percent of P450s
Terpenes	CYP714,CYP9,CYP108,CYP101,CYP106,CYP105,CYP97,CYP716,CYP87,CYP706,CYP705,CYP82,CYP81,CYP79,CYP76,CYP71	16	21
Lignin constituents	CYP5141,CYP5139,CYP5152,CYP5144,CYP5037,CYP98,CYP84,CYP76,CYP73	9	12
Fatty acids	CYP152,CYP2,CYP704,CYP94,CYP86,CYP63,CYP52,CYP4,CYP709,CYP6,CYP164,CYP147,CYP124,CYP107,CYP505,CYP102,CYP74,CYP110,CYP78	19	25
Steroids	CYP121,CYP17,CYP2,CYP72,CYP142,CYP125,CYP124,CYP46,CYP106,CYP512,CYP90, CYP61,CYP51,CYP27,CYP11,CYP39,CYP19,CYP8,CYP7,CYP705,CYP92,CYP76	22	29
Alkanes	CYP96,CYP52,CYP153	3	4
Vitamins	CYP107,CYP26,CYP27,CYP24	4	5
Eicosanoids	CYP4, CYP5	2	3

Animal P450s that are grouped into clans such as 2, 3, 7 and 13 are involved in steroid hydroxylation. However, based on functional data, a functional divergence was observed for CYP4, CYP12 and CYP325 P450s, as these P450s are known for their role in insecticide resistance. CYP55, grouped under clan 6 that performs denitrification, shows functional

divergence compared to other members that are involved in oxidation of aliphatic hydrocarbons (see Table 2.11). Functional divergence of P450s in terms of their function can be observed for P450s grouped across clans, as they involved in different xenobiotics oxidation. Clan 8 harbors a single P450 CYP157 that is known to lack the EXXR motif. One of the P450 CYP121 belonging to bacteria that are exclusively confined to the MTBC complex has been found to be highly specific towards its known substrate (McLean *et al*., 2015 & 2008),

Based on the above data we conclude that functional conservation of P450s is quite common, considering P450s' prime evolution is to serve organisms through involvement in different biological reactions of physiological importance and during evolution, in response to constant pressure, these P450s acquired/evolved capabilities such as enhancing their substrate specificity and performing different catalytic reactions. Functional conservation of P450s from different biological kingdoms grouped into the same clan further strengthens the hypothesis of common ancestral origin of these P450s. The results presented in this article enhance our understanding of the molecular evolutionary analysis of P450s in terms of their dynamic nature (both at protein and gene level) across biological kingdoms.

2.4. REFERENCES

- Agnew, C.R., Warrilow, A.G., Burton, N.M., Lamb, D.C., Kelly, S.L. and Brady, R.L., 2012. An enlarged, adaptable active site in CYP164 family P450 enzymes, the sole P450 in Mycobacterium leprae. *Antimicrobial agents and chemotherapy*, *56*(1), pp.391-402.
- Batard, Y., LeRet, M., Schalk, M., Robineau, T., Durst, F. and Werck-Reichhart, D., 1998. Molecular cloning and functional expression in yeast of CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin O-de-ethylase from Helianthus tuberosus. *The Plant Journal*, *14*(1), pp.111-120.
- Bhattarai, S., Liou, K. and Oh, T.J., 2013. Hydroxylation of long chain fatty acids by CYP147F1, a new cytochrome P450 subfamily protein from Streptomyces peucetius. *Archives of biochemistry and biophysics*, *539*(1), pp.63-69.
- Bingham, J. and Sudarsanam, S., 2000. Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics*, *16*(7), pp.660-661.
- Bojja, R.S., Cerny, R., Proctor, R.H. and Du, L., 2004. Determining the biosynthetic sequence in the early steps of the fumonisin pathway by use of three gene-disruption mutants of Fusarium verticillioides. *Ronald Cerny Publications*, p.6.
- Brezna, B., Kweon, O., Stingley, R.L., Freeman, J.P., Khan, A.A., Polek, B., Jones, R.C. and Cerniglia, C.E., 2006. Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading Mycobacterium vanbaalenii PYR-1. *Applied microbiology and biotechnology*, *71*(4), p.522.
- Cankar, K., van Houwelingen, A., Goedbloed, M., Renirie, R., de Jong, R.M., Bouwmeester, H., Bosch, D., Sonke, T. and Beekwilder, J., 2014. Valencene oxidase

CYP706M1 from Alaska cedar (Callitropsis nootkatensis). *FEBS letters*, *588*(6), pp.1001-1007.

- Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W.R., Eltis, L.D. and Mohn, W.W., 2009. Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. *Journal of Biological Chemistry*, *284*(51), pp.35534-35542.
- Chen, W., Lee, M.K., Jefcoate, C., Kim, S.C., Chen, F. and Yu, J.H., 2014. Fungal cytochrome p450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution*, *6*(7), pp.1620-1634.
- Chen, W., Lee, M.K., Jefcoate, C., Kim, S.C., Chen, F. and Yu, J.H., 2014. Fungal cytochrome p450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution*, *6*(7), pp.1620-1634.
- Christ, B., Süssenbacher, I., Moser, S., Bichsel, N., Egert, A., Müller, T., Kräutler, B. and Hörtensteiner, S., 2013. Cytochrome P450 CYP89A9 is involved in the formation of major chlorophyll catabolites during leaf senescence in Arabidopsis. *The Plant Cell*, *25*(5), pp.1868-1880.
- Cohen, J.C., Cali, J.J., Jelinek, D.F., Mehrabian, M., Sparkes, R.S., Lusis, A.J., Russell, D.W. and Hobbs, H.H., 1992. Cloning of the human cholesterol 7α-hydroxylase gene (CYP7) and localization to chromosome 8q11–q12. *Genomics*, *14*(1), pp.153-161.
- Corbin, C.J., Graham-Lorence, S., McPhaul, M., Mason, J.I., Mendelson, C.R. and Simpson, E.R., 1988. Isolation of a full-length cDNA insert encoding human aromatase system cytochrome P-450 and its expression in nonsteroidogenic cells. *Proceedings of the National Academy of Sciences*, *85*(23), pp.8948-8952.

- Cryle, M.J. and De Voss, J.J., 2004. Carbon–carbon bond cleavage by cytochrome P450 BioI (CYP107H1). *Chemical communications*, (1), pp.86-87.
- Cupp-Vickery, J.R. and Poulos, T.L., 1995. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Structural & Molecular Biology*, *2*(2), pp.144-153.
- Darban-Sarokhalil, D., Fooladi, A., Bameri, Z., Nasiri, M. and Feizabadi, M., 2011. Cytochrome CYP141: a new target for direct detection of Mycobacterium tuberculosis from clinical specimens. *Acta microbiologica et immunologica Hungarica*, *58*(3), pp.211-217.
- David, J.P., Strode, C., Vontas, J., Nikou, D., Vaughan, A., Pignatelli, P.M., Louis, C., Hemingway, J. and Ranson, H., 2005. The Anopheles gambiae detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(11), pp.4080-4084.
- Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E., Leys, D. and Munro, A.W., 2010. Structural and biochemical characterization of Mycobacterium tuberculosis CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *Journal of Biological Chemistry*, *285*(49), pp.38270-38282.
- Du, Y., Chu, H., Chu, I.K. and Lo, C., 2010. CYP93G2 is a flavanone 2-hydroxylase required for C-glycosyl-flavone biosynthesis in rice. *Plant physiology*, pp.pp-110.
- Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Comput Biol*, *7*(10), p.e1002195.

- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, *5*(1), p.113.
- Ehlting, J., Hamberger, B., Million-Rousseau, R. and Werck-Reichhart, D., 2006. Cytochromes P450 in phenolic metabolism. *Phytochemistry Reviews*, *5*(2-3), pp.239-270.
- Ehrlich, K.C., Chang, P.K., Yu, J. and Cotty, P.J., 2004. Aflatoxin biosynthesis cluster gene cypA is required for G aflatoxin formation. *Applied and Environmental Microbiology*, *70*(11), pp.6518-6524.
- Feyereisen, R., 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*(1), pp.19-28.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J. and Sonnhammer, E.L., 2014. Pfam: the protein families database. *Nucleic acids research*, *42*(D1), pp.D222-D230.
- Frank, D.J., Madrona, Y. and de Montellano, P.R.O., 2014. Cholesterol ester oxidation by mycobacterial cytochrome P450. *Journal of Biological Chemistry*, *289*(44), pp.30417-30425.
- Fruetel, J.A., Mackman, R.L., Peterson, J.A. and de Montellano, P.O., 1994. Relationship of active site topology to substrate specificity for cytochrome P450terp (CYP108). *Journal of Biological Chemistry*, *269*(46), pp.28815-28821.
- Funhoff, E.G., Bauer, U., García-Rubio, I., Witholt, B. and van Beilen, J.B., 2006. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. *Journal of bacteriology*, *188*(14), pp.5220-5227.

- Galagan, J.E., Sisk, P., Stolte, C., Weiner, B., Koehrsen, M., Wymore, F., Reddy, T.B.K., Zucker, J.D., Engels, R., Gellesch, M. and Hubble, J., 2010. TB database 2010: overview and update. *Tuberculosis*, *90*(4), pp.225-235.
- Girhard, M., Bakkes, P.J., Mahmoud, O. and Urlacher, V.B., 2015. P450 Biotechnology. In *Cytochrome P450* (pp. 451-520). Springer International Publishing.
- Grechkin, A.N., 2002. Hydroperoxide lyase and divinyl ether synthase. *Prostaglandins & other lipid mediators*, *68*, pp.457-470.
- Greer, S., Wen, M., Bird, D., Wu, X., Samuels, L., Kunst, L. and Jetter, R., 2007. The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of Arabidopsis. *Plant Physiology*, *145*(3), pp.653-667.
- Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F. and Smirnova, T., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*(D1), pp.D699-D704.
- Guillén, F., Martínez, M.J., Gutiérrez, A. and Del Rio, J., 2005. Biodegradation of lignocellu-losics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *International Microbiology*, *8*, pp.195-204.
- Hamberger, B. & Bak, S. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Phil. Trans. R. Soc. B. Biol. Sci*. **368,** 20120426 (2013).
- Hamberger, B. and Bak, S., 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1612), p.20120426.

- Hirosue, S., Tazaki, M., Hiratsuka, N., Yanai, S., Kabumoto, H., Shinkyo, R., Arisawa, A., Sakaki, T., Tsunekawa, H., Johdo, O. and Ichinose, H., 2011. Insight into functional diversity of cytochrome P450 in the white-rot basidiomycete Phanerochaete chrysosporium: involvement of versatile monooxygenase. *Biochemical and biophysical research communications*, *407*(1), pp.118-123.

- Hirosue, S., Tazaki, M., Hiratsuka, N., Yanai, S., Kabumoto, H., Shinkyo, R., Arisawa, A., Sakaki, T., Tsunekawa, H., Johdo, O. and Ichinose, H., 2011. Insight into functional diversity of cytochrome P450 in the white-rot basidiomycete Phanerochaete chrysosporium: involvement of versatile monooxygenase. *Biochemical and biophysical research communications*, *407*(1), pp.118-123.

- Höfer, R., Boachon, B., Renault, H., Gavira, C., Miesch, L., Iglesias, J., Ginglinger, J.F., Allouche, L., Miesch, M., Grec, S. and Larbat, R., 2014. Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant physiology*, *166*(3), pp.1149-1161.
- Holsclaw Cynthia, M., Sogi, K.M., Gilmore, S.A., Schelle, M.W., Leavell, M.D., Bertozzi, C.R. and Leary, J.A., 2008. Structural characterization of a novel sulfated menaquinone produced by stf3 from Mycobacterium tuberculosis. *ACS chemical biology*, *3*(10), p.619.
- Ide, M., Ichinose, H. and Wariishi, H., 2012. Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete Postia placenta. *Archives of microbiology*, *194*(4), pp.243-253.

- Ide, M., Ichinose, H. and Wariishi, H., 2012. Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete Postia placenta. *Archives of microbiology*, *194*(4), pp.243-253.
- Imaishi, H., Matsuo, S., Eri, S.W.A.I. and Ohkawa, H., 2000. CYP78A1 preferentially expressed in developing inflorescences of Zea mays encoded a cytochrome P450-dependent lauric acid 12-monooxygenase. *Bioscience, biotechnology, and biochemistry*, *64*(8), pp.1696-1701.
- Irmisch, S., Zeltner, P., Handrick, V., Gershenzon, J. and Köllner, T.G., 2015. The maize cytochrome P450 CYP79A61 produces phenylacetaldoxime and indole-3-acetaldoxime in heterologous systems and might contribute to plant defense and auxin formation. *BMC plant biology*, *15*(1), p.128.
- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and Syed, K., 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. *PloS one*, *9*(9), p.e107209.
- Johnston, J.B., Kells, P.M., Podust, L.M. and de Montellano, P.R.O., 2009. Biochemical and structural characterization of CYP124: A methyl-branched lipid ω-hydroxylase from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, *106*(49), pp.20687-20692.

- Johnston, J.B., Kells, P.M., Podust, L.M. and de Montellano, P.R.O., 2009. Biochemical and structural characterization of CYP124: A methyl-branched lipid ω-hydroxylase from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, *106*(49), pp.20687-20692.

- Johnston, J.B., Ouellet, H. and de Montellano, P.R.O., 2010. Functional redundancy of steroid C26-monooxygenase activity in Mycobacterium tuberculosis revealed by biochemical and genetic analyses. *Journal of Biological Chemistry*, *285*(47), pp.36352-36360.

- Johnston, J.B., Ouellet, H. and de Montellano, P.R.O., 2010. Functional redundancy of steroid C26-monooxygenase activity in Mycobacterium tuberculosis revealed by biochemical and genetic analyses. *Journal of Biological Chemistry*, *285*(47), pp.36352-36360.

- Kai, K., Hashidzume, H., Yoshimura, K., Suzuki, H., Sakurai, N., Shibata, D. and Ohta, D., 2009. Metabolomics for the characterization of cytochromes P450-dependent fatty acid hydroxylation reactions in Arabidopsis. *Plant Biotechnology*, *26*(1), pp.175-182.
- Kaltenbach, M., Schröder, G., Schmelzer, E., Lutz, V. and Schröder, J., 1999. Flavonoid hydroxylase from Catharanthus roseus: cDNA, heterologous expression, enzyme properties and cell-type specific expression in plants. *The Plant Journal*, *19*(2), pp.183-193.
- Kandel, S., Morant, M., Benveniste, I., Blée, E., Werck-Reichhart, D. and Pinot, F., 2005. Cloning, Functional Expression, and Characterization of CYP709C1, the First Sub-terminal Hydroxylase of Long Chain Fatty Acid in Plants INDUCTION BY CHEMICALS AND METHYL JASMONATE. *Journal of Biological Chemistry*, *280*(43), pp.35881-35889.
- Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp.27-30.

- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M., 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, p.gkv1070.

- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M., 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, p.gkv1070.

- Kang, J.G., Yun, J., Kim, D.H., Chung, K.S., Fujioka, S., Kim, J.I., Dae, H.W., Yoshida, S., Takatsuto, S., Song, P.S. and Park, C.M., 2001. Light and brassinosteroid signals are integrated via a dark-induced small G protein in etiolated seedling growth. *Cell*, *105*(5), pp.625-636.
- Katsumata, T., Fukazawa, J., Magome, H., Jikumaru, Y., Kamiya, Y., Natsume, M., Kawaide, H. and Yamaguchi, S., 2011. Involvement of the CYP78A subfamily of cytochrome P450 monooxygenases in protonema growth and gametophore formation in the moss Physcomitrella patens. *Bioscience, biotechnology, and biochemistry*, *75*(2), pp.331-336.
- Kawashima, A. and Satta, Y., 2014. Substrate-dependent evolution of cytochrome P450: Rapid turnover of the detoxification-type and conservation of the biosynthesis-type. *PloS one*, *9*(6), p.e100059.
- Kelly, S.L. and Kelly, D.E., 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us. *Phil. Trans. R. Soc. B*, *368*(1612), p.20120476.
- Kelly, S.L. and Kelly, D.E., 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?. *Phil. Trans. R. Soc. B*, *368*(1612), p.20120476.

- Kelly, S.L., Lamb, D.C., Baldwin, B.C., Corran, A.J. and Kelly, D.E., 1997. Characterization of Saccharomyces cerevisiae CYP61, sterol Δ22-desaturase, and inhibition by azole antifungal agents. *Journal of Biological Chemistry*, *272*(15), pp.9986-9988.
- Kgosiemang, I.K.R., Mashele, S.S. and Syed, K., 2014. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum Saccharomycotina. *J. Pure Appl. Microbiol*, *8*, pp.291-302.
- Kim, D., Cryle, M.J., De Voss, J.J. and de Montellano, P.R.O., 2007. Functional expression and characterization of cytochrome P450 52A21 from Candida albicans. *Archives of biochemistry and biophysics*, *464*(2), pp.213-220.
- Kimura, M., Tokai, T., Takahashi-Ando, N., Ohsato, S. and Fujimura, M., 2007. Molecular and genetic studies of Fusarium trichothecene biosynthesis: pathways, genes, and evolution. *Bioscience, biotechnology, and biochemistry*, *71*(9), pp.2105-2123.
- Kiss, F.M., Schmitz, D., Zapp, J., Dier, T.K., Volmer, D.A. and Bernhardt, R., 2015. Comparison of CYP106A1 and CYP106A2 from Bacillus megaterium–identification of a novel 11-oxidase activity. *Applied microbiology and biotechnology*, *99*(20), pp.8495-8514.
- Kitazume, T., Takaya, N., Nakayama, N. and Shoun, H., 2000. Fusarium oxysporum fatty-acid subterminal hydroxylase (CYP505) is a membrane-bound eukaryotic counterpart of Bacillus megaterium cytochrome P450BM3. *Journal of Biological Chemistry*, *275*(50), pp.39734-39740.
- Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, p.msw054.

- Lee, S., Badieyan, S., Bevan, D.R., Herde, M., Gatz, C. and Tholl, D., 2010. Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proceedings of the National Academy of Sciences*, *107*(49), pp.21205-21210.
- Lepesheva, G.I. and Waterman, M.R., 2007. Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1770*(3), pp.467-477.
- Letunic, I. and Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), pp.127-128.
- Li, H. and Poulos, T.L., 1997. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural & Molecular Biology*, *4*(2), pp.140-146.
- Li, H., Pinot, F., Sauveplane, V., Werck-Reichhart, D., Diehl, P., Schreiber, L., Franke, R., Zhang, P., Chen, L., Gao, Y. and Liang, W., 2010. Cytochrome P450 family member CYP704B2 catalyzes the ω-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. *The Plant Cell*, *22*(1), pp.173-190.
- Lio, P. and Goldman, N., 1998. Models of molecular evolution and phylogeny. *Genome research*, *8*(12), pp.1233-1244.

- Lio, P. and Goldman, N., 1998. Models of molecular evolution and phylogeny. *Genome research*, *8*(12), pp.1233-1244.

- Lund, E.G., Guileyardo, J.M. and Russell, D.W., 1999. cDNA cloning of cholesterol 24-hydroxylase, a mediator of cholesterol homeostasis in the brain. *Proceedings of the National Academy of Sciences*, *96*(13), pp.7238-7243.

- Maier, T., Förster, H.H., Asperger, O. and Hahn, U., 2001. Molecular characterization of the 56-kDa CYP153 from Acinetobacter sp. EB104. *Biochemical and biophysical research communications*, *286*(3), pp.652-658.
- McKinnon, R.A., Sorich, M.J. and Ward, M.B., 2008. Cytochrome P450 part 1: multiplicity and function. *Journal of Pharmacy Practice and Research*, *38*(1), pp.55-57.
- McLean, K. J., Leys, D. & Munro, A. W. Microbial cytochrome P450s in *Cytochrome P450: Structure, mechanism, and biochemistry* 4th edn. (eds Oritz de Montellano, P. R.) Ch. 6, 261-407. (Springer International Publishing, 2015).
- McLean, K.J., Carroll, P., Lewis, D.G., Dunford, A.J., Seward, H.E., Neeli, R., Cheesman, M.R., Marsollier, L., Douglas, P., Smith, W.E. and Rosenkrands, I., 2008. Characterization of active site structure in CYP121 a cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *Journal of Biological Chemistry*, *283*(48), pp.33406-33416.
- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and Munro, A.W., 2009. The structure of Mycobacterium tuberculosis CYP125 molecular basis for cholesterol binding in a P450 needed for host infection. *Journal of Biological Chemistry*, *284*(51), pp.35524-35533.
- Mizutani, M. and Ohta, D., 2010. Diversification of P450 genes during land plant evolution. *Annual review of plant biology*, *61*, pp.291-315.
- Moktali, V., Park, J., Fedorova-Abrams, N.D., Park, B., Choi, J., Lee, Y.H. and Kang, S., 2012. Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC genomics*, *13*(1), p.525.

- Moody, S.C. and Loveridge, E.J., 2014. CYP105—diverse structures, functions and roles in an intriguing family of enzymes in Streptomyces. *Journal of applied microbiology*, *117*(6), pp.1549-1563.
- Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D. and Goossens, A., 2015. Unraveling the triterpenoid saponin biosynthesis of the African shrub Maesa lanceolata. *Molecular plant*, *8*(1), pp.122-135.
- Mve-Obiang, A., Lee, R.E., Umstot, E.S., Trott, K.A., Grammer, T.C., Parker, J.M., Ranger, B.S., Grainger, R., Mahrous, E.A. and Small, P.L.C., 2005. A newly discovered mycobacterial pathogen isolated from laboratory colonies of Xenopus species with lethal infections produces a novel form of mycolactone, the Mycobacterium ulcerans macrolide toxin. *Infection and immunity*, *73*(6), pp.3307-3312.
- Nafisi, M., Goregaoker, S., Botanga, C.J., Glawischnig, E., Olsen, C.E., Halkier, B.A. and Glazebrook, J., 2007. Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *The Plant Cell*, *19*(6), pp.2039-2052.
- Nagy, I., Schoofs, G., Compernolle, F., Proost, P., Vanderleyden, J. and De Mot, R., 1995. Degradation of the thiocarbamate herbicide EPTC (S-ethyl dipropylcarbamothioate) and biosafening by Rhodococcus sp. strain NI86/21 involve an inducible cytochrome P-450 system and aldehyde dehydrogenase. *Journal of bacteriology*, *177*(3), pp.676-687.
- Neff, M.M., Nguyen, S.M., Malancharuvil, E.J., Fujioka, S., Noguchi, T., Seto, H., Tsubuki, M., Honda, T., Takatsuto, S., Yoshida, S. and Chory, J., 1999. BAS1: A gene

regulating brassinosteroid levels and light responsiveness in Arabidopsis. *Proceedings of the National Academy of Sciences*, *96*(26), pp.15316-15323.

- Nelson, D.R. and Strobel, H.W., 1987. Evolution of cytochrome P-450 proteins. *Molecular biology and evolution*, *4*(6), pp.572-593.
- Nelson, D.R., 1998. Metazoan cytochrome P450 evolution. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, *121*(1), pp.15-22.
- Nelson, D.R., 1999. Cytochrome P450 and the individuality of species. *Archives of biochemistry and biophysics*, *369*(1), pp.1-10.
- Nelson, D.R., 2009. The cytochrome p450 homepage. *Human genomics*, *4*(1), p.59.
- Nelson, D.R., 2011. Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*(1), pp.14-18.
- Noble, M.A., Miles, C.S., Reid, G.A., Chapman, S.K. and Munro, A.W., 1999. Catalytic properties of key active site mutants of flavocytochrome P-450 BM3. *Biochemical Society Transactions*, *27*(1), pp.A44-A44.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2010. The Mycobacterium tuberculosis cytochrome P450 system. *Archives of biochemistry and biophysics*, *493*(1), pp.82-95.
- Ozbakir, H.F., Sambade, D.A., Majumdar, S., Linday, L.A. and Banta, S., 2015. Detectio of25-Hydroxyvitamin D3 with an Enzyme modified Electrode. *J Biosens Bioelectron*, *7*(193), p.2.
- Pei, J., Kim, B.H. and Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research*, *36*(7), pp.2295-2300.

- Pfalz, M., Mikkelsen, M.D., Bednarek, P., Olsen, C.E., Halkier, B.A. and Kroymann, J., 2011. Metabolic engineering in Nicotiana benthamiana reveals key enzyme functions in Arabidopsis indole glucosinolate modification. *The Plant Cell*, *23*(2), pp.716-729.
- Pierrel, M.A., Batard, Y., Kazmaier, M., Mignotte-Vieux, C., Durst, F. and Werck-Reichhart, D., 1994. Catalytic properties of the plant cytochrome P450 CYP73 expressed in yeast. *European journal of biochemistry*, *224*(3), pp.835-844.
- Poulos, T. L. & Johnson, E. F. Structures of cytochrome P450 enzymes in *Cytochrome P450: Structure, mechanism, and biochemistry* 4th edn. (eds Oritz de Montellano, P. R.) Ch. 1, 3-32. (Springer International Publishing, 2015.
- Price, M.N., Dehal, P.S. and Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, *26*(7), pp.1641-1650.
- Proctor, R.H., Plattner, R.D., Desjardins, A.E., Busman, M. and Butchko, R.A., 2006. Fumonisin production in the maize pathogen Fusarium verticillioides: genetic basis of naturally occurring chemical variation. *Journal of Agricultural and Food Chemistry*, *54*(6), pp.2424-2430.
- Qhanya, L.B., Matowane, G., Chen, W., Sun, Y., Letsimo, E.M., Parvez, M., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete biotrophic plant pathogens. *PloS one*, *10*(11), p.e0142100.
- Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., Young, S., Wajnberg, E., Fricaux, T., Taquet, N., Blomquist, G.J. and Feyereisen, R., 2012. An insect-specific P450

oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proceedings of the National Academy of Sciences*, *109*(37), pp.14858-14863.

- Reddy, B.N., Rao, B.P., Prasad, G.B.K.S. and Raghavendra, K., 2012. Identification and classification of detoxification enzymes from Culex quinquefasciatus (Diptera: Culicidae). *Bioinformation*, *8*(9), p.430.
- Reddy, T.B.K., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H. and Koehrsen, M., 2009. TB database: an integrated platform for tuberculosis research. *Nucleic acids research*, *37*(suppl 1), pp.D499-D508.
- Rienksma, R.A., Suarez-Diez, M., Mollenkopf, H.J., Dolganov, G.M., Dorhoi, A., Schoolnik, G.K., dos Santos, V.A.M., Kaufmann, S.H., Schaap, P.J. and Gengenbacher, M., 2015. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. *BMC genomics*, *16*(1), p.34.
- Rude, M.A., Baron, T.S., Brubaker, S., Alibhai, M., Del Cardayre, S.B. and Schirmer, A., 2011. Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from Jeotgalicoccus species. *Applied and environmental microbiology*, *77*(5), pp.1718-1727.
- Ruiz-Sola, M.Á. and Rodríguez-Concepción, M., 2012. Carotenoid biosynthesis in Arabidopsis: a colorful pathway. *The Arabidopsis Book*, p.e0158.
- Rupasinghe, S., Schuler, M.A., Kagawa, N., Yuan, H., Lei, L., Zhao, B., Kelly, S.L., Waterman, M.R. and Lamb, D.C., 2006. The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. *FEBS letters*, *580*(27), pp.6338-6342.

- Saito, S., Hirai, N., Matsumoto, C., Ohigashi, H., Ohta, D., Sakata, K. and Mizutani, M., 2004. Arabidopsis CYP707As encode (+)-abscisic acid 8′-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. *Plant Physiology*, *134*(4), pp.1439-1449.
- Sakaki, T., Sugimoto, H., Hayashi, K., Yasuda, K., Munetsuna, E., Kamakura, M., Ikushiro, S. and Shiro, Y., 2011. Bioconversion of vitamin D to its active form by bacterial or mammalian cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*(1), pp.249-256.
- Sassetti, C.M., Boyd, D.H. and Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, *48*(1), pp.77-84.
- Schuler, M.A., 2015. P450s in plants, insects, and their fungal pathogens. In *Cytochrome P450* (pp. 409-449). Springer International Publishing.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific reports*, *5*, p.11572.
- Sezutsu, H., Le Goff, G. and Feyereisen, R., 2013. Origins of P450 diversity. *Phil. Trans. R. Soc. B*, *368*(1612), p.20120428.
- Sirim, D., Wagner, F., Lisitsa, A. and Pleiss, J., 2009. The cytochrome P450 engineering database: Integration of biochemical properties. *BMC biochemistry*, *10*(1), p.27.
- Sogi, K.M., Holsclaw, C.M., Fragiadakis, G.K., Nomura, D.K., Leary, J.A. and Bertozzi, C.R., (2016) Biosynthesis and regulation of Sulfomenaquinone, a metabolite associated with virulence in Mycobacterium tuberculosis. ACS Infect Dis. 2, pp. 800-806.

- Stumpe, M. and Feussner, I., 2006. Formation of oxylipins by CYP74 enzymes. *Phytochemistry Reviews*, *5*(2-3), pp.347-357.
- Sugawara, S., Hishiyama, S., Jikumaru, Y., Hanada, A., Nishimura, T., Koshiba, T., Zhao, Y., Kamiya, Y. and Kasahara, H., 2009. Biochemical analyses of indole-3-acetaldoxime-dependent auxin biosynthesis in Arabidopsis. *Proceedings of the National Academy of Sciences*, *106*(13), pp.5430-5435.
- Syed, K. and Mashele, S.S., 2014. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. *PLoS One*, *9*(4), p.e95616.
- Syed, K., Nelson, D.R., Riley, R. and Yadav, J.S., 2013. Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore species Bjerkandera adusta, Ganoderma sp. and Phlebia brevispora. *Mycologia*, *105*(6), pp.1445-1455.
- Syed, K., Porollo, A., Lam, Y.W., Grimmett, P.E. and Yadav, J.S., 2013. CYP63A2, a catalytically versatile fungal P450 monooxygenase capable of oxidizing higher-molecular-weight polycyclic aromatic hydrocarbons, alkylphenols, and alkanes. *Applied and environmental microbiology*, *79*(8), pp.2692-2702.
- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and Pagadala, N.S., 2014. Genome-wide identification, annotation and characterization of novel thermostable cytochrome P450 monooxygenases from the thermophilic biomass-degrading fungi Thielavia terrestris and Myceliophthora thermophila. *Genes & Genomics*, *36*(3), pp.321-333.

- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One*, *9*(1), p.e86683.
- Szekeres, M., Németh, K., Koncz-Kálmán, Z., Mathur, J., Kauschmann, A., Altmann, T., Rédei, G.P., Nagy, F., Schell, J. and Koncz, C., 1996. Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in Arabidopsis. *Cell*, *85*(2), pp.171-182.
- Tamura, K. and Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, *10*(3), pp.512-526.
- Teutsch, H.G., Hasenfratz, M.P., Lesot, A., Stoltz, C., Garnier, J.M., Jeltsch, J.M., Durst, F. and Werck-Reichhart, D., 1993. Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. *Proceedings of the National Academy of Sciences*, *90*(9), pp.4102-4106.
- Tomura, D., Obika, K., Fukamizu, A. and Shoun, H., 1994. Nitric oxide reductase cytochrome P-450 gene, CYP 55, of the fungus Fusarium oxysporum containing a potential binding-site for FNR, the transcription factor involved in the regulation of anaerobic growth of Escherichia coli. *The Journal of Biochemistry*, *116*(1), pp.88-94.
- Torres, S., Fjetland, C.R. and Lammers, P.J., 2005. Alkane-induced expression, substrate binding profile, and immunolocalization of a cytochrome P450 encoded on the nifD excision element of Anabaena 7120. *BMC microbiology*, *5*(1), p.16.

- Tortoli, E., 2012. Phylogeny of the genus Mycobacterium: many doubts, few certainties. *Infection, Genetics and Evolution*, *12*(4), pp.827-831.
- Van Bogaert, I.N., Demey, M., Develter, D., Soetaert, W. and Vandamme, E.J., 2009. Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast Candida bombicola. *FEMS yeast research*, *9*(1), pp.87-94.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W., 2010. Lignin biosynthesis and structure. *Plant physiology*, *153*(3), pp.895-905.
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. and van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews*, *71*(3), pp.495-548.
- Wagenknecht, H.A. and Woggon, W.D., 1997. Identification of intermediates in the catalytic cycle of chloroperoxidase. *Chemistry & biology*, *4*(5), pp.367-372.
- Warman, A.J., Robinson, J.W., Luciakova, D., Lawrence, A.D., Marshall, K.R., Warren, M.J., Cheesman, M.R., Rigby, S.E., Munro, A.W. and McLean, K.J., 2012. Characterization of Cupriavidus metallidurans CYP116B1–A thiocarbamate herbicide oxygenating P450–phthalate dioxygenase reductase fusion protein. *FEBS Journal*, *279*(9), pp.1675-1693.
- Yokoyama, C., Yabuki, T., Inoue, H., Tone, Y., Hara, S., Hatae, T., Nagata, M., Takahashi, E.I. and Tanabe, T., 1996. Human gene encoding prostacyclin synthase (PTGIS): genomic organization, chromosomal localization, and promoter activity. *Genomics*, *36*(2), pp.296-304.
- Yoshida, Y., Aoyama, Y., Noshiro, M. and Gotoh, O., 2000. Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450

gene superfamily. *Biochemical and biophysical research communications*, *273*(3), pp.799-804.

- Yoshida, Y., Noshiro, M., Aoyama, Y., Kawamoto, T., Horiuchi, T. and Gotoh, O., 1997. Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the most conserved P450 monooxygenase: II. Evolutionary analysis of protein and gene structures. *Journal of biochemistry*, *122*(6), pp.1122-1128.m

CHAPTER 3

IN SILICO STRUCTURAL ANALYSIS OF CYP123A1 OF *MYCOBACTERIUM TUBERCULOSIS* H37RV

3.1. INTRODUCTION

Genome sequencing analysis of the mycobacterial species give researchers an opportunity to look for novel drug targets against these pathogens. It is of great significant to analyse the three-dimensional (3D) structure of a protein/enzyme to develop an inhibitor based on its active site cavity and residues involved in binding of the substrates (Vyas *et al*., 2012). Among different techniques used in understanding the protein structure, 3D modelling, also known as homology modelling, is now widely used (Vyas *et al*., 2012). 3D modelling is gaining momentum because of its simplicity compared to nuclear magnetic resonance (NMR) or X-ray-crystallography. In this technique, a 3D model of a protein is constructed experimentally using a template, usually a homologous protein whose structure has been revealed by crystallography. The 3D model of the protein construct helps to identify the active site structure and its dynamics in terms of interacting with a ligand or substrate, especially in drug discovery and design, a core aspect of pharmaceutical research (Vyas *et al*., 2012).

Mycobacterium tuberculosis genome sequencing revealed the presence of 20 P450s in its genome (Cole *et al*., 1998) and these are considered as novel drug targets against this pathogen (Hudson *et al*., 2012). However, only six out of the 20 P450s have been structurally characterised: CYP51B1 (Bellamine *et al*., 1999), CYP121A1 (Mc Lean *et al*., 2002), CYP124A1 (Johnston *et al*., 2009), CYP125A1 (Mc Lean *et al*., 2009), CYP130A1 (Ouellet *et al*., 2008) and CYP142A1 (Table 3.1) (Driscoll *et al*., 2010). One P450 CYP144A1 was successfully expressed and its binding affinity to different azole drugs was shown (Driscoll *et*

al., 2011). Moreover, CYP144 has been structurally characterised by *in silico*-based 3D homology modelling (Driscoll *et al*., 2010).

Table 3.1. Information on *M. tuberculosis* P450 enzymes' properties.

P450	Crystal structure Available	High affinity for azole	Reference
CYP121	YES	YES	Mc Lean *et al*., 2002
CYP124	YES	YES	Johnston *et al.*, 2009
CYP125	YES	NO	Mc Lean *et al*., 2009
CYP130	YES	YES	Ouellet *et al.*, 2008
CYP142	YES	YES	Driscoll *et al.*, 2010
CYP144	NO	YES	Driscoll *et al.*, 2011
CYP51	YES	YES	Bellamine *et al.*, 1999

Considering that CYP123A1 is a novel drug target for both active and latent *M. tuberculosis*, this study is aimed at understanding the CYP123A1 structure by *in silico*-based 3D homology modelling and also at assessing binding affinity with different azole drugs.

3.2. METHODOLOGY

3.2.1. Homology modelling

The primary sequence of CYP123A1 (Kegg ID: Rv0766c) was blasted at the NCBI using PDB databank option to select a template. BLASTp was used for identifying the homolog protein. The protein 3A4G (cytochrome P450 vdh from *Pseudonocardia autotrophica*) (Yasutake et al., 2010) was selected on the basis of maximum identity of 38% with query cover 96%. The sequence alignment of model CYP123A1 with the template 3A4G was accomplished using ClustalX2 2.0 (http://www.clustal.org/). The 3D structure of CYP123A1 was generated by homology modelling using SWISS-MODEL (https://swissmodel.expasy.org/). Both the template 3A4G and the modelled structure of CYP123A1 were aligned using SPDBV 4.1.0 with carbon-alpha (CA) backbone with RMS value 1.81 Å. The 3D structure of CYP123A1 obtained was visualised by Maestro (https://www.schrodinger.com/maestro). The generated 3D structure of model CYP123A1 was submitted to PDBsum database **(**http://www.ebi.ac.uk/thornton-srv/databases/pdbsum**).** Thus, the quality of the model CYP123A1 was analysed by PROCHECK, which uses the Ramachandran map.

3.2.2. Binding site analysis

As no data is available on the binding site of CYP123A1, the binding site pocket was obtained by using FlexX software. The binding cavity includes amino acids, namely Leu87, Ala239, Lys246, Asn250, Gln287, Pro314, Val341 and Ser342. These residues were used to generate the grid box, which describes the binding pocket of model CYP123A1 for docking calculation. The grid covered grid map dimensions 46Å×46Å×46Å with grid spacing of 0.375Å.

3.2.3. Ligand database

The eight ligands were retrieved from PubChem and the ligand database was built using Avogadro for conversion of SDF files to PDB file. The ligands were clotrimazole (PubChem CID: 2812), econazole (PubChem CID: 3198), fluconazole (PubChem CID: 3365), itraconazole (PubChem CID: 55283), ketoconazole (PubChem CID: 47576), miconazole (PubChem CID: 4189), posaconazole (PubChem CID: 468595) and voriconazole (PubChem CID: 71616). These were further prepared for docking in AutoDock tools 1.5.6.

3.2.4. Molecular docking

The modelled protein CYP123A1 was prepared for docking in AutoDock tools 1.5.6. Molecular docking of model protein CYP123A1 with the respective ligands was carried out using AutoDock Vina. In AutoDock Vina the ligands are docked in the binding site cavity. All structures were docked successfully and the results were visualized in Maestro version 11.0.014 ((https://www.schrodinger.com/maestro)).

3.3. RESULTS AND DISCUSSION

3.3.1. Sequence alignment

The sequence alignment of CYP123A1 and the template 3A4G shows that leucine, arginine, aspartic acid, alanine, glutamic acid and proline are the best conserved amino acid residues among the residues (Fig. 3.1).

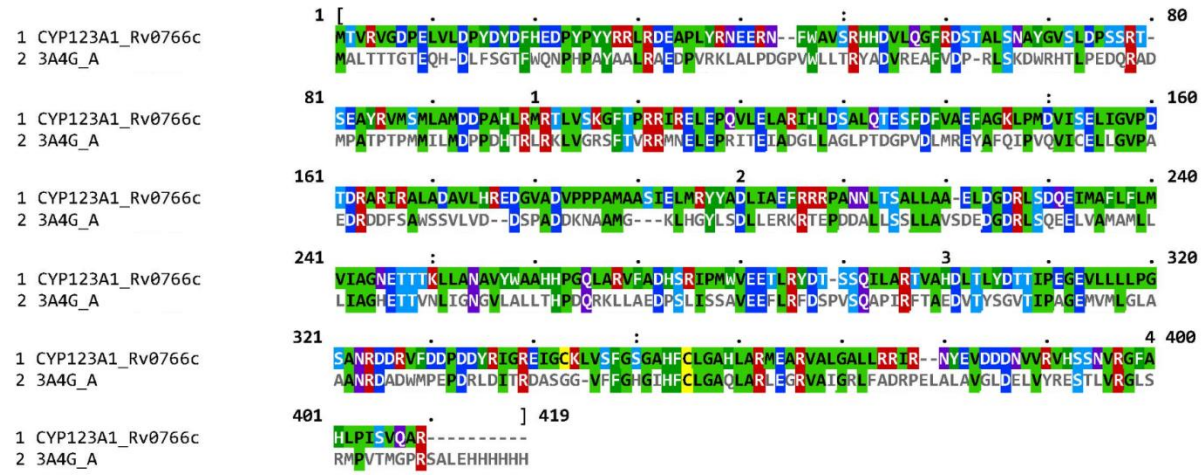

Figure 3.1. Sequence alignment of protein CYP123A1 with template 3A4G.

3.3.2. Construction of CYP123A1 model

At present no crystallised structures of CYP123A1 are available. Therefore, to analyse ligand-protein interactions, a model of CYP123A1 protein has been constructed (Figure 3.2). The template 3A4G was obtained as a result of protein blast. Homology modelling was achieved using SWISS-MODEL and its validation was carried out.

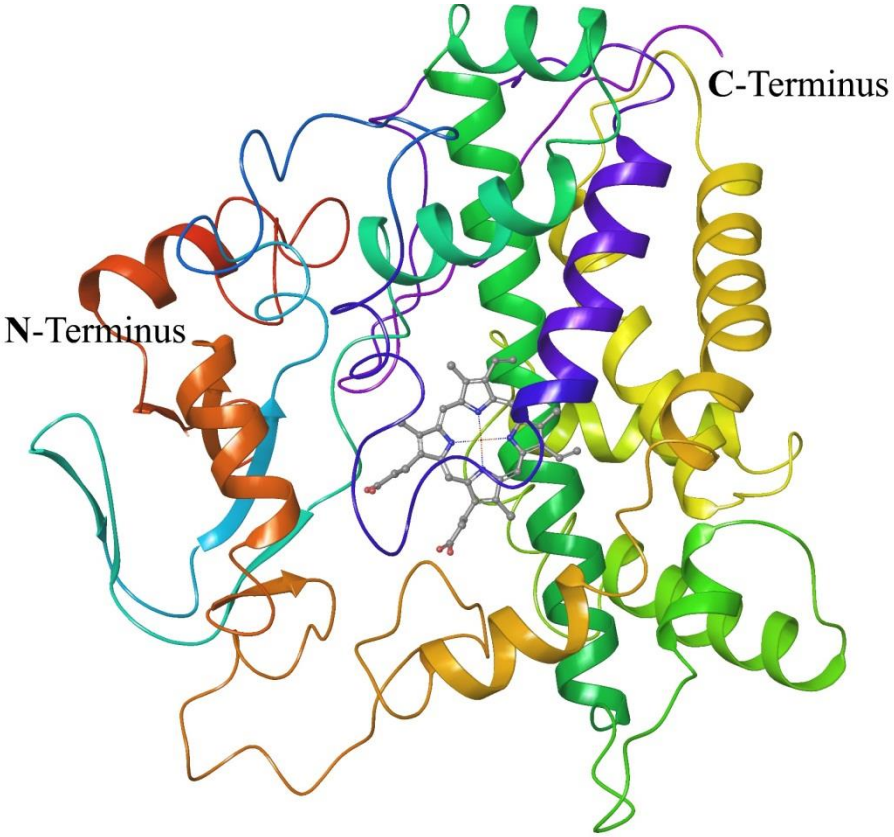

Figure 3.2. 3D model of CYP123A1 constructed using template 3A4G

3.3.3. Validation of homology model

The model of CYP123A1 generated by homology modelling was validated by PDBsum and the Ramachandran plot (Figure 3.3) was obtained. A secondary structure of CYP123A1 was obtained as well (Figure 3.4) and PROMOTIF documentation was prepared, as shown below (Figure 3.5). The root mean square derivative (RMSD) value of the model of CYP123A1 is 1.81 Å, which was obtained in SPDBV with CA backbone.

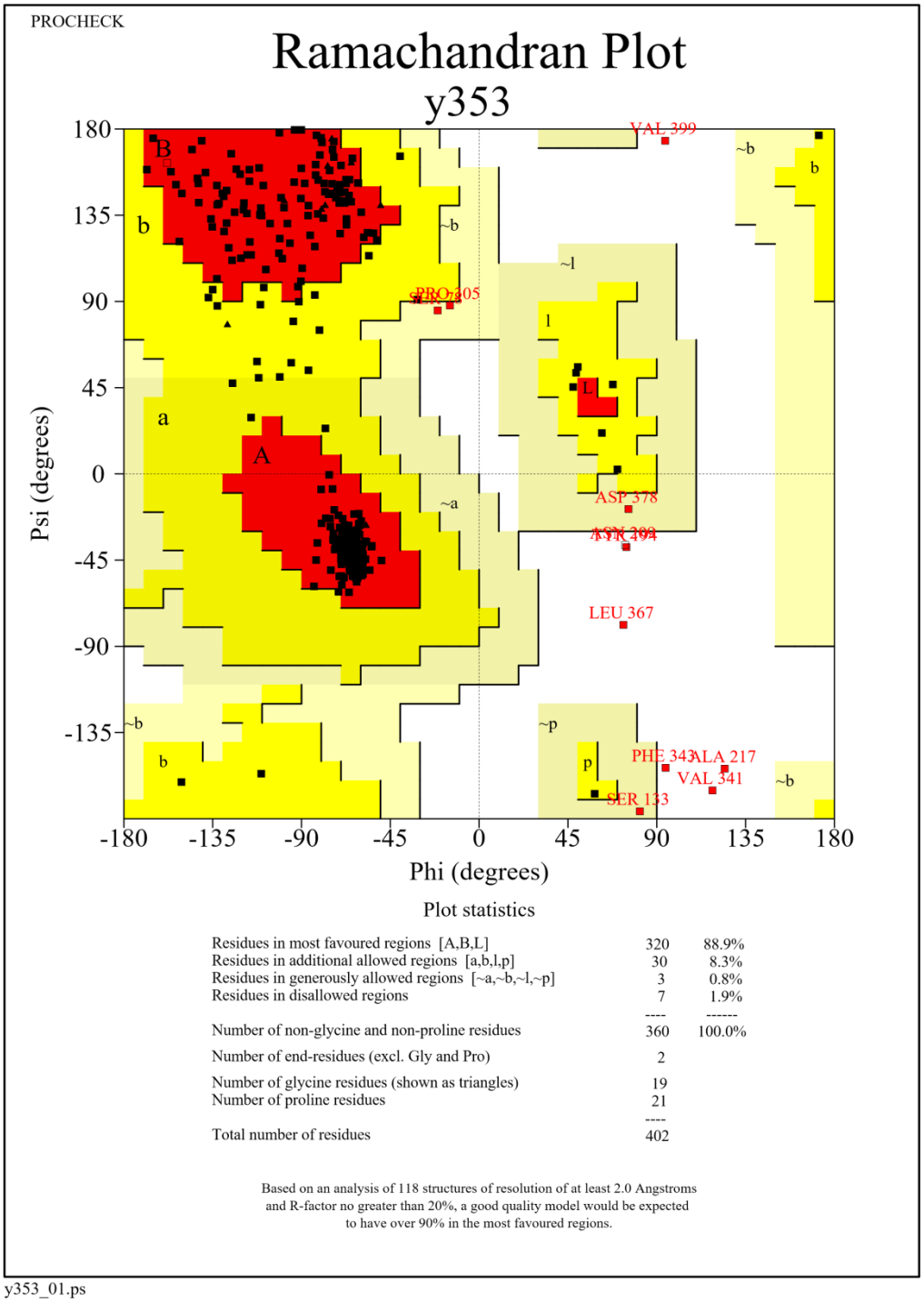

Figure 3.3. Ramachandran plot of CYP123A1 model.

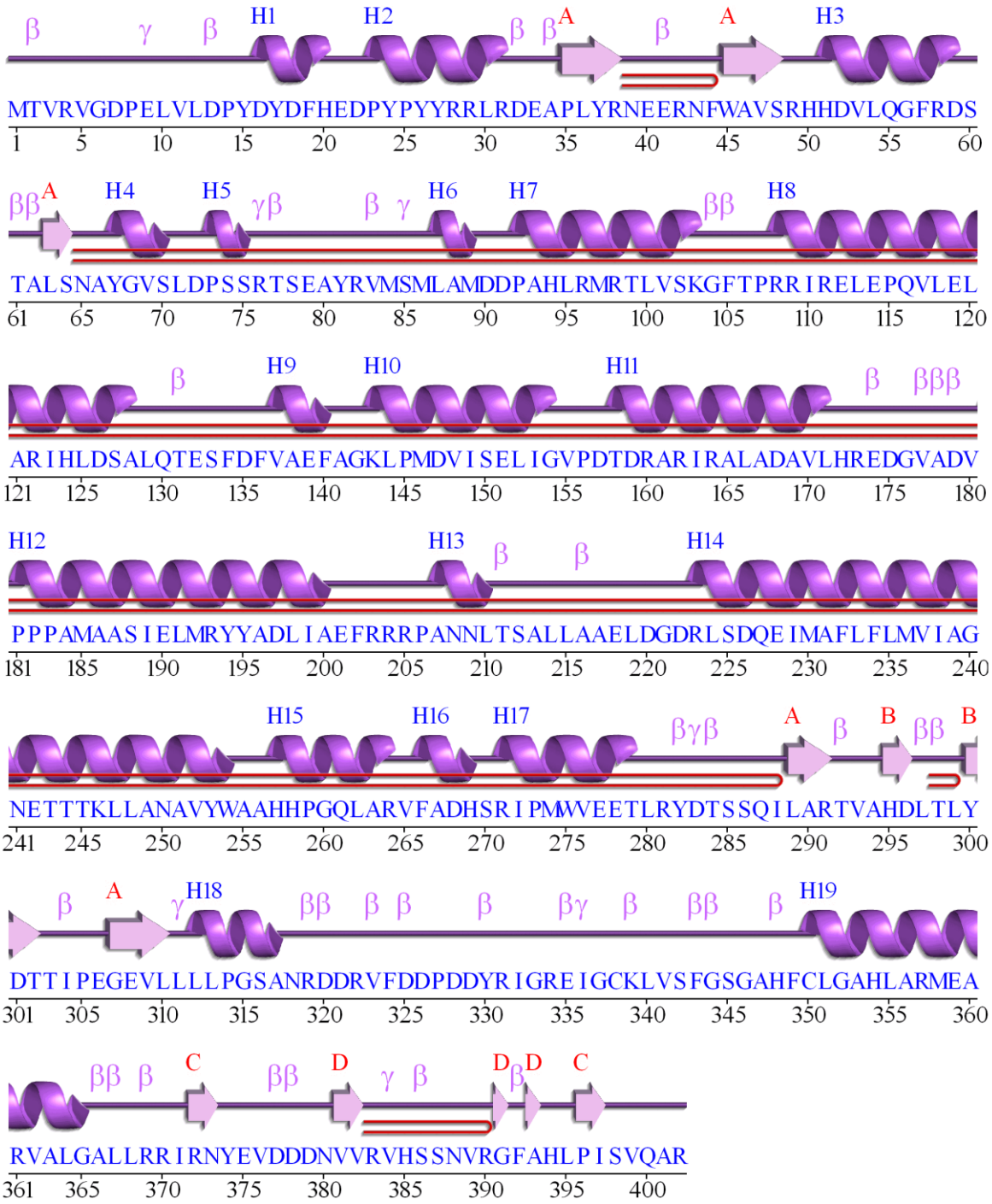

Figure 3.4. Secondary structure of the model CYP123A1.

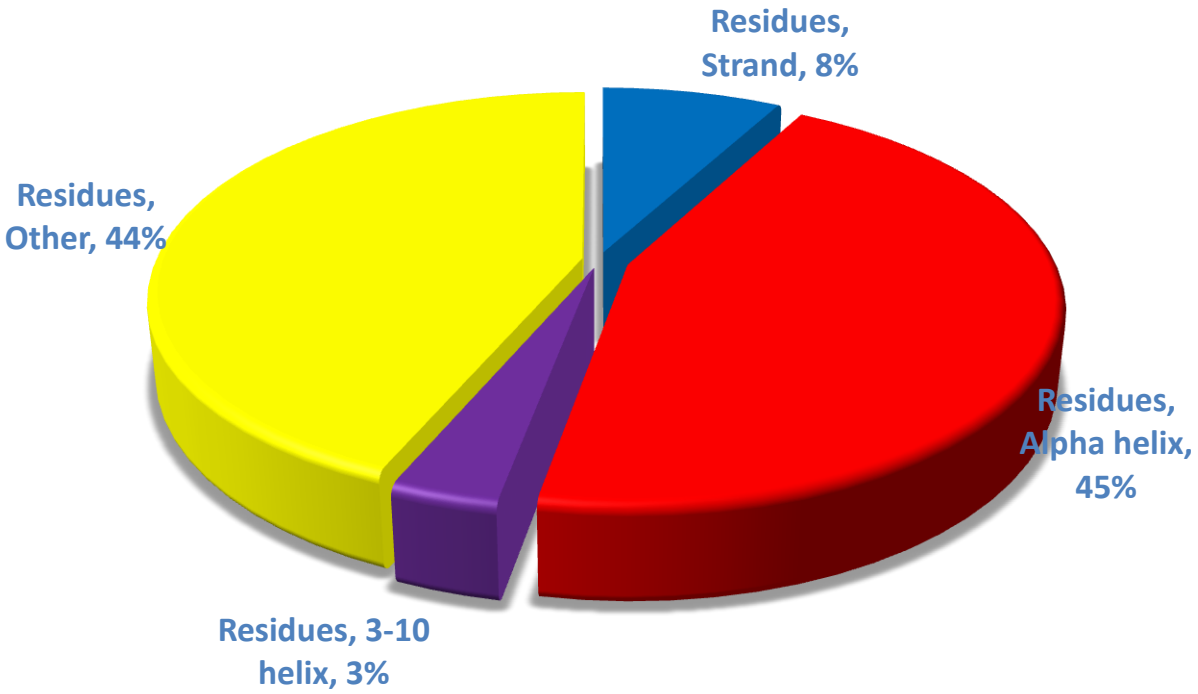

Figure 3.5. Secondary structure summary for CYP123A1 model.

3.3.4. Model-template alignment

Figure 3.6 shows alignment of model CYP123A1 (green colour) with the template 3A4G (magenta colour) with RMSD 1.81 Å, which was obtained using Maestro version 11.0.014 (Schrodinger). The RMSD of 1.81 Å indicates that the 3D model of CYP123A1 is of good quality.

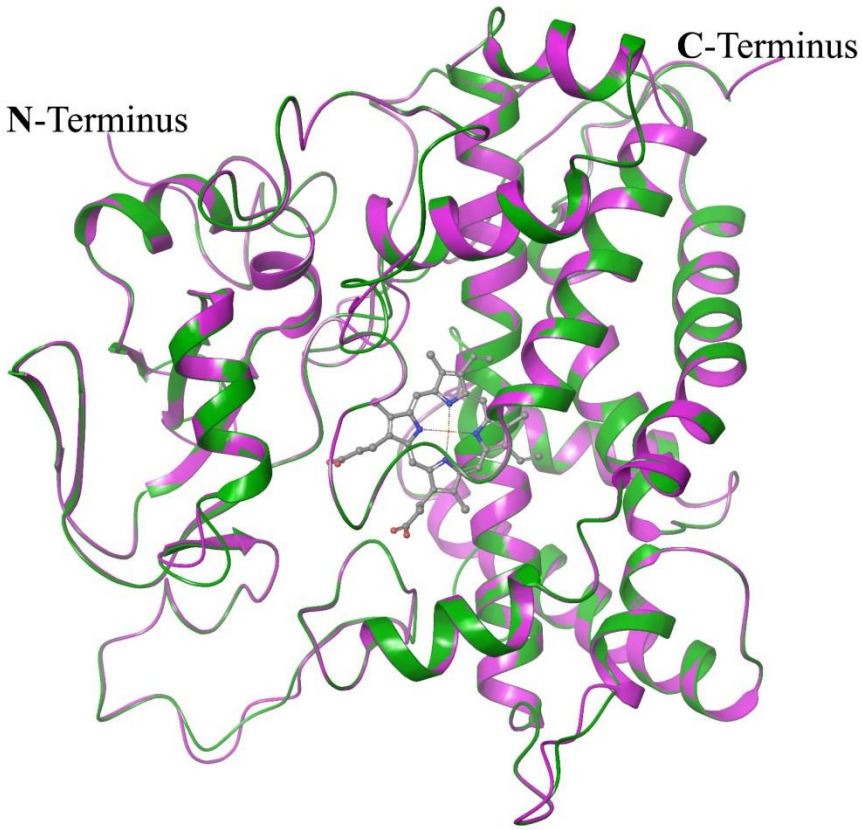

Figure 3.6. Alignment of template (PDB ID: 3A4G) with CYP123A1 model

3.3.5. Molecular docking study

The docking of model CYP123A1 with the ligands clotrimazole, econazole, fluconazole, itraconazole, ketoconazole, miconazole, posaconazole and voriconazole was performed in Autodock vina. The best results and conformation were selected by comparing their lowest binding energy, number of hydrogen bonds and bond distance. The results are shown in Table 3.1. The molecular docking studies showed that ketoconazole forms an excellent complex, better than others, with the best interaction rate and the lowest energy of -9.0 kcal/mol. Ketoconazole forms a tight hydrogen bond with the residue Lys246 CYP123A1 model

(Figure 3.7). Clotrimazole did not show interaction with the 3D model of CYP123A1 (Figure 3.8). Lys246 forms a hydrogen bond with fluconazole; HEM also forms two Pi-Pi stacking with fluconazole (Figure 3.9). Voriconazole makes one Pi-Cation interaction and hydrogen bond with the CYP123A1 residues Arg82 (Figure 3.10).

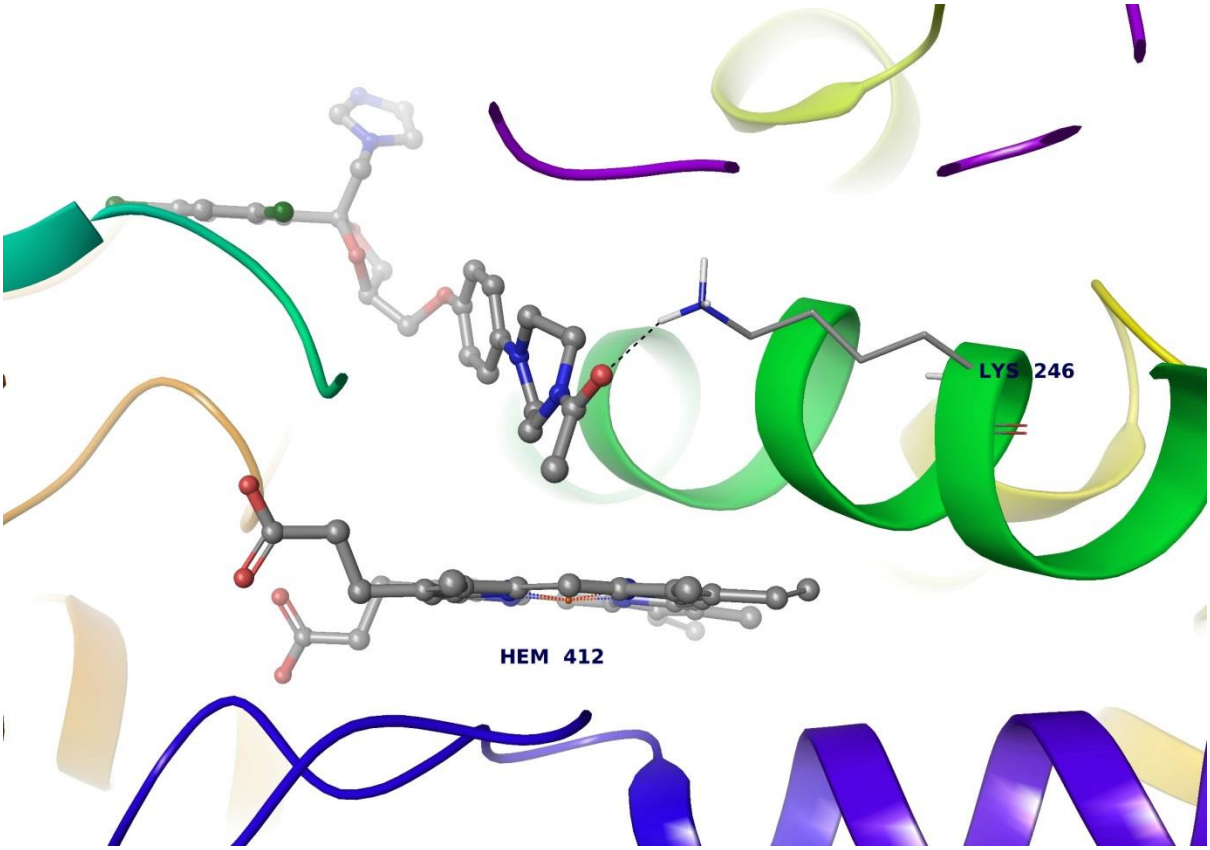

Figure 3.7. Ketoconazole forms a hydrogen bond with residue Lys246 of CYP123A1 model.

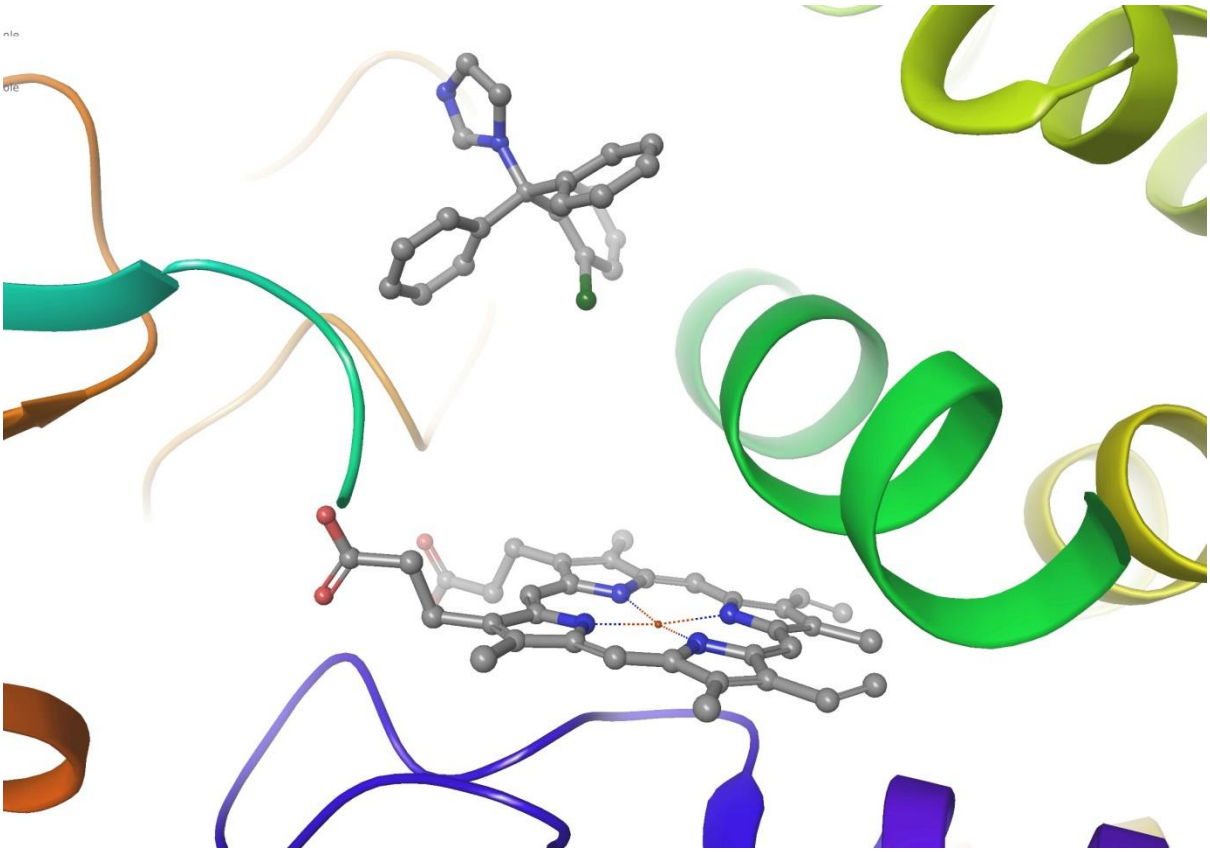

Figure 3.8. Clotrimazole not showing any interaction with CYP123A1 model.

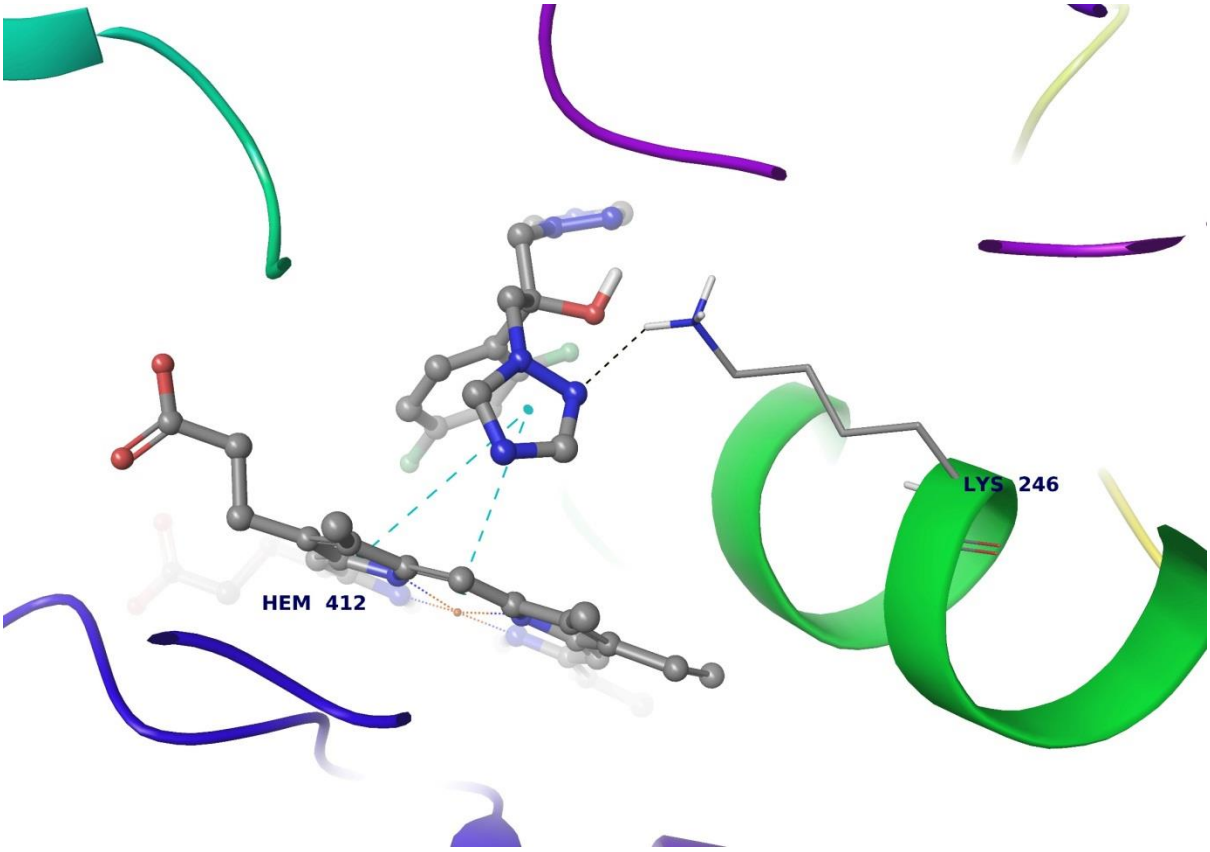

Figure 3.9. Docking analysis of fluconazole with CYP123A1 model

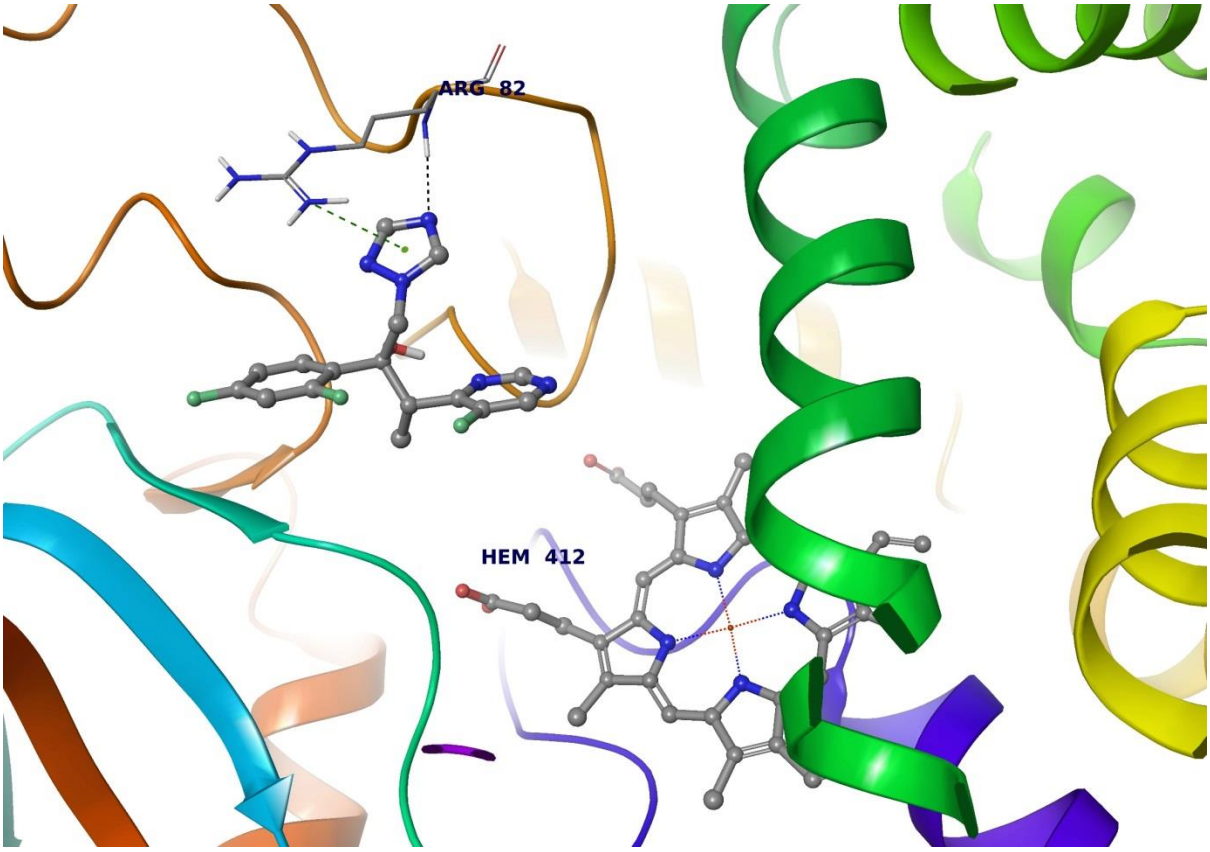

Figure 3.10. Voriconazole makes one Pi-Cation interaction and hydrogen bond with the CYP123A1 residues Arg82.

The sulphur-containing residue Ser386 forms a hydrogen bond with both the miconazole (Figure 3.11) and econazole (Figure 3.12) compounds with a binding affinity of -7.3 kcal/mol and -7.0 kcal/mol respectively. Itraconazole has a binding affinity of -7.6 kcal/mol and forms hydrogen bond interaction with Lys246 (Figure 3.13). The basic amino acid residue His385 forms a hydrogen bond with posaconazole, with binding affinity of -7.1 kcal/mol (Figure 3.14). Figure 3.15 shows that binding interaction of HEM with the

CYP123A1 model. This interaction reveals that only the basic amino acids interact with HEM, namely arginine (98) and histidine (348). Arg291 forms two hydrogen bonds and salt bridge with HEM, Arg98 interacts with HEM forming one hydrogen bond, and His348 also forms one hydrogen bond and makes one salt bridge.

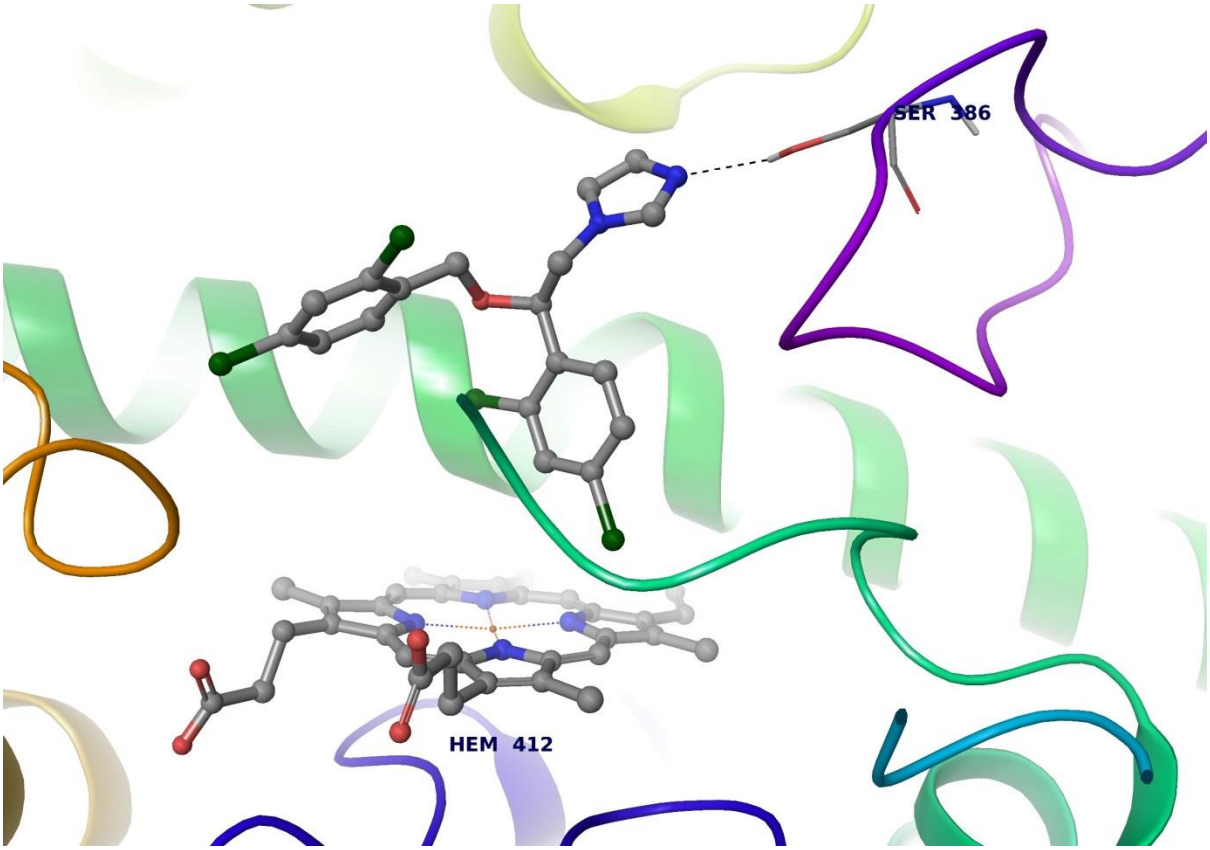

Figure 3.11. Interaction of miconazole with CYP123A1 model.

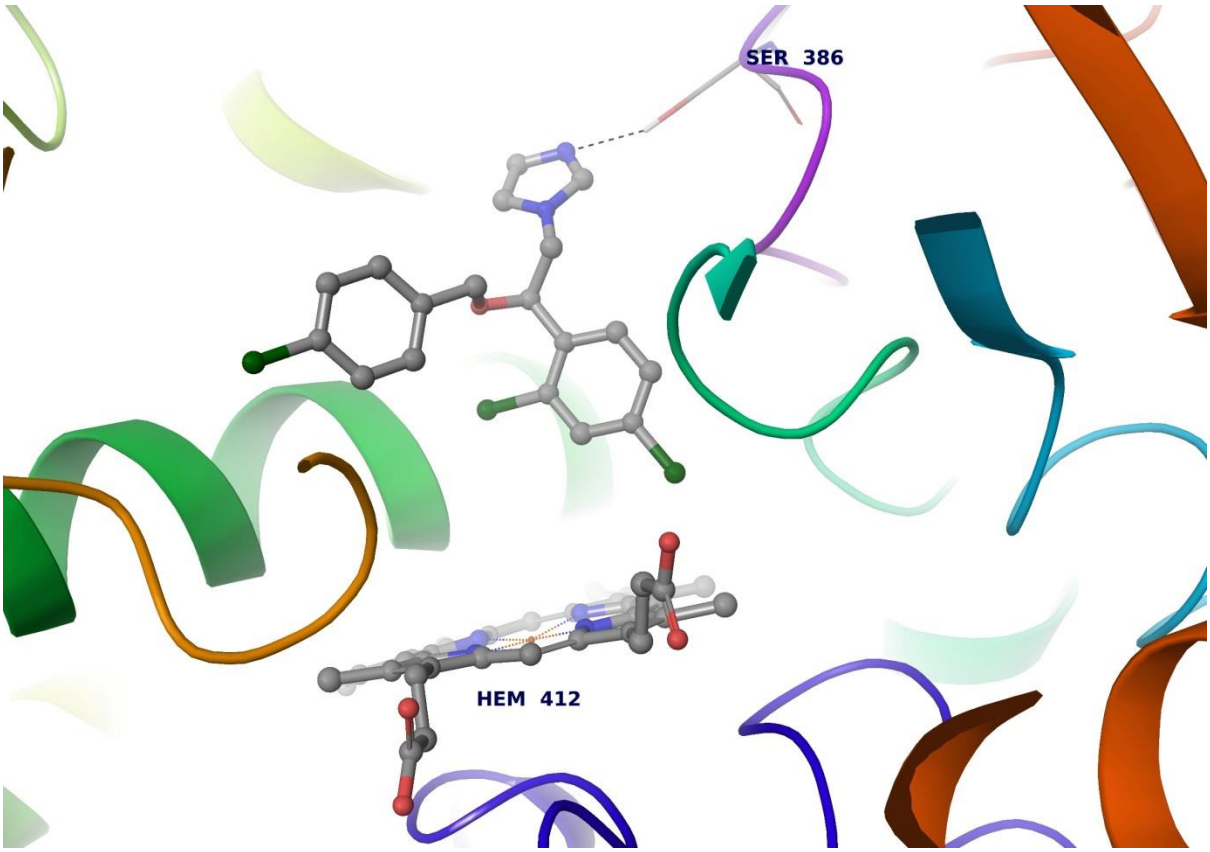

Figure 3.12. Interaction of econazole with CYP123A1 model.

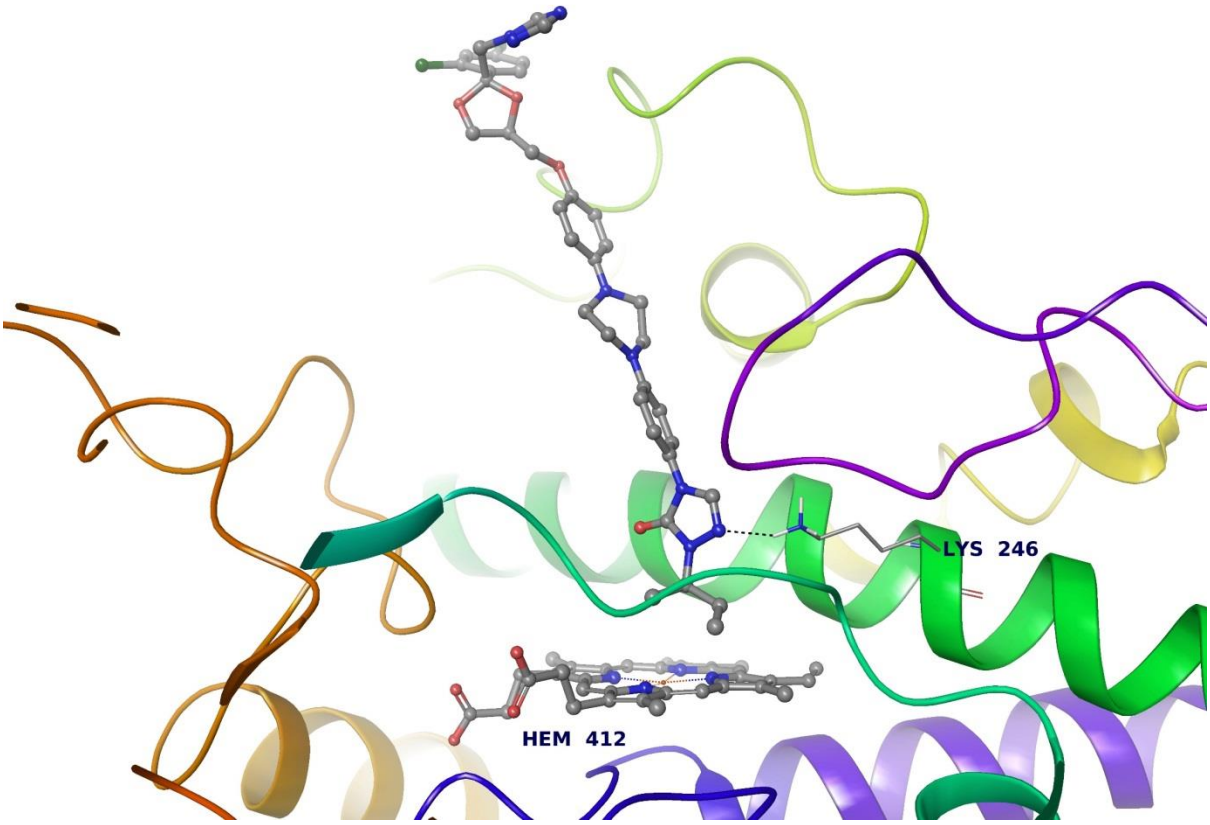

Figure 3.13. Interaction of itraconazole with CYP123A1 model.

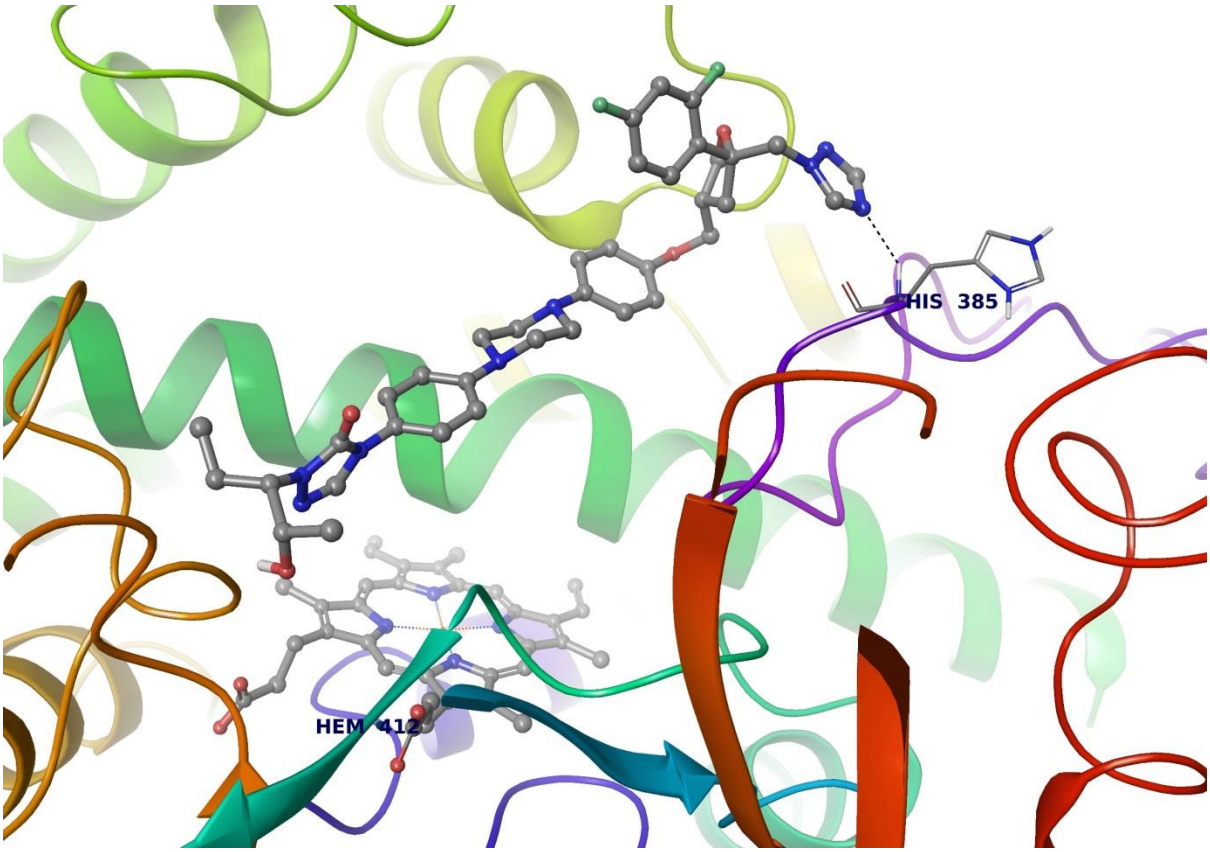

Figure 3.14. Interaction of posoconazole with CYP123A1 model.

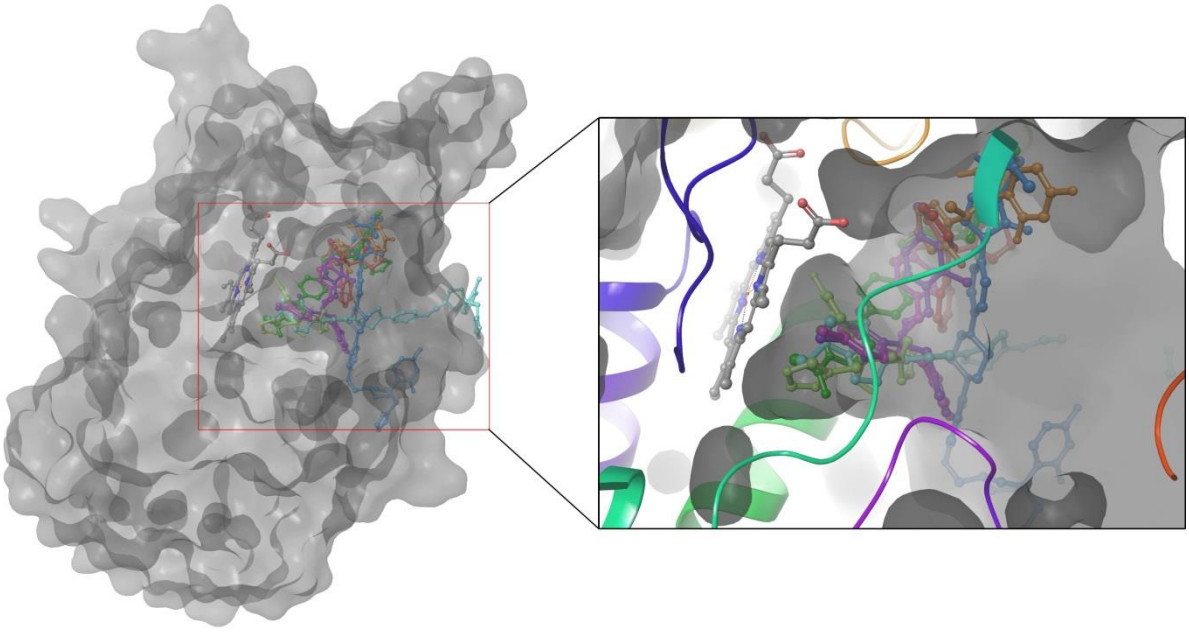

Figure 3.15. Binding cavity with HEM and azole derivatives

Table. 3.2. Binding affinity analysis of different azole drugs with CYP123A1 model.

P450	PDB ID	Binding affinity							
		Clotrimazole	Fluconazole	Voriconazole	Miconazole	Econazole	Ketoconazole	Itraconazole	Posaconazole
CYP123A1_Model	3A4G	-6.0	-7.6	-7.3	-7.0	-8.4	-9.0	-7.6	-7.1
Interacting Residues	Ligand	No	Lys246 (1HB), Hem (2 Pi-Pi)	Arg82 (1HB) & (1Pi-Cation)	Ser386 (1HB)	Ser386 (1HB)	Lys246 (1HB)	Lys246 (1HB)	His385 (1HB)
Interacting Residues	Hem	Arg291 (2HB) and one Salt Bridge, Arg98 (1HB), His348 (1HB) and one Salt Bridge							

3.4. CONCLUSION

In conclusion, in this study, a 3D model of CYP123A1 is constructed and its binding affinity with different azole drugs is analysed. The study revealed that ketoconazole forms an excellent complex, better than other azole drugs, and clotrimazole did not show interaction with the 3D model of CYP123A1. Results generated in this study are based on *in silico* analysis and these results should be validated using *in vitro* binding studies with different azole drugs using purified CYP123A1 protein.

3.5. REFERENCES

- Bellamine, A., Mangla, A.T., Nes, W.D. and Waterman, M.R., 1999. Characterization and catalytic properties of the sterol 14α-demethylase from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *96*(16), pp. 8937-8942.
- Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., II III, C.B. and Tekaia, F., 1998. Erratum: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *396*(6707), p. 190.
- Driscoll, M.D., McLean, K.J., Cheesman, M.R., Jowitt, T.A., Howard, M., Carroll, P., Parish, T. and Munro, A.W., 2011. Expression and characterization of Mycobacterium tuberculosis CYP144: common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. *Biochimica et biophysica acta (BBA) - Proteins and proteomics*, *1814*(1), pp. 76-87.
- Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E., Leys, D. and Munro, A.W., 2010. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *Journal of biological chemistry*, *285*(49), pp. 38270-38282.
- Hudson, S.A., McLean, K.J., Munro, A.W. and Abell, C., 2012. *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets. *Biochem Soc Trans*, *40*(3), pp. 573-9.
- Johnston, J.B., Kells, P.M., Podust, L.M. and de Montellano, P.R.O., 2009. Biochemical and structural characterization of CYP124: a methyl-branched lipid ω-hydroxylase from

Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, *106*(49), pp. 20687-20692.

- McLean, K.J., Cheesman, M.R., Rivers, S.L., Richmond, A., Leys, D., Chapman, S.K., Reid, G.A., Price, N.C., Kelly, S.M., Clarkson, J. and Smith, W.E., 2002. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. *Journal of inorganic biochemistry*, *91*(4), pp. 527-541.
- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and Munro, A.W., 2009. The structure of *Mycobacterium tuberculosis* CYP125 molecular basis for cholesterol binding in a P450 needed for host infection. *Journal of biological chemistry*, *284*(51), pp. 35524-35533.
- Ouellet, H., Podust, L.M. and de Montellano, P.R.O., 2008. *Mycobacterium tuberculosis* CYP130 crystal structure, biophysical characterization, and interactions with antifungal azole drugs. *Journal of biological chemistry*, *283*(8), pp. 5069-5080.
- Vyas, V.K., Ukawala, R.D., Ghate, M. and Chintha, C., 2012. Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, *74*(1), p. 1.
- Yasutake, Y., Fujii, Y., Nishioka, T., Cheon, W.K., Arisawa, A. and Tamura, T., 2010. Structural evidence for enhancement of sequential vitamin D3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D3 hydroxylase. *Journal of Biological Chemistry*, *285*(41), pp.31193-31201.

CHAPTER 4

CLONING AND GENERATION OF RECOMBINANT *ESCHERICHIA COLI* CELLS CONTAINING *CYP123A1* IN pINK-d EXPRESSION VECTOR

4.1. INTRODUCTION

Genome-sequencing analysis of *Mycobacterium tuberculosis* revealed the presence of 20 cytochrome P450 monooxygenases (P450s/CYPs) in its genome (Cole *et al*., 1998). Among 20 P450s, five *M. tuberculosis* P450 enzymes have been structurally and functionally characterised, namely CYP51B1 (Bellamine *et al*., 1999), CYP121A1 (McLean *et al*., 2002), CYP124A1 (Johnston *et al*., 2009), CYP125A1 (McLean *et al*., 2009; Johnston *et al*., 2010), and CYP142A1 (Driscoll *et al*., 2010). One P450 CYP130A1 has only been structurally characterised (Ouellet *et al*., 2008), one P450 CYP1441 has been successfully expressed and its binding affinity to different azole drugs has been shown (Driscoll *et al*., 2011). The remaining P450s are still orphans with no function and structure.

In order to understand the function/structure of *M. tuberculosis* P450s, they were expressed in *E. coli* using different expression vectors (Table 4.1.)

Table 4.1. Vector systems used for expression of *M. tuberculosis* P450s in *E. coli*.

M. tuberculosis P450	Expression vector	References
CYP121A1	pET11a	McLean *et al*., 2002
CYP124A1	pCW	Johnston *et al*., 2009
CYP125A1	pET15b	McLean *et al*., 2009
CYP130A1	pCW	Ouellet *et al*., 2008

CYP142A1	pET15b	Driscoll *et al*., 2010
CYP51B1	pET17b	Bellamine *et al*., 1999
CYP144A1	pET15b	Driscoll *et al*., 2011

Considering that quite a number of *M. tuberculosis* P450s are orphans and that it is difficult to express these P450s (Driscoll, 2011) in this study, the pINK-A expression vector system that has been proven to be efficient to express proteins was used (Karim *et al*., 1993). In this chapter, a strategy was developed to clone *CYP123A1* in the modified pINK-A vector (named pINK-d) and recombinant cells containing the *CYP123A1* in the pINK-d vector was generated.

4.2. METHODOLOGY

4.2.1. CYP123A1 cDNA sequence

The cDNA sequence of CYP123A1 was obtained from Kyoto Encyclopedia of Genes and Genomes database (Kanehisa *et al*., 2017) using its gene ID *Rv0766c*.

4.2.2. Information on expression vector

The expression vector (pINK-A) was a kind gift from Dr Naheed Kaderbhai, Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, SY23 3DD, United Kingdom. The vector's multiple cloning site (MCS) was modified and used in this study. Since the work is going to be patented and commercial aspects are involved, details on the expression vector are not given.

4.2.3. Restriction enzyme analysis

Restriction enzyme analysis of genes and the expression vector was performed using freely available pDRAW DNA analysis software (http://www.acaclone.com/). Restriction enzymes and their sequences for creating a new multiple cloning site in the expression vector were downloaded from the New England Biolabs (NEB) website (https://www.neb.com/products/restriction-endonucleases).

4.2.4. Primer design

Primers for the cloning of *M. tuberculosis* P450s in the expression vector were designed *in silico* using IDT DNA OligoAnalyzer 3.1 (http://eu.idtdna.com/calc/analyzer). Forward and reverse primer sequences were selected and after incorporating the appropriate restriction enzyme sequence, their melting temperature was measured using OligoAnalyzer. Primers were designed in such a way that forward and reverse primers had matching melting temperatures.

4.2.5. Strains, plasmids, chemicals and kits

E. coli DH5α strain was used in this study. *E. coli* cells were cultured on Luria-Bertani (LB) broth and LB-agar (LB broth supplemented with 10 g/L agar). For selection of recombinant *E. coli* cells LB was supplied with the antibiotic ampicillin. LB-antibiotic plates were prepared by adding ampicillin at 100 µg/ml final concentration. Ampicillin stock solution was prepared by dissolving 100 mg of ampicillin (Catalog No. A6140, Sigma-Aldrich, USA) in 1 ml of DNase and RNase free water (Catalog No. L3152, Sigma-Aldrich, USA). The ampicillin stock solution was stored at -20^{o}C. The pINK-d expression vector was used for cloning of *CYP123A1*. All chemicals used were of high quality and were purchased from Sigma-Aldrich and Merck. The plasmid isolation kit was purchased from Qiagen, USA.

4.2.6. Synthesis and cloning of *CYP123A1*

The *M. tuberculosis* P450 CYP123 cDNA sequence along with the modified vector sequence (pINK-d) and cloning strategy were submitted to GenScript (GenScript USA Inc, USA). GenScript synthesised *CYP123A1* and cloned this into the modified vector and the constructs were received.

4.2.7. Preparation of competent cells and transformation

Recombinant plasmids were propagated using chemically competent *E. coli* DH5α, according to the methods described by Inoue *et al.* (1990). To prepare competent cells, LB broth was inoculated with a glycerol stock of the relevant cells, and incubated at 37°C for approximately 10 hours. This pre-culture was then used to inoculate 250 ml of fresh SOB medium (20 g/L tryptone, 5 g/L yeast extract, 0.584 g/L NaCl, 0.186 g/L KCl, 2.034 g/L MgCl2, 2.464 g/L MgSO4), which was incubated overnight at 18-20°C until OD_{600} ~0.55. The culture was then incubated on ice for 10 minutes, followed by centrifugation at 2500 x g for 10 minutes at 4°C. The cell pellet was resuspended on ice in 80 ml of ice-cold TB buffer (10 mM HEPES, pH 6.7; 15mM CaCl2; 250 mM KCl; 55 mM MnCl2), and incubated on ice for 10 minutes. The resuspension was then centrifuged at 2500 g for 10 minutes at 4°C, followed by resuspension in 20 ml ice-cold TB buffer supplemented with DMSO (7% v/v). This was incubated on ice for 10 minutes, before aliquoting 100 µl per 1.5 ml microcentrifuge tube.

To transform the cells, 1 µl plasmid solution containing *CYP123A1* gene, cloned in expression vector (obtained from GenScript), was added to the 100 µl competent cells aliquot and incubated on ice for 20 minutes. Cells were heat-shocked at 42°C for one minute, followed by cold-shock on ice for two minutes. Thereafter, 250 µl SOC medium (SOB medium supplemented with 20 mM glucose) was added to the cells, which were then

incubated at 37°C for one hour. The solution was then centrifuged at 2500 g for five minutes, after which 150 µl supernatant was removed and the cells re-suspended in the remaining 100 µl. The re-suspended cells were streaked on LB agar plates supplemented with 100 µg/ml ampicillin and incubated at 37°C for 16 hours.

4.2.8. Plasmid isolation and purification

Plasmid isolation and purification of the recombinant cells were carried out using QIAprep Spin Miniprep Kit (Catalog No. 27104, Qiagen, Germany) following the manufacturer's protocol. Plasmid DNA concentration was carried out using the SimpliNano microvolume spectrophotometer (Catalog No. GE29-0617-13, Sigma-Aldrich, USA).

4.2.9. Restriction enzyme analysis of plasmids

The above isolated plasmids from recombinant *E. coli* cells were subjected to restriction enzyme digestion to check the presence of the inserts and the correct size of the cloned CYP123A1 cDNA. All restriction enzymes used in this study were purchased from New England Biolabs, South Africa. Digested DNA fragments were analysed on 1% agarose gels. Visualisation of DNA fragments was carried out using SYBR® Sae DNA gel stain (Catalog No. S33102, Thermo Fisher Scientific, USA). The agarose gels were photographed using the Gel Doc™ EZ System (Bio-Rad, South Africa).

4.3. RESULTS AND DISCUSSION

4.3.1. Modifying the multiple cloning site of the expression vector

The expression vector contained only a few restriction enzymes' recognition sites in its MCS. Therefore, the MCS was re-engineered by incorporating more recognition sites in order to clone *M. tuberculosis* P450s. The vector can also be used in future for cloning other genes. The expression vector sequence was analysed using pDRAW and restriction enzymes that did

not have recognition sites on the expression vector's sequence were noted. Based on the above criteria, a few of those restriction enzymes were selected and incorporated in the MCS of the expression vector. The restriction enzymes selected are listed in Table 4.2. The revised vector is named pINK-d and its map and other features are shown in Figure 4.1.

Table 4.2 List of selected restriction enzymes incorporated in the multiple cloning site of the expression vector. The restriction enzyme and its recognition sequences are shown in the table.

Restriction enzyme	Recognition Sequences (5' to 3')
AbsI	CC'TCGAGG
AflII	C'TTAAG
AgeI	A'CCGGT
AscI	GG'CGCGCC
AvrII	C'CTAGG
BglII	A'GATCT
BsiWI	C'GTACG
BspEI	T'CCGGA
BssHII	G'CGCGC
FseI	GGCCGG'CC
KasI	G'GCGCC
MfeI	C'AATTG
NcoI	C'CATGG
PluTI	GGCGC'C
SacI	GAGCT'C
SbfI	CCTGCA'GG

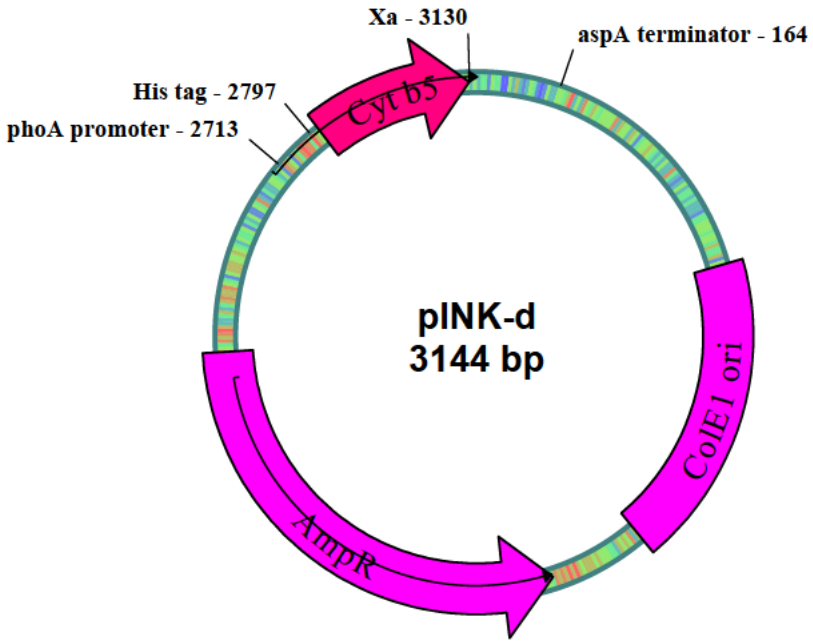

Figure 4.1. Vector map of pINK-d.

4.3.2. Strategy for cloning of *CYP123A1* in expression vector

In order to design a strategy to clone CYP123A1 cDNA in the expression vector, restriction enzyme profiling of the *CYP123A1* was carried out using pDRAW. Table 4.3 shows the restriction profiling of the *CYP123A1* against pINK-d MCS enzymes. The restriction enzymes that do not cut *CYP123A1* and are part of the expression vector's MCS were selected for in-frame cloning of *CYP123A1* (Table 4.3).

Table 4.3. Restriction enzyme profiling of *CYP123A1*. The red-filled boxes indicate that the corresponding restriction enzymes can cleave the gene. Two restriction enzymes with no cleaving ability were selected for *CYP123A1*, for forward and reverse primer design, and are shown as yellow-filled boxes.

Restriction enzyme	CYP123A1
BamHI	
BstXI	
ApaI	
EcoRV	
KpnI	
MIuI	
SphI	
XbaI	

The selected restriction enzymes were made part of the forward and reverse primer of *CYP123A1*, as shown in Table 4.4. Forward and reverse primers for each *M. tuberculosis* P450 were designed as described in the methodology. The primers were designed in such a way that they had almost the same melting temperatures (Table 4.4.).

Table 4.4. Primer sequences and restriction enzymes selected for cloning of *CYP123A1*. The shaded regions represent the restriction enzyme recognition site. The appropriate restriction enzyme in the forward primer (FP) and reverse primer (RP) is also presented in the table. Tm represents melting temperature.

Direction of primer	**Name of the primer**	**Primer sequence**	**Tm**
Forward primer	CYP123A1-FP-KpnI	TATATAGGTACCATGACCGTCCGCGTCGGTGACCCCG	69.3
Reverse primer	CYP123A1-RP-XbaI	GCCTTCTAGATTACCTGGCCTGCACGCTGATCGGC	69.0

In theory, after gene synthesis with the designed forward and reverse primers, the synthesised products will be digested and cloned into the expression vector with the appropriate restriction enzymes. The vector maps with cloned *CYP123A1* are shown in Figure 4.2.

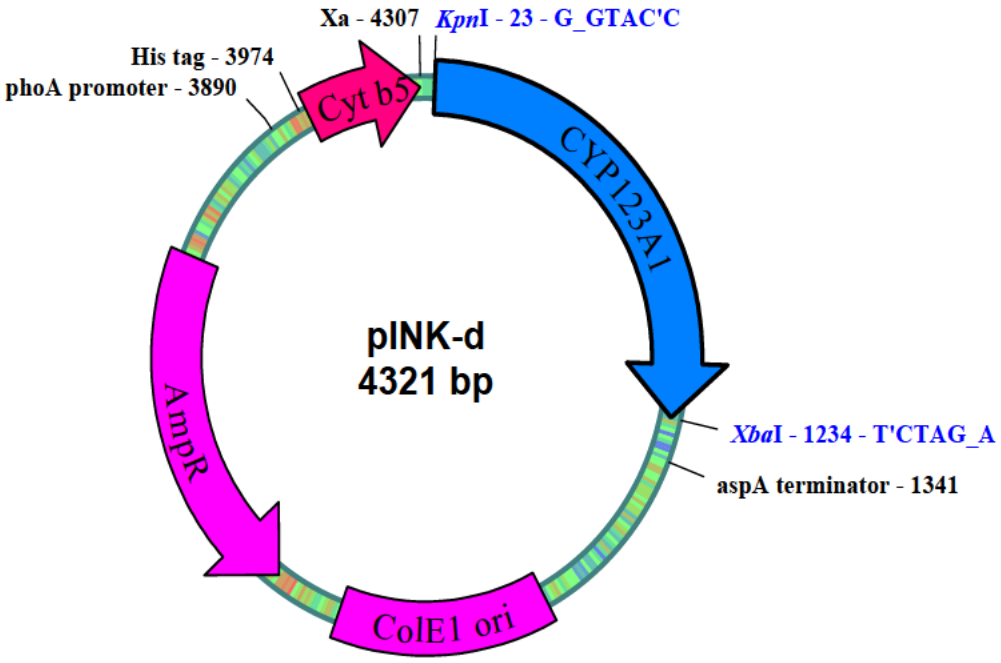

Figure 4.2. Vector maps showing the cloning of *CYP123A1* in pINK-d.

4.3.3. GenScript synthesis of CYP123A1 cDNA and cloning into expression vector

CYP123A1 cDNA, along with the expression vector, was synthesised by GenScript, USA. The constructs carrying CYP123A1 cDNA and empty expression vector were received in a microtiter plate. The constructs and the expression vector were resuspended in 100 µl of Tris-Hcl buffer (pH: 8.0) and 1 µl of this solution was used for transformation.

4.3.4. Generation of recombinant *E. coli* cells containing *CYP123A1*

The pINK-d construct containing CYP123A1 cDNA was transformed into *E. coli* DH5α using the transformation method described in the "Materials and methods" section (Inoue *et al.,* 1990). The recombinant cells were selected on LB-agar ampicillin plates. The figure below (Figure 4.3) indicates the growth of recombinant cells on ampicillin-containing plates.

This indicates that the constructs and expression vector were successfully transformed into *E. coli*.

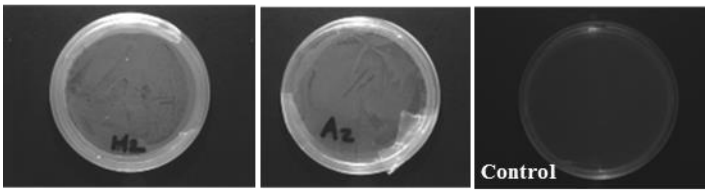

Figure 4.3. Selection of transformed *E. coli* DH5α on LB medium plates containing ampicillin antibiotic. H2: pINK-d_CYP123A1; A2: pINK-d vector and the control 1 plate: plate speared with only *E. coli* cells. No growth on control plate indicates no contamination during transformation.

From the above plates, a single colony was picked and plated on a separate LB-agar plate containing ampicillin. Figure 4.4 below shows the propagation of selected recombinant cells.

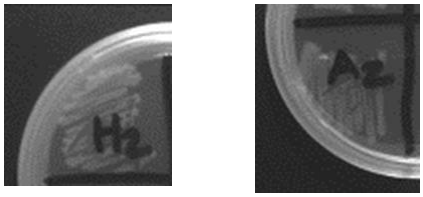

Figure 4.4. Propagation of selected recombinant *E. coli* cells containing CYP123A1 and pINK-d vector. The labels on plates follow *E. coli* cells containing the respective genes: A2: pINK-d_CYP123A1; H2: pINK-d vector.

4.3.5. Plasmid isolation and confirmation of the presence of *CYP123A1* cDNA

The above selected recombinant *E. coli* cells carrying CYP123A1 cDNA and expression vector were inoculated into 10 ml of LB broth and incubated at 37^{o}C and 150 rpm overnight.

The overnight bacterial culture was used for the isolation of plasmids. The plasmid DNAs were extracted using a QIAprep Spin Miniprep Kit and the concentration of the plasmid DNA was measured (Table 4.5) before it was subjected to restriction enzyme digestion analysis with enzymes specified in Table 4.5.

Table 4.5. Recombinant plasmid DNA concentration and enzymes used for releasing the inserted cDNA.

Name of the vector	Yield (µg/ml)	Restriction enzymes
pINK-d_CYP123A1	340	KpnI and XbaI
pINK-d	278.3	EcoRI and HindIII

After digestion of different vectors, the digested products were run on a 1% agarose gel. The figure below (Figure 4.5) shows the results of the gel electrophoresis, which indicate that CYP123A1 cDNA was cloned using the suggested restriction enzymes and thus the correct size of insert was released upon digestion.

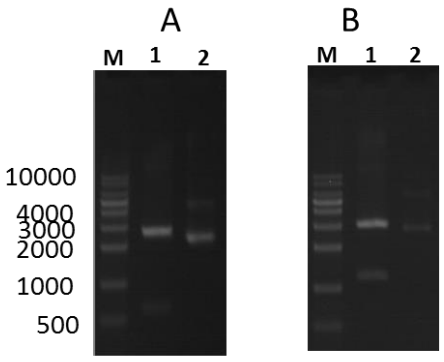

Figure 4.5. Restriction enzyme digestion analysis of recombinant *M. tuberculosis* P450 cDNA plasmids. 300 ng of each plasmid was digested in a water bath at 37°C for 40 min, and then later run on a 1% agarose gel. Panels A and B indicate: A: pINK-d, B: pINK-

d_CYP123A1. In the panels, Lane M represents molecular weight markers (1 KB DNA ladder) and Lanes 1 and 2 represent digested and undigested plasmid, respectively. The respective enzymes used for digestion of plasmids were shown in Table 4.5.

4.4. CONCLUSION

In conclusion, the multiple cloning site of the pINK-A vector was successfully modified by adding more suitable restriction enzymes and the revised vector was named pINK-d. In total 16 restriction enzymes were added in such a way that they did not cause any shift or change in the reading frame of the vector. The new vector can be used in future for cloning other P450s. Forward and reverse primers were also carefully designed to contain appropriate restriction enzyme sites. The developed strategy will be used to synthesise vectors and genes and clone the genes in the expression vector.

Furthermore, CYP123A1 cDNA and the modified expression vector were successfully synthesised by GenScript. Synthesised *CYP123A1* was further cloned into the expression vector. The construct containing the CYP123A1 gene and the pINK-d vector was transformed into *E. coli* cells and recombinant cells were selected. Plasmids were isolated and the presence of the correct size insert (*CYP123A1*) was verified by restriction enzyme digestion analysis. Recombinant *E. coli* cells carrying *CYP123A1* and the expression vector were stored at -80^{o}C for further expression analysis of CYP123A1.

4.5 REFERENCES

- Bellamine, A., Mangla, A.T., Nes, W.D. and Waterman, M.R., 1999. Characterization and catalytic properties of the sterol 14α-demethylase from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *96*(16), pp. 8937-8942.
- Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., II III, C.B. and Tekaia, F., 1998. Erratum: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *396*(6707), p. 190.
- Driscoll, M., 2011. Investigating orphan cytochromes P450 from *Mycobacterium tuberculosis*: The search for potential drug targets. https://www.escholar.manchester.ac.uk/uk-ac-man-scw:122243
- Driscoll, M.D., McLean, K.J., Cheesman, M.R., Jowitt, T.A., Howard, M., Carroll, P., Parish, T. and Munro, A.W., 2011. Expression and characterization of Mycobacterium tuberculosis CYP144: common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. *Biochimica et biophysica acta (BBA) - Proteins and proteomics*, *1814*(1), pp. 76-87.
- Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E., Leys, D. and Munro, A.W., 2010. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *Journal of biological chemistry*, *285*(49), pp. 38270-38282.
- Inoue, H., Nojima, H. and Okayama, H., 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96, pp. 23-28.

- Johnston, J.B., Kells, P.M., Podust, L.M. and de Montellano, P.R.O., 2009. Biochemical and structural characterization of CYP124: A methyl-branched lipid ω-hydroxylase from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *106*(49), pp. 20687-20692.
- Johnston, J.B., Ouellet, H. and de Montellano, P.R.O., 2010. Functional redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. *Journal of biological chemistry*, *285*(47), pp. 36352-36360.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, 45(D1), pp. D353-D361.
- Karim, A., Kaderbhai, N., Evans, A., Harding, V. and Kaderbhai, M.A., 1993. Efficient bacterial export of a eukaryotic cytoplasmic cytochrome. *Bio/technology*, *11*.
- McLean, K.J., Cheesman, M.R., Rivers, S.L., Richmond, A., Leys, D., Chapman, S.K., Reid, G.A., Price, N.C., Kelly, S.M., Clarkson, J. and Smith, W.E., 2002. Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*. *Journal of inorganic biochemistry*, *91*(4), pp. 527-541.
- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and Munro, A.W., 2009. The structure of *Mycobacterium tuberculosis* CYP125 molecular basis for cholesterol binding in a P450 needed for host infection. *Journal of biological chemistry*, *284*(51), pp. 35524-35533.
- Ouellet, H., Podust, L.M. and de Montellano, P.R.O., 2008. *Mycobacterium tuberculosis* CYP130 crystal structure, biophysical characterization, and interactions with antifungal azole drugs. *Journal of biological chemistry*, *283*(8), pp. 5069-5080.

CHAPTER 5

CONCLUSION AND FUTURE PERSPECTIVES

In this study, molecular evolutionary dynamics of P450s were carried out with special focus on mycobacterial P450s. Furthermore, *in silico* structural analysis of CYP123A1 was carried out and the binding of different azole drugs was assessed. Furthermore, CYP123A1 is successfully cloned in a novel expression vector pINK-d and recombinant cells carrying this P450 were successfully generated. Future work includes expression analysis of CYP123A1 in *E. coli*, purification and assessment of the binding affinity with different azole drugs, including substrate identification.

CHAPTER 6

RESEARCH OUTPUTS

PUBLICATIONS

1. **Parvez, M.,** Qhanya, L.B., Mthakathi, N.T., Kgosiemang, I.K.R., Bamal, H.D., Pagadala, N.S., Xie, T., Yang, H., Chen, H., Theron, C.W. and Monyaki, R., 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: special focus on mycobacterial P450s. *Scientific reports*, 6, p.33099.

2. Qhanya, L.B., Matowane, G., Chen, W., Sun, Y., Letsimo, E.M., **Parvez, M.,** Yu, J.H., Mashele, S.S. and Syed, K., 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete biotrophic plant pathogens. *PloS one*, *10*(11), p.e0142100.

3. Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., **Parvez, M.,** Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific reports*, 5. p. 11572.

CONFERENCE ABSTRACT/POSTER PRESENTATIONS

1. **M Parvez,** SS Mashele, K Syed (2017) Structure analysis of cytochrome P450 monooxygenase CYP123A1 from *Mycobacterium tuberculosis* H37Rv. The Annual South African Pharmacology Conference. Faculty of Health Sciences, University of the Free State, Bloemfontein, 01-04 October, 2017.

2. **M Parvez**, SS Mashele, K Syed (2016) Genomewide identification and annotation of cytochrome P450 monooxygenase CYP123A1 in sixty mycobacteria. The 13th International Symposium on Cytochrome P450 Biodiversity and Biotechnology, 22-26 July 2016, Vancouver, BC, Canada.

3. **M Parvez,** SS Mashele, K Syed (2015) Genome-wide identification and annotation of cytochrome P450 monooxygenase CYP123A1 in sixty mycobacterial species. International Symposium on Methods for Studying Drug Metabolism and Transport, and African

Traditional Medicines (METHODS-2015) from 23-25 November 2015 at St Georges Hotel and Conference Center, Pretoria, South Africa.

4. MM Sello, N Jafta, DR Nelson, W Chen, JH Yu, **M Parvez,** IKR Kgosiemang, R Monyaki, SC Raselemane, LB Qhanya, NT Mthakathi, SS Masheel, K Syed (2015) Genome-wide annotation and phylogenetic analysis of cytochrome P450 monooxygenases in pathogenic oomycetes. International Symposium on Methods for Studying Drug Metabolism and Transport, and African Traditional Medicines (METHODS-2015) from 23-25 November 2015 at St Georges Hotel and Conference Center, Pretoria, South Africa.

5. LB Qhanya, G Matowane, W Chen, Y Sun, EM Letsimo, **M Parvez,** JH Yu, SS Mashele, K Syed (2015) Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in basidiomycete biotrophic plant pathogens. International Symposium on Methods for Studying Drug Metabolism and Transport, and African Traditional Medicines (METHODS-2015) from 23-25 November 2015 at St Georges Hotel and Conference Center, Pretoria, South Africa.

6. WI Booi, P Nkolanyane, L Wieteska, **M Parvez,** M van Wyk, SS Mashele, D Grontand, K Syed (2015) *In silico* approach to understand the molecular basis of ketoconazole resistance in chronic granulomatous infectious fungus *Sporothrix schenckii*. International Symposium on Methods for Studying Drug Metabolism and Transport, and African Traditional Medicines (METHODS-2015) from 23-25 November 2015 at St Georges Hotel and Conference Center, Pretoria, South Africa.

MEDIA COVERAGE

pressreader

CUT's groundbreaking discovery

New discovery may help fight aquatic animal infection

leading: Research team leader Dr Khajamohiddin Syed and faculty of environmental science acting dean professor Sam Mashele.

pioneers: The CUT science students who helped discover the drug that may help fight aquatic animal infection.

Acting dean of the faculty Prof Sam Mashele said some of the deadliest pathogens in the world were found in the fish farming industry.

He said the discovery was made by researchers from the unit for drug discovery in the institution.

"For many years researchers across the world have been trying to understand these micro-organisms in order to control the disease and develop novel drugs against these pathogens, and CUT researchers are leading the way in finding solutions

that will bring an end to this socio-economic challenge facing aquatic farming," Mashele said.

He said microorganisms were widely known in the scientific world and they continued to wreak havoc on the aqua farming sector worldwide.

"They are considered the deadliest of pathogens, causing diminished production of aquatic food," Mashele said.

The team of researchers who discovered the drug were led by Dr Khajamohiddin Syed.

"We collaborated with highly acclaimed international scientists such as professors David Nelson from the University of Tennessee in the US, ae-Hyuk Yu from the University of Wis-consin Madison and Dr Wanping

University in China," Mashele said.

"The university researchers are investigating solutions that would sustain the aquatic resources while helping to increase high production levels of aqua farming for commercial purposes, food security and poverty alleviation."

He said their work highlighted the important role aqua farming plays in promoting healthy living and in fighting poverty and hunger.

Syed said aqua farming was a big industry across the world and it in-

prawns, squid and octopus and is considered by the UN as an important sector that can provide a livelihood for more than 60 million people in Africa and Asia.

"Consumption of these animals remains a vital source of protein and essential nutrients, especially for developing countries, where they constitute almost half of the total value of their traded communities," Syed said.

"The results of this study have been accepted for publication in the Nature Publication Group journal Scientific Reports, a prestigious multidisciplinary scientific international journal with an impact factor of 5.1."

Navorsers by SUT maak deurbraak