Cytochrome P450 monooxygenase CYP53 family in fungi: Structural analysis of CYP53A and its redox partner from the thermophilic fungus *Thielavia terrestris*

By

Poojah Jawallapersand

Dissertation submitted in fulfilment of the requirements of the degree

Magister Technologiae (M.Tech) in Biomedical Technology

Department of Health Sciences

Central University of Technology, Free State

03 August 2015

Promoter: Dr. Khajamohiddin Syed

Co-Promoter: Prof. Samson Sitheni Mashele

DECLARATION

I, **POOJAH JAWALLAPERSAND**, hereby certify that the dissertation submitted by me for the degree MAGISTER TECHNOLOGIAE (M.Tech): 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).

POOJAH JAWALLAPERSAND **DATE**

ACKNOWLEDGEMENTS

I would like to extend my sincere and heartfelt gratitude to the following people and organizations. The completion and success of my Masters study would not have been possible without any of them.

- My supervisor, Dr. Khajamohiddin Syed, for the opportunity to be part of his research team, sharing insight on molecular biology and bioinformatics, constant guidance, mentorship, invaluable constructive criticism and for allowing me to grow and become an autonomous scientist.

- Prof. Samson Setheni Mashele for his mentorship, guidance and support.

- Dr. Nada Kraševec, National Institute of Chemistry, Ljubljana, Slovenia and her research group for their valued expertise and input on protein modelling.

- Dr. Nataraj Sekhar Pagadala and Dr Jack Tuszynski, University of Alberta, Canada for their assistance and advice on protein modelling.

- Research and Innovation Fund, Central University of Technology, Free State for funding my studies and enabling dreams to be realised.

- Dr. Pankaj Sharma, for showing me the way when I was at crossroads, for introducing me to the Central University of Technology, Free State and its wonderful people, as well as for encouraging me to study further for my Masters degree. I certainly would never have been here without your valuable advice and suggestions.

- My colleagues at the Faculty of Health and Environmental Sciences, especially the Department of Clinical Sciences and Department of Biomedical Technology at the Central University of Technology for their continuous support, assistance and

encouragement. I am very blessed that the list is too long to mention, I am grateful to all of you.

- Prof. Hesta Frederick-Nel, Mrs Gloriah Molefe and the Emergency Medical Care team for always supporting me and uplifting my spirit when I needed it the most.

- To my dearest friends and family, for being my ardent aficionados. Their continuous faith in me, prayers, moral support and motivation, gave me strength and inspiration during my studies. I am grateful for your love and support.

- To my dearest Aunt Kavita Ragoobar and Aunt Pratima Ragoobar for being my guardian angels and well-wishers. To Binta Devi and Jankee Jawallapersand for encouraging me to pursue my dreams and believing in my capabilities.

- To my late uncles Bhai Ramesswar, Bhai Assoodeo and grandmother Sarodah, this thesis is dedicated to you as a legacy to the family name, not a day goes by where you are all not thought of and dearly missed.

- To my father, mother and sister for always believing in me, being my pillars of support, loving me unconditionally and respecting my decisions. Thank you for nurturing me into who I am today. I am blessed to have all of you.

- To Siddhartha Gautama, for giving me enlightenment, blessing me with grace and for being my Bodhisattva in this lifetime.

- To my Creator for His guidance, gracious blessings, loving protection and for inner peace. I glorify thee. Fear, ignorance and delusion were enemies that I encountered along the way during this intellectual journey, but from that self-experience, I realised that man is his own enemy. However, faith, hope and perseverance prove that the soul indeed has

its own intrinsic Divinity and that nothing is impossible, no matter how dark the shadows are that surround you. Truth triumphs at the end.

"Arise, awake, sleep no more: within each of you there is the power to remove all wants and all miseries. Believe this, and that power will be manifested."

- Swami Vivekananda

CONTENTS

		Page
LIST OF ABBREVIATIONS AND ACRONYMS		**X**
LIST OF FIGURES		**XVII**
LIST OF TABLES		**XIX**
ABSTRACT		**XXI**
CHAPTER 1: LITERATURE REVIEW		
1.1.	**Introduction to cytochrome P450s**	**1**
1.1.1.	Cytochrome P450 monooxygenase: General aspects	**1**
1.1.2.	Terminology and nomenclature of cytochrome P450s	**2**
1.1.3.	Cytochrome P450 monooxygenases and the chemistry of oxygen activation	**4**
1.1.4.	Cytochrome P450 reductase (CPR)	**5**
1.2.	**Role of P450s in the generation of human valuables**	**12**
1.3.	**Thermostable P450s**	**15**
1.3.1.	Introduction to thermostable P450s	**15**
1.3.2.	Unravelling the lower eukaryote fungal genomes for thermostable P450s	**19**
1.4.	**Rationale and objectives of the study**	**26**

References **29**

CHAPTER 2: GENOME-WIDE IDENTIFICATION, ANNOTATION AND COMPARATIVE ANALYSIS OF CYP53A FAMILY IN FUNGI

2.1. Introduction **40**

2.2. Materials and methods **47**

2.2.1. Genome data mining and annotation of CYP53 members **47**

2.2.2. Phylogenetic analysis **48**

2.2.3. Intron-exon analysis **49**

2.2.4. Analysis of homology **49**

2.2.5. Analysis of amino acid conservation **50**

2.3. Results and discussion **50**

2.3.1. CYP53 distribution in fungi **50**

2.3.2. Phylogenetic analysis of CYP53 P450 family **52**

2.3.3. High conservation of primary structure of CYP53 members in ascomycota **54**

2.3.4. Gene conservation and genome duplications of CYP53 members **59**

2.3.5. Functional significance of CYP53 family and its potential role as a common antifungal drug target **63**

2.4. Conclusion **66**

References **68**

Appendix A **76**

CHAPTER 3: STRUCTURAL ANALYSIS OF CYTOCHROME P450 MONOOXYGENASE CYP53A FROM THE THERMOPHILIC FUNGUS *THIELAVIA TERRESTRIS*

3.1. Introduction **97**

3.2. Materials and methods **99**

3.2.1. Homology modelling **99**

3.2.2. Active site residue mapping **100**

3.3. Results and discussion **101**

3.3.1. Homology modelling **101**

3.3.2. Secondary structure analysis of CYP53A *Tter* **103**

3.3.3. Active site cavity and its amino acid residues identification **105**

3.4. Conclusion **107**

References **109**

CHAPTER 4: STRUCTURAL ANALYSIS OF CYTOCHROME P450 REDUCTASE (CPR) FROM THE THERMOPHILIC FUNGUS *THIELAVIA TERRESTRIS*

4.1. Introduction **112**

4.2. Materials and methods **114**

4.2.1. Selection of template **114**

4.2.2. Homology modeling and 3D model validation **115**

4.2.3. Active site identification and ligand docking **115**

4.3. Results and discussion **116**

4.3.1. CPR *Tter* modeling and 3D model validation **116**

4.3.2. Comparative structural analysis of CPR *Tter* and CPR *Scer* **118**

4.3.3. Identification of amino acids binding to cofactors in CPR *Tter* **119**

4.4. Conclusion **122**

References **123**

CHAPTER 5: CONCLUSION **128**

RESEARCH PUBLICATION

LIST OF ABBREVIATIONS

$	Dollar unit of currency
%	Percentage
≥	Greater or equal to
>	Greater than
±	Plus –minus
°C	Degree Celsius
3D	Three-dimensional
Å	Angstrom
Abi	*Agaricus bisporus*
Acl	*Aspergillus clavatus*
Ade	*Auricularia delicata*
Afl	*Aspergillus flavus*
Afu	*Aspergillus fumigatus*
Ani	*Aspergillus niger*
Anid	*Aspergillus nidulans*
Aor	*Aspergillus oryzae*
Ate	*Aspergillus terreus*
ATM	Atom
Bad	*Bjerkandera adusta*
C	Carbon

CASTp	Computed Atlas of Surface Topography of proteins
C-C	Carbon-carbon bond
chemcorp	Chemical Corporation
Cim	*Coccidioides immitis*
Clu	*Cochliobolus lunatus*
ClustalW2	Multiple sequence alignment program
CO	Carbon monoxide
C-O	Carbon-oxygen bond
CPR	Cytochrome P450 reductase
Cpu	*Coniophora puteana*
Csu	*Ceriporiopsis subvermispora*
CYP	Cytochrome P450
dDFIRE	Updated energy function of DFIRE
DFIRE	Distance-scaled, finite ideal-gas reference
DFIRE2	Updated energy function of DFIRE
DNA	Deoxyribonucleic acid
Dsq	*Dichomitus squalens*
ER	Endoplastic reticulum
ERRAT	Server detecting errors in protein models
et al.	*Et alia* (and others)
FAD	Flavin adenine dinucleotide

FASTA	File format for DNA and protein sequences
FdR	Ferredoxin reductase
Fdx	Ferredoxin
Fe^{2+}	Iron (II) cation
Fe-S	Iron-sulphur
Fgr	*Fusarium graminearum*
Fldx	Flavodoxin
Fme	*Fomitiporia mediterranea*
FMN	Flavin mononucleotide
Fox	*Fusarium oxysporum*
Fpi	*Fomitopsis pinicola*
Fve	*Fusarium verticillioides*
G	Glycine
Glu	*Ganoderma lucidum*
Gsp	*Ganoderma* sp.
Gtr	*Gloeophyllum trabeum*
H	Hydrogen
HEM	Heme group
HIV	Human immunodeficiency virus
I	Isoleucine
ID	Identity

JGI	Joint Genome Institute
kDa	kilodalton
LLC	Limited liability company
Mfi	*Mycosphaerella fijiensis*
Mgr	*Magnaporthe grisea*
MH	Membrane helix
MIT	Massachusetts Institute of Technology
MOE	Molecular Operating Environment
Mth	*Myceliophthora thermophila*
N-	Substitution of nitrogen atom
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
Ncr	*Neurospora crassa*
Ndi	*Neurospora discreta*
Nfi	*Neosartorya fischeri*
Nha	*Nectria haematococca*
nm	Nanometre
NS	New subfamily
N-terminal	Amino terminal end
O-	Substitution of oxygen atom

OFOR	2-Oxoacid: ferredoxin oxidoreductase
P450	Cytochrome P450
PAH	Polycyclic aromatic hydrocarbons
Pbr	*Phlebia brevispora*
Pca	*Phanerochaete carnosa*
Pch	*Phanerochaete chrysosporium*
PDB	Protein Data Bank
PFOR	Phthatate-family oxygenase reductase
Pgr	*Puccinia graminis*
pH	Power of hydrogen
Ppl	*Postia placenta*
PROCHECK	A program that checks the stereochemical quality of a protein structure
PROMALS3D	PROfile Multiple Alignment with predicted Local Structures and 3D constraints
Pst	*Punctularia strigosozonata*
Q	Glutamine
QMEAN	Qualitative Model Energy ANalysis
RDX	Hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine
RH	Substrate
R-OH	Hydroxylated product
S-	Substitution of sulphur atom

Scer	*Saccharomyces cerevisiae*
Shi	*Stereum hirsutum*
Sla	*Serpula lacrymans*
sp.	Species
Sro	*Sporobolomyces roseus*
SRS	Substrate recognition site
TAFF	Thermally activated flux flow
T_m	Melting temperature
TMH	Transmembrane helix
Tte	*Thielavia terrestris*
Tve	*Trametes versicolor*
Uma	*Ustilago maydis*
Ure	*Uncinocarpus reesii*
US	United States of America
US-DOE	United States of America Department of Energy
Wco	*Wolfiporia cocos*
WHAT_CHECK	A system for protein structure validation derived from the WHAT IF program
WHAT IF	A protein structure analysis program that may be used for mutant prediction, structure verification and molecular graphics
www	World wide web
Y	Tyrosine

YASARA	Yet Another Scientific Artificial Reality Application
Z-score	Indicates overall model quality

LIST OF FIGURES

		Page
Figure 1.1	Typical cytochrome P450 reduced-CO difference spectrum.	**3**
Figure 1.2	A typical cytochrome P450 catalytic reaction.	**5**
Figure 1.3	Schematic representations of different types of P450 redox systems.	**9**
Figure 1.4	Phylogenetic tree of the P450ome of *T. terrestris* (*Thite*-red) *and M. thermophila* (*Spoth*-blue).	**20**
Figure 2.1	Phylogenetic analysis of CYP53 family in fungi.	**53**
Figure 2.2	Analysis of amino acid conservations in CYP53 family members of ascomycota (A) and basidiomycota (B).	**57**
Figure 2.3	Analysis of amino acid conservations in CYP53C subfamily of basidiomycota.	**58**
Figure 2.4	Gene-structure analysis of CYP53 family in ascomycete species.	**61**
Figure 2.5	Gene-structure analysis of CYP53 family in basidiomycete species.	**62**

Figure 3.1	Structural analysis of CYP53A *Tter*.	**102**
Figure 3.2	Secondary structure analysis of CYP53A *Tter* and CYP51 (4LXJ).	**104**
Figure 3.3	CYP53A *Tter* active site cavity mapping and analysis of the nature of the active site amino acids.	**105**
Figure 4.1	Three dimensional structure of cytochrome P450 reductase from *T. terrestris.*	**117**
Figure 4.2	Superimposition of CPR *Tter* with template (CPR *Scer*).	**119**
Figure 4.3	Ligand protein interactions of FAD, FMN and NAD(P).	**121**

LIST OF TABLES

		Page
Table 1.1.	Different classes of P450 redox systems.	**10**
Table 1.2.	Examples of extremophiles and the industrial applications of their isolated enzymes.	**17**
Table 1.3.	Genome-wide comparative analysis of the P450omes of thermophilic and mesophilic ascomycetes.	**21**
Table 1.4.	Analysis of thermostability of the P450omes of *T. terrestris* and *M. thermophila*.	**24**
Table 2.1.	Genome-wide comparative analysis of CYP53 family in fungi.	**41**
Table 2.2.	Analysis of homology between CYP53 members in fungi.	**54**
Table 3.1.	Validation of CYP53A *Tter* 3D-model.	**102**
Table 3.2.	Analysis of amino acid conservation in CYP53A *Tter* active site cavity.	**106**
Table 4.1.	Crystal structures and their residues of cytochrome P450 reductases from the Protein Data Bank (www.pdb.org).	**113**

Table 4.2. Validation of CPR *Tter* 3D-model. **117**

ABSTRACT

Cytochrome P450 monooxygenases (CYPs or P450s) are heme-thiolate proteins distributed across biological kingdoms. P450s show multiple and diverse catalytic activity on a wide range of substrates, and as such the use of these enzymes have been prompted in various areas such as the production of human valuable chemicals, pharmaceutical compounds, antibiotics, fragrances and the degradation of xenobiotic compounds. If P450s are to be used on a large-scale production, it is imperative that they are able to resist extreme industrial conditions, including thermostability. Considering the fact that P450s are weak and less stable enzymes, research has focused on identifying thermostable P450s. Furthermore, P450s have been used as a drug target against pathogens. However, a study revealed that pathogens are developing drug resistance against currently available drugs. To tackle and address this emerging dilemma to drug resistance, novel drug targets need to be discovered and identified.

The current study is the first of its kind that focuses on two aspects, which are identifying a common alternative anti-fungal drug target and structural characterization of thermostable P450 CYP53A and its redox partner cytochrome P450 reductase (CPR) from biomass-degrading thermophilic ascomycete *Thielavia terrestris*. This study also marks the beginning of our understanding on thermostable P450s from eukaryotes.

Part of the study has been published as an article in PLoS ONE journal (impact factor 3.5). Below are the details of the manuscript:

Jawallapersand P, Mashele SS, Kovačič L, Stojan J, Komel R, Pakala SB, Kraševec N, Syed K. 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. 2014 Sep 15;9(9):e107209. doi: 10.1371/journal.pone.0107209. eCollection 2014.

CHAPTER 1

LITERATURE REVIEW

1.1. Introduction to cytochrome P450s

1.1.1. Cytochrome P450 monooxygenases: General aspects

Cytochrome P450 monooxygenases (CYPs or P450s) comprise of a very extensive superfamily of diverse heme-thiolate proteins (Nelson, 2013). P450s exist in nature in all phylogenetic domains of life ranging from microscopic prokaryotes such as archaea and bacteria to lower eukaryotes such as protists, fungi and insects and ultimately to more complex eukaryotes including plants and animals (Nelson, 2013). Plants have the greatest number of P450 genes with animals, fungi, protists, archaea and bacteria having lower numbers of P450s (Nelson, 2011). Fungal genomes have lower numbers of P450s compared to plants. However, fungal genomes express the highest P450 diversity among all other phylogenetic domains, with about 399 P450 families located throughout the 2784 annotated fungal P450s (as of 2011) in comparison to plant genomes having only 129 P450 families located throughout 4267 annotated plant P450s (Nelson, 2009; Nelson, 2011).

Currently, there are more than 21039 P450s (as of 2013) that have been sequenced, annotated and described (Nelson, 2009). P450s have existed for billions of years and have played pivotal roles from insecticide resistance in mosquito disease-carrying vectors to bio-remediation of hexahydro-1,3,5-trinitro-1,3,5-triazine (Royal Demolition Explosive, RDX) on contaminated soil (Rylott *et al*., 2011 (a, b); Nelson, 2013). P450s are indeed primordial and have played key role in organisms adapting to various types of ecological niches, by even adapting to extreme conditions such as elevated hydrostatic pressures and temperatures of solfataric hot springs (Park *et al*., 2002; Nelson, 2013) and adaption to utilize

photosynthetically fixed carbon such as wood (Syed *et al*., 2014a). Hence, if P450s were wiped out, life on Earth would probably relapse to only a pool of primitive organisms (bacteria) with all other life forms dwindling or ceasing to exist (Nelson, 2013).

1.1.2. Terminology and nomenclature of cytochrome P450s

The term "Cytochrome P450" was derived by Omura and Sato (1962-1964) who referred this protein as 'pigment 450'. This is due to P450s are hemeproteins by nature and display unusual spectral properties in a reduced state of CO-bound (carbon monoxide-bound) complex at a maximum absorption peak at a wavelength of 450 nm (Figure 1.1) (Omura and Sato, 1962; Omura and Sato, 1964; Klingenberg, 2003; Bernhardt, 2006; Luthra *et al*., 2011). Hence, "Cytochrome P450" actually stands for: cytochrome meaning hemoprotein, P is an abbreviation for pigment and 450 indicates the absorption peak spectra of the CO-bound complex at a wavelength of 450 nm (Figure 1.1) (Bernhardt, 2006; Guengerich *et al*., 2009; Luthra *et al*., 2011). The cysteine-thiolate group is the prime reason for this phenomenal spectral display observed in P450s, whereby there is formation of the fifth ligand of the heme-iron. Hence, the name heme-thiolate proteins or hemoproteins is given to P450 enzymes (Hannemann *et al*., 2007). The first occurrence of the P450 absorption spectrum was described in 1958 (Klingenberg, 1958).

The nomenclature of P450s is dependent and classified according to the percentage homology of their individual amino acid sequence, and by precept, are assigned the letters CYP for cytochrome P450 followed by a family number (≥ 40% sequence homology), an alphabetical letter for the subfamily (>55% sequence homology) and a unique number for the individual enzyme within the designated subfamily (Nebert *et al*., 1987; Nebert *et al*., 1991; Nelson, 2009). Although P450s show low sequence similarity, their secondary topography

and structural fold are highly conserved (Poulos *et al.*, 1987) indicting that P450s share similar mechanisms of oxygen activation (Denisov *et al.*, 2005). Hence, P450s are categorized according to their respective clans (large group of P450s families or genes) instead, which is becoming more acknowledged, since genes within respective clans have not originated from a single common ancestor, but rather diverged from a common ancestor and are thus more likely to share mutual functional traits and features (Nelson, 1998 (a, b); Nelson, 1999). Nevertheless, a conventional criterion concerning the classification of P450s into their respective clans has not firmly been established yet (Nelson, 1998b; Nelson, 1999; Nelson, 2006).

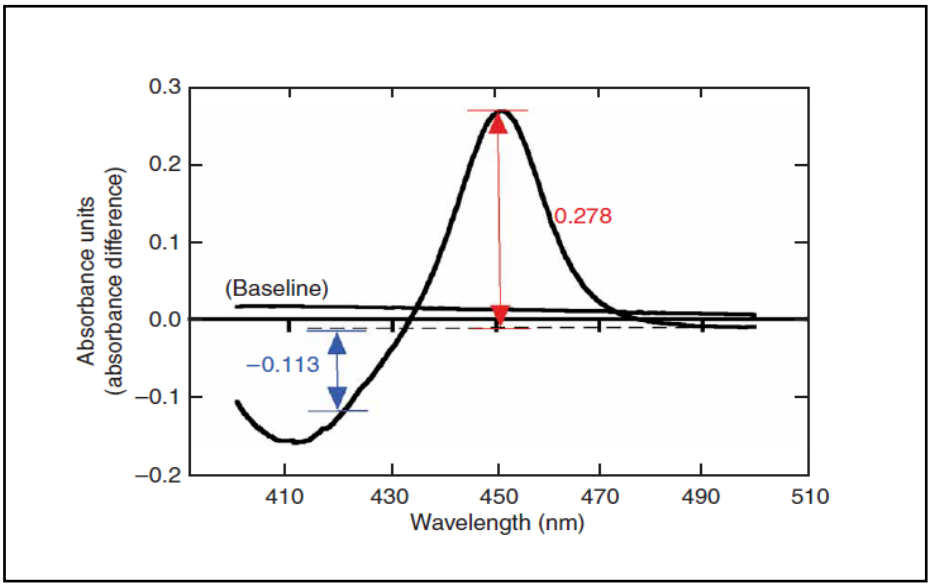

Figure 1.1. Typical cytochrome P450 reduced-CO difference spectrum. The P450 spectral assay is based on the principle that the ferrous form (Fe^{2+}) of the hemoprotein reacts with carbon monoxide (CO) to form a CO-bound complex that distinctively generates a spectrum with a maximum absorption at a wavelength of 450 nm (shown in red), as a result of the cysteine-thiolate axial ligand bound to the heme iron molecules present in the P450 enzymes. (Taken from Guengerich *et al.*, 2009).

1.1.3. Cytochrome P450 monooxygenases and the chemistry of oxygen activation

P450s are well known for their monooxygenation of substrates hence they are classified or described as monooxygenases in literature (Urlacher and Girhard, 2012). Monooxygenases (previously known as mixed function oxidases) are unique enzymes that catalyse molecular oxygen by inserting the first oxygen atom into a substrate with the second oxygen atom reduced to a water molecule, whereby two electrons are used in the process provided by the cofactors NADH or NADPH (reduced nicotinamide adenine dinucleotide phosphate) by means of an external reductase protein (Figure 1.2) (Denisov *et al.,* 2005). P450s are biocatalysts that also catalyse a wide variety of chemo-, regio- and stereo-specific oxidation reactions such as the activation of sp^3 hybridized C atoms, dehalogenation, hydroxylation of aliphatic and aromatic hydrocarbons, sulphoxidation, deamination, desulphuration, epoxidation, peroxidation, N-, O- and S- dealkylation (Sono *et al*., 1996; Bernhardt, 2006). Additionally, P450s carry out more unusual chemical reactions, examples being C-C and C-O phenol coupling, C-C bond cleavage, Baeyer-Villiger oxidation, re-arrangement reactions (such as ring formation and oxidative aryl migration)and isomerisation (Sono *et al*., 1996; Bernhardt, 2006).

Diverse catalytic reactions performed by P450s suggest that P450s have the remarkable ability to accept and bind to a diverse range of substrates. The assortment of substrates include important natural compounds like fatty acids, terpenes, steroids, prostaglandins, eicosanoids, fat-soluble vitamins, bile acids, aliphatic compounds, aromatic compounds and hetero-aromatic compounds as well as xenobiotics such as drugs, organic solvents, polycyclic aromatic hydrocarbons (PAHs), antibiotics, pesticides, herbicides, anaesthetics, alkyl aryl hydrocarbon products, ethanol, carcinogens, toxins and even explosive military compounds

(Roberts *et al*., 2002; Bernhardt, 2006; Urlacher and Eiben, 2006;Rylott*et al.*, 2011(a, b); Urlacher and Girhard, 2012; Syed *et al*., 2014a). Over the past five decades, studies have proven that P450s are indeed "versatile biocatalysts" with the presence of flexible substrate recognition sequences/regions (SRS). Hence, P450s are able to exhibit extraordinary chemical reactivity and catalyse a diverse range of substrates (Carmichael and Wong, 2001; Meunier *et al*., 2004; Bernhardt, 2006; Guengerich, 2006; Urlacher and Eiben, 2006; Hanneman *et al*., 2007; Grogan, 2011; Urlacher and Girhard, 2012).

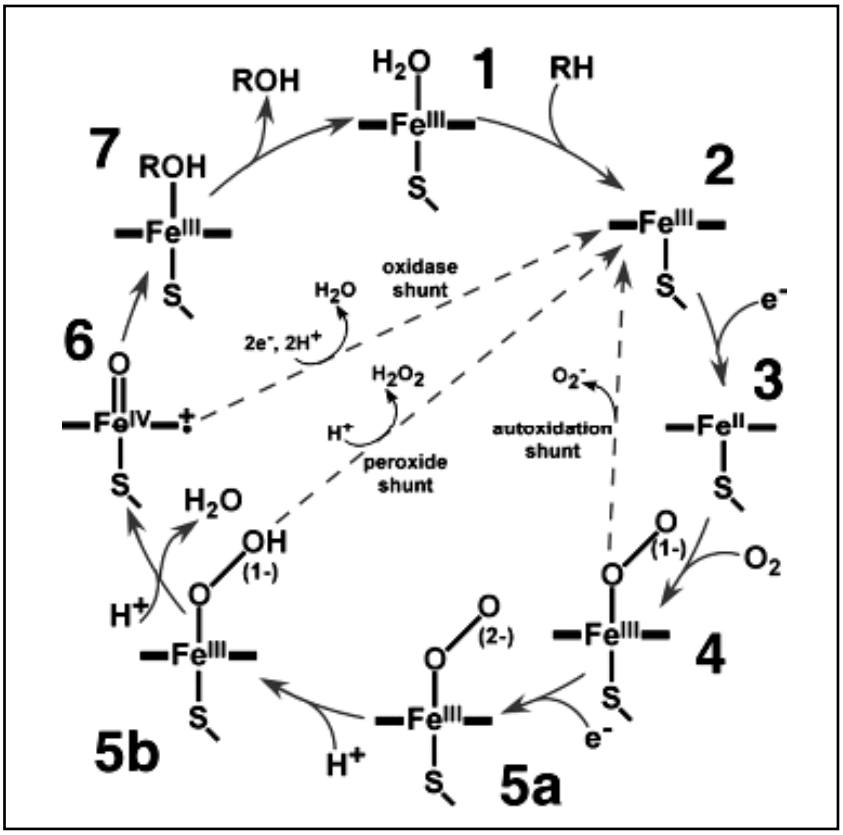

Figure 1.2. A typical cytochrome P450 catalytic reaction. RH and ROH represent the substrate and product respectively. (Taken from Denisov *et al.*, 2005).

1.1.4. Cytochrome P450 reductase (CPR)

In general, nearly all eukaryotic P450s require one or more redox partner proteins also known as cytochrome P450 reductases (CPRs) during catalytic reactions for a plethora of substrates

for the transfer of electron equivalents provided by pyridine cofactors NADH or NADPH to the heme centre of P450s (Denisov *et al*., 2005). The general chemical scheme applied by P450s is shown as:

$$RH + O_2 + 2e^- + 2H^+ \rightarrow R\text{–}OH + H_2O$$

where RH represents the substrate, R–OH is the hydroxylated product and NAD(P)H is usually the electron donor (Munro *et al*., 2013). However, the chemical equation can be sometimes misleading, as the mechanisms employed by P450s is much more complex (Munro *et al*., 2013).

P450s function as terminal electron acceptors in various reactions involving multiple constituent P450 contingent monooxygenation systems (Hannemann *et al*., 2007; Lah *et al*., 2008). In eukaryotic organisms both the P450 and CPR are situated in the membrane of the endoplasmic reticulum (ER) by a single transmembrane anchor with the catalytic domain positioned towards the cytosol and the hydrophobic area of the catalytic domain protruding into the phospholipid bilayer (Cojocaru *et al*., 2011). The P450 is anchored by a monotopic N-terminal transmembrane α-helix (Cojocaru *et al*., 2011). Experiments conducted using site-directed antibodies against peptides of the human CYP2B1 and the rabbit CYP2B4, proposed a suitable model that shows the microsomal topography of the P450 within the ER (De Lemos-Chiarandini *et al*., 1987; Black *et al*., 1994).

Previous studies conducted on P450s, suggested that these enzymes fall into two distinct classes namely, the mitochondrial/prokaryote type (class I) and the microsomal type (class II) (Mclean *et al*., 2005; Bernhardt, 2006; Hannemann *et al*., 2007). At present, P450s are categorized into different classes (Figure 1.3 and Table 1.1) and sub-classes based on their

auxiliary external redox partner proteins (Benhardt, 2006; Hannemann *et al.*, 2007; Urlacher and Girhard, 2012). The different P450 redox classes are discussed here briefly.

Class I P450s are situated in the inner mitochondrial membrane and are predominantly bacterial P450s (Bernhardt, 2006). Mitochondrial P450s are soluble enzymes of the cell matrix, and acquire electrons for catalysis from a NADH-dependent FAD-containing reductase constituting of an iron-sulphur [2Fe-2S] ferredoxin (adrenodoxin) and a flavin adenine dinucleotide (FAD) containing reductase (adrenodoxin reductase) (Mclean *et al.*, 2005; Bernhardt, 2006; Hannemann *et al.*, 2007; Urlacher and Girhard, 2012). Historically, the first bacterial P450 system belonged to a camphor hydroxylase, CYP101 (P450cam) from *Pseudomonas putida*, which was also the first microbial P450 to be enzymatically and structurally characterized (Mclean *et al.*, 2005; Bernhardt, 2006; Hannemann *et al.*, 2007). The electron donating mechanism in CYP101 involves electrons being transferred from NADH *via* a FAD-containing reductase (putidaredoxin reductase) and an iron–sulphur protein (putidaredoxin) to CYP101. Bacterial P450s are responsible for the metabolism of carbon compounds, xenobiotics, fatty acids and for the synthesis of secondary intermediates such as antibiotics and anti-fungal compounds (Hannemann *et al.*, 2007). Mitochondrial P450s in mammals catalyze the side-chain cleavage of cholesterol, 11β-hydroxylation of 11-deoxycortisol and are involved in the biosynthesis of vitamin D and aldosterone (Bernhardt, 2006).

In contrast, class II P450s are integral membrane proteins that bind to the endoplasmic reticulum by an N-terminal membrane anchor and obtain electrons *via* membrane-anchored NADPH-dependent reductase containing both FAD and FMN cofactors (CPR) (Mclean *et al.*, 2005; Bernhardt, 2006; Hannemann *et al.*, 2007).

However, the discovery of CYP102 (P450-BM3) isolated from *Bacillus megaterium* in the 1980s, revealed a unique, yet unusual characteristic, whereby the P450 system was that of a fused P450 protein to a reductase component, into one catalytically self-sufficient single-polypeptide chain (Roberts *et al.*, 2002; Mclean *et al.*, 2005; Bernhardt, 2006; Hannemann *et al.*, 2007). Consequently, since then, an explosive range and variety of P450s have been characterized, with novel and different P450 redox systems still being discovered today, that belong to neither class I or class II, but rather a class of their own (Figure 1.3) (Mclean *et al.*, 2005; Hannemann *et al.*, 2007; Urlacher and Girhard, 2012).

Apart from microsomal reductase such as CPR, some microsomal P450s are able to take a second electron from cytochrome b5 (Bernhardt, 2006). Cytochrome b5 is a 17 kDa heme protein coupled mainly with the microsomal and mitochondrial components of eukaryotic cells together with its redox protein partner NADH-cytochrome b5 reductase (Lederer *et al.*, 1983; Hannemann *et al.*, 2007). Cytochrome b5 and its reductase component are both electron transport elements that partake in various biochemical process from electron transfer to desaturases involved in the biosynthesis of unsaturated fatty acids, plasmalogen and sterols (Vergeres and Waskell., 1995; Hannemann *et al.*, 2007).

P450s, with the exception of self-sufficient P450s, interact with their corresponding redox protein partners of the catalytic cycle, as this specificity and criterion assures that an adequate reaction rate of catalysis results that prevents the systems from changing to shunt pathways (Bernhardt, 2006). The recognition of the auxiliary reductase by the P450 enzyme occurs as a result of salt bridges (Bernhardt, 2006). These salt bridges allow discrimination between potential electron donors and acceptors, whereby the correct orientation of the P450 to the reductase will be favoured (Bernhardt, 1996 & 2006).

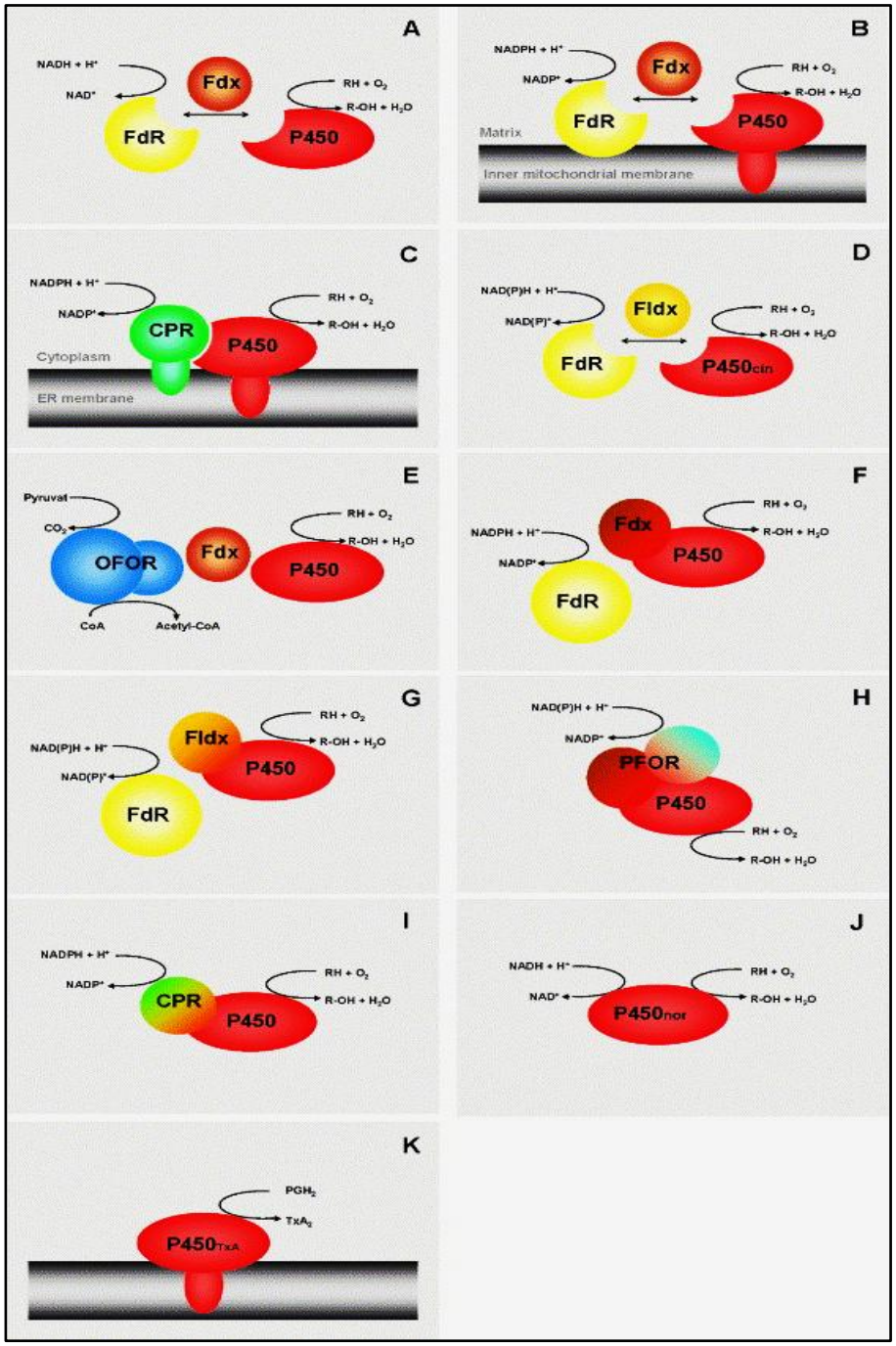

Figure 1.3. Schematic representations of different types of P450 redox systems. (A) Class I, bacterial system., (B) Class I, mitochondrial system., (C) Class II microsomal system.,(D) Class III, bacterial system., example P450cin., (E) Class IV, bacterial thermophilic system., (F) Class V, bacterial [Fdx]–[P450] fusion system., (G) Class VI, bacterial [Fldx]–[P450] fusion system., (H) Class VII, bacterial [PFOR]–[P450] fusion system., (I) Class VIII, bacterial [CPR]–[P450] fusion system., (J) Class IX, soluble eukaryotic P450nor., (K) Independent eukaryotic system, example P450TxA. (Taken from Hannemann *et al.*, 2007).

Table 1.1. Different classes of P450 redox systems. (Taken from Hannemann *et al.*, 2007).

Class/source	Electron transport chain	Localization/remarks
Class I		
Bacterial	NAD(P)H → [FdR] → [Fdx] [a] → [P450]	Cytosolic, soluble
Mitochondrial	NADPH → [FdR] → [Fdx] → [P450]	P450: inner mitochondrial membrane FdR: membrane associated Fdx: mitochondrial matrix, soluble
Class II		
Bacterial	NADH → [CPR] → [P450]	Cytosolic, soluble, *Streptomyces carbophilus*
Microsomal A	NADPH → [CPR] → [P450]	Membrane anchored, ER
Microsomal B	NADPH → [CPR] → [cytb5] → [P450]	Membrane anchored, ER
Microsomal C	NADH → [cytb5Red] → [cytb5] → [P450]	Membrane anchored, ER
Class III		
Bacterial	NAD(P)H → [FdR] → [Fldx] → [P450]	Cytosolic, soluble, *Citrobacter braakii*
Class IV		
Bacterial	Pyruvate, CoA → [OFOR] → [Fdx] → [P450]	Cytosolic, soluble, *Sulfolobus tokadaii*
Class V		
Bacterial	NADH → [FdR] → [Fdx–P450]	Cytosolic, soluble, *Methylococcus capsulatus*
Class VI		
Bacterial	NAD(P)H → [FdR] → [Fldx–P450]	Cytosolic, soluble, *Rhodococcus rhodochrous* strain 11Y
Class VII		
Bacterial	NADH → [PFOR–P450]	Cytosolic, soluble, *Rhodococcussp strain* NCIMB 9784, *Burkholderia sp.*, *Ralstonia metallidurans*
Class VIII		
Bacteria, fungi	NADPH → [CPR–P450]	Cytosolic, soluble, *Bacillus megaterium.*, *Fusarium oxysporum*
Class IX		
Only NADH dependent, fungi	NADH → [P450]	Cytosolic, soluble, *Fusarium oxysporum*
Class X		
Independent in plants/mammals	[P450]	Membrane bound, ER

Table notes:

Abbreviated protein components with redox centres:

Fdx, Ferredoxin (iron–sulphur-cluster)

[a]Fdx, Ferredoxin containing iron–sulphur-cluster of [2Fe–2S], [3Fe–4S], [4Fe–4S], [3Fe–4S]/ [4Fe–4S] type.

FdR, Ferredoxin reductase (FAD)

CPR, Cytochrome P450 reductase (FAD, FMN)

Fldx, Flavodoxin (FMN)

OFOR, 2-Oxoacid ferredoxin oxidoreductase (thiamin pyrophosphate, [4Fe–4S] cluster)

PFOR, Phthatate-family oxygenase reductase (FMN, [2Fe–2S] cluster).

1.2. Role of P450s in the generation of human valuables

P450s have immense biotechnological potential, due to their innate ability to catalyze important reactions on a diverse range of substrates (Bernhardt, 2006). Today, research on P450s has become a field of extreme interest and curiosity because of the remarkable chemical properties of P450s, especially in areas such as biotechnology, pharmacokinetics, pharmacology, toxicology and environmental sciences (Miners, 2002, Carmichael and Wong, 2001; Meunier *et al*., 2004; Bernhardt, 2006; Guengerich, 2006; Urlacher and Eiben, 2006; Hanneman *et al*., 2007; Grogan, 2011; Urlacher and Girhard, 2012). P450s have been used for the production of fine chemicals, fragrances, pharmaceutical products as well as in bio-fuel production, bio-sensing and bioremediation (Miners, 2002, Carmichael and Wong, 2001; Meunier *et al*., 2004; Bernhardt, 2006; Guengerich, 2006; Urlacher and Eiben, 2006; Hanneman *et al*., 2007; Grogan, 2011; Zhang *et al*., 2011; Urlacher and Girhard, 2012). Here the role and importance of P450s have been explored and briefly described.

A practical example whereby P450s have been used for large-scale market production is of transgenic plants, namely, the blue roses (Bernhardt, 2006). Scientists from Florigene, Australia and from Suntory, Japan managed to synthesize delphinidin, which currently accounts for the blue colour, in roses (Holton *et al*., 1993; Ogata *et al*., 2005; Bernhardt, 2006). The technique used involves the incorporation of the respective P450 gene (CYP75A). This initiative has taken flower cultivation to an entirely new dimension, where not only can the colour of roses, but also carnations be successfully changed to different shades ranging from blue to purple depending on the clientele (Bernhardt, 2006). The demand for cut flowers is significantly high, whereby the global flower market is valued at an approximate amount of US$27 billion per annum, which demonstrates that flower production does indeed have a huge economic impact (Bernhardt, 2006).

P450s have also been widely applied in the specialized field of drug discovery and development, for the biotransformation of steroids to drugs (Bernhardt, 2006; Guengerich, 2006; Urlacher and Eiben, 2006). In the *Mycobacterium* sp. strain HXN-1500, hydroxylation of 1-limonene in the 7-position into perillyl alcohol having anti-carcinogenic properties is produced (van Beilen *et al*., 2005). The transformation of compactin into pravastatin by microbial oxidation in *Streptomyces* sp., produces a drug that lowers the level of cholesterol in blood by inhibition of biosynthesis of cholesterol (Park *et al*., 2002). In the *Curvularia* sp., Reichstein S is transformed into hydrocortisone, due to the process of 11β-hydroxylation, where there is enantio- and regio- selective insertion of an oxygen molecule into Reichstein S (van Beilen *et al*., 2003). Another example is the biotransformation of progesterone to cortisone (Peterson *et al*., 1952; Hogg, 1992; Syed and Yadav, 2012). An example of an antibiotic producing P450, CYP107A1 (P450eryF) from *Saccharopolyspora erythraea*, involved in the initial hydroxylation of 6*S*-hydroxylation of 6-deoxyerythronolide B in a multiple erythromycin synthesis pathway, eventually converting 6-deoxyerythronolide B into erythromycin (Cupp-Vickery and Poulos, 1995; Sen and Thiel; 2014). The role of P450s for the biosynthesis of biological molecules of medicinal value is indeed exceptional and current innovative results clearly demonstrate the importance of P450s in pharmaceutical development and technology.

P450s are applied *in vivo* to test drug toxicity to evaluate the effects of prodrugs, toxic substances and xenobiotic compounds on human metabolism (Isin and Guengerich, 2007). P450s have also been applied for the design of biosensors, a technology with potential long-term benefits (Paternolli *et al*., 2004). Currently, the average cost of introducing a novel drug to the market takes about 10-15 years of laborious development and clinical trials, at an estimated cost and investment of US$750 million to more than US$5 billion, with the regrettable probability that 95% of the experimental drugs that are studied in humans, being

ineffective and unsafe (Myszka and Rich, 2000; Herper, 2013). The solution to this vicious cycle could be the implementation of biosensors based on P450s as potential and cost-effective screening tools in the multi-million dollar pharmaceutical industry. The reason why P450s are suitable candidates for the design of biosensors is due to the fact that more than 1,000,000 different xenobiotic and endobiotic liphophilic substrates as well as approximately 75% of all drug compounds are metabolized by P450s (Baj-Rossi *et al*., 2011). Therefore, biosensors based on P450s could be used in the multiple phases of the drug discovery procedure to increase sample throughput and reduce costs simultaneously producing substantial new information about potential drug candidates, as the biosensors could be designed to be more sensitive and specific for the detection of drug molecules (Myszka and Rich, 2000; Rouse and Hardiman, 2003). Such an example are biosensors that have been developed using mammalian P450s namely; CYP1A2, CYP2B4 and CYP11A for the detection of drugs (clozapine), xenobiotics (styrene) and steroids (cholesterol) (Paternolli *et al*., 2004). In an interesting study, different P450 isoforms were used to detect drug compounds using nanoparticles or carbon nanotubes for enhancing the device sensitivity so as to target therapeutic ranges in the serum of patients, whereby the drug concentration was indirectly measured *via* a reduction potential obtained with the cyclic voltammetry technique (Baj-Rossi *et al*., 2011). The remarkable abilities of P450s are not limited to drug synthesis alone. On the contrary P450s have been introduced in bioremediation for the removal of toxic and synthetic compounds from the environment. CYP101 of *P. putida* capable of hydroxylating camphor was engineered using site-directed mutagenesis (Poulos *et al*., 1987; Bernhardt, 2006). Various mutations on the known three-dimensional structure to induced and re-orientate the stereochemistry have been reported, such as mutations in F87 and Y96 that optimized the activity of CYP101 for the oxidation of polycyclic aromatic hydrocarbons such as phenanthrene, fluoranthene and pyrene, as well as polychlorinated benzenes

(Urlacher *et al*., 2004; Benhardt, 2006). Strategic bio-engineering of both CYP101 and CYP102 have been reported, whereby the substrate diversity catalyzed by these enzymes have been increased for the purpose of bioremediation of environmental pollutants (Harford-Cross *et al*., 2000; Carmichael and Wong, 2001). XplA (Corynebacterineae: *Rhodococcus* sp.) is another unique example of a P450 used in bioremediation (Rylott *et al*., 2011a). XplA is involved in the microbial biodegradation of the synthetic military explosive and pollutant RDX on contaminated soil (Rylott *et al*., 2011(a, b)).

For the past few decades scientists have been striving to find renewable sources of energy that will reduce world petroleum exploitation as well as decrease carbon dioxide emissions, which are also the main cause of global warming (Zhang *et al*., 2011). Biofuel production using P450s is a promising area of research and a potential solution against the struggle for finding biodegradable, sustainable and renewable biofuels. Mutated P450, CYP153A6 from *Mycobacterium* sp. HXN-1500 is capable of terminal alkane hydroxylation of alkanes to 1-alkanols such as the biotransformation of butane to 1-butanol (Funhoff *et al*., 2006). In contrast, OleTJE an engineered P450, from the CYP152 family from the bacteria *Jeotgalicoccus*, is able to decarboxylate and to hydroxylate fatty acids (Rude *et al*., 2011).

1.3. Thermostable P450s

1.3.1. Introduction to thermostable P450s

There exists a unique and elite group of remarkable organisms that thrive in deep-sea hydrothermal vents, hot springs and solfataric areas called extremophiles that have adapted to harsh enviromental conditions through the gradual process of evolution and natural selection (Demirjian *et al*., 2001). Extremophiles consist of various classes such as thermophiles, acidophiles and peizophiles (barophiles) to mention a few (Table 1.2) (Demirjian *et al*., 2001). Thermostable enzymes are categorized into three respective groups based on the

property of temperature stability of these enzymes namely: low temperature (35-45°C), moderate thermophiles (temperature range: 45-65°C), thermophiles (65-85°C), and hyperthermophiles (>85°C) (Demirjian *et al*., 2001). Thermostable enzymes are found in thermophilic organisms and are proteins that are able to resist high temperatures, without undergoing denaturation as the structural and chemical properties of these proteins are not affected.

In the past, the need to culture thermophilic microorganisms for research purposes such as isolation of genetic material, proteins or even for observation was quite challenging, laborious, time-consuming and the microbial cultures did not survive long enough (Kushnar, 1978, Rothschild and Mancinelli, 2001). Today, with the help of genomic tools, it has become easier to study thermophilic microorganisms. Extremophilic/thermophilic P450s were the first proteins to be unpredictably discovered by genomics in the species *Sulfolobus solfataricus* (CYP119) (Wright *et al*., 1996; Yano *et al*., 2000). In this section, focus will be on thermophilic enzymes namely thermostable P450s, which are important not only because of their bio catalytic activities, but because they possess a higher magnitude of thermotolerance compared to other P450s (Matsumura *et al*., 2008).

Table 1.2. Examples of extremophiles and the industrial applications of their isolated enzymes. (Taken from Demirjian *et al*., 2001; Van den Burg, 2003).

Extremophile	Habitat	Enzymes	Applications
Thermophile	High temperature (35- 45°C) Moderate thermophiles (45-65°C) Thermophiles (65-85°C) Hyperthermophiles (<85°C)	Amylase Xylanases Proteases DNA polymerases	Glucose and fructose for sweeteners Paper bleaching Baking, brewing, detergents Molecular biology (PCR), genetic engineering
Psychrophile	Low temperature (<15°C)	Proteases Dehydrogenases Amylases	Cheese maturation, dairy production Biosensors Polymer degradation in detergents
Acidophile	Low pH (pH<2-3)	Sulfur oxidation Chalcopyrite concentrate	Desulfurization of coal Valuable metals recovery
Alkalophile	High pH (pH>9)	Cellulases	Polymer degradation in detergents, Ion exchange resin regenerant disposal
Halophile	High salt concentration	Proteases	Peptide synthesis
Piezophile	High pressure (≤130MPa)	Whole micro-organism	Formation of gels and starch granules, Food processing and antibiotic production
Metalophile	High metal concentration	Whole micro-organism	Ore-bioleaching, bioremediation, biomineralization
Radiophile	High radiation levels	Whole micro-organism	Bioremediation of radionuclide contaminated sites
Microaerophile	Growth in<21% O_2	To be defined	To be defined

As previously mentioned, P450s have immense catalytic activities and varied substrate recognition, which have encouraged the utilization of these enzymes as prospective

biocatalysts for the production of chemicals, pharmaceutical substances, antibiotics, fragrances, food flavourings and the detoxification of carcinogens/mutagens (Ingelman-Sundberg, 2004; Guengerich, 2006, Urlacher and Eiben, 2006; Syed *et al*., 2014a). The criteria for biotechnological exploitation of P450s on an industrial scale requires that these enzymes express resistance to extreme industrial reaction conditions such as high pressure, high temperature, pH and organic solvents, in doing so the rate of reaction would be increased, reaction specificity could be controlled and the chances of contamination by micro-organisms could be reduced (Niehaus, 1999; Zeikus *et al*., 1998, Yano and Poulos, 2003). However, due to complexity, lack of thermal and chemical stability as well as the limited substrate specificity of P450 enzymes, usage of these enzymes for industrial and biotechnological purposes is hindered (O'Reilly *et al*., 2011). Moreover, thermostable P450s used at industrial conditions should not only be stable *in vivo*, but also *in vitro*, if they are to be used in future for biotechnological purposes. In view of the fact that P450s are structurally and chemically less stable at extreme conditions, research has focused on discovering and identifying novel thermostable P450s that exhibit thermostability (Nishida and Ortiz de Montellano, 2005). Thus, thermophilic P450 enzymes are of potential interest and can serve as potential biocatalysts in terms of their protein structure, catalytic mechanisms and biotechnological/industrial applications such as the biosynthesis of important organic metabolites or intermediates (Nishida and Ortiz de Montellano, 2005; Bernhardt, 2006).

Currently, only a few thermostable P450s have been described in the literature and have been characterized, whereby most of them are of archaeal [CYP119, CYP119A2 (P450st) and CYP231A2] (Wright *et al*., 1996; Koo *et al*., 2000; Kawarabayasi *et al*. 2001; Ho *et al*., 2008) and bacterial origin [CYP175A1] (Yano *et al*., 2003). These extremophiles have ecological niches in geothermal vents and areas of high solfataric activity (Wright *et al*., 1996; Ho *et al*., 2008). CYP119 was first identified in the archael species *S. solfataricus*

(Wright *et al*., 1996). CYP119 has a melting temperature (T_m) of 91°C and the substrates that this enzyme binds to specifically are lauric acid, *cis*- and *trans*-*β*-methylstyrenes and styrene (Koo *et al*., 2000). CYP119A2 or P450st was first identified in the archael species *S. tokodaii* strain 7 (Kawarabayasi *et al*., 2001). The T_m of P450st is yet to be determined and the substrates that this enzyme binds to specifically are lauric acid and styrene (Matsumura *et al*., 2011; Matsumura *et al*., 2008; Kawarabayasi *et al*., 2001). CYP231A2 was identified in the archael species *Picrophilus torridus* and has a T_m of 65°C (Ho *et al*., 2008). CYP175A1 found in the bacterial species *Thermus thermophilus* HB27 and has a T_m of 88°C (Mandai *et al*., 2009a; Yano *et al*., 2003). CYP175A1 binds specifically to *β*-carotene, however with bioengineering attempts of the 175RF domain and introduction of one mutant (Q67G/Y68I), CYP175A1 is now capable of hydroxylating and accepting testosterone as a substrate (Mandai *et al*., 2009b).

1.3.2. Unravelling the lower eukaryote fungal genomes for thermostable P450s

To date, no thermostable P450s from other biological kingdoms have been reported. Until, recently a study conducted by Syed *et al*. (2014b), unravelled a large number of thermostable P450s of biotechnological potential from the two thermophilic ascomycetes, *Myceliophthora thermophila* (*Sporotrichum thermophile*) (abbreviated as *M. thermophila*) and *Thielavia terrestris* (abbreviated as *T. terrestris*). In the study, comparative genome-wide P450 analysis was used to "mine" for P450s exhibiting both the P450 signature domain motifs namely EXXR (heme-binding) and FXXGXXXCXG (oxygen-binding), whereby 50 families and 56 subfamilies (*T. terrestris*) and 49 families and 53 subfamilies (*M. thermophila*) were identified (Table 1.3) (Syed *et al*., 2014b). Out of 70 and 79 P450s reported, only 14 and 11

P450s were considered as thermostable based on an aliphatic index cut-off of > 90, in *M. thermophila* and *T. terrestris* respectively (Figure 1.4 and Table 1.4) (Syed *et al.*, 2014b). Among the P450s identified as exhibiting thermotolerance in *T. terrestris*, is CYP53A, where the melting temperature of CYP53A was predicted to be 55-65°C (Syed *et al*., 2014b). CYP53A belongs to the CYP53 family known as benzoate *para*-hydroxylase, an enzyme capable of hydroxylating benzoate (Faber *et al*., 2001).

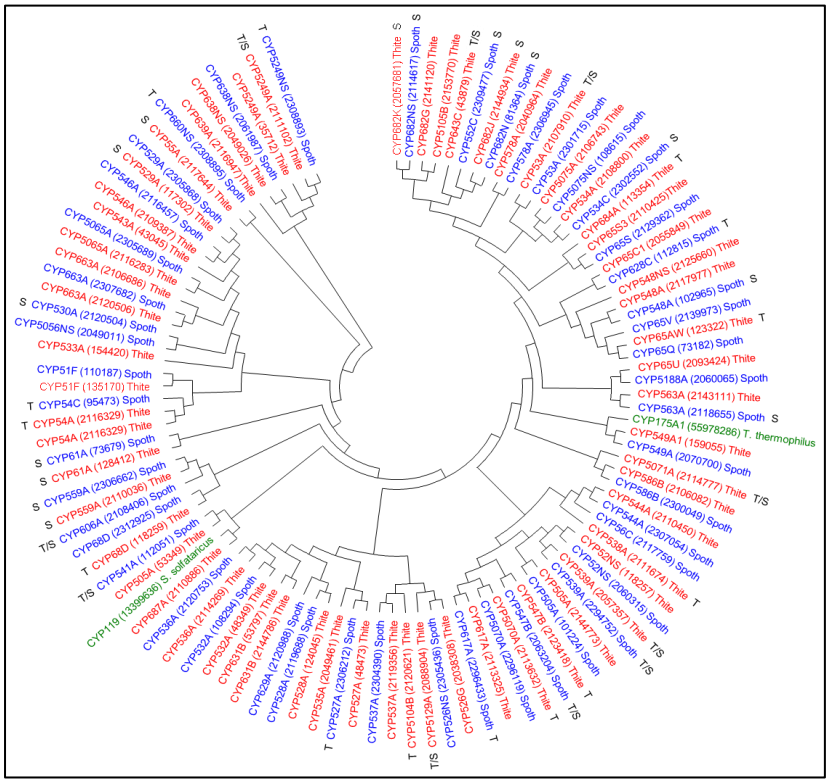

Figure 1.4. Phylogenetic tree of the P450ome of *T. terrestris* (*Thite*-red) *and M. thermophila* (*Spoth*-blue). A total of 108 sequences were analysed, whereby thermostable P450s CYP119 and CYP175A1 of archael and bacterial origin (green), were also included for evolutionary analysis. Protein IDs are presented in parenthesis, with the symbols S and T, next to the P450s depicting stability of the P450 in vitro and the thermostability respectively. (Taken from Syed *et al*., 2014b).

Table 1.3. Genome-wide comparative analysis of the P450omes of thermophilic and mesophilic ascomycetes. P450omes of thermophilic ascomycetes *T. terrestris* (TT) and *M. thermophila* (MT) were compared with P450omes of the 14 mesophilic ascomycete species, whereby the number of P450s present is indicated. *Mycosphaerella fijiensis* (MF), *Uncinocarpus reesii* (UR), *Histoplasma capsulatum* (HC), *Coccidioides immitis* (CI), *Aspergillus clavatus* (AC), *Aspergillus niger* (AN), *Aspergillus flavus* (AF), *Aspergillus oryzae* (AO), *Aspergillus terreus* (AT), *Aspergillus fumigatus* (AFu), *Neurospora crassa* (NC), *Neurospora discrete* (ND), *Fusarium graminearum* (FG) and *Fusarium oxysporum* (FO). (Taken from Syed *et al*., 2014b).

P450 family	Thermophilic Pezizomycetes		Mesophilic Pezizomycetes													
	TT	MT	MF	UR	HC	CI	AC	AN	AF	AO	AT	AFu	NC	ND	FG	FO
CYP51	1	1	1	2	2	1	2	2	4	3	3	2	1	1	3	3
CYP52	1	1	1	1			2	5	4	5	4	3				
CYP53	1	1	1	1		1	1	1	1	2	1	1	1	1	3	3
CYP54	2	1											1	1	1	1
CYP55	1			1	1			1	1		1		1	1	1	4
CYP56		1		2		1	1		1	1						
CYP61	1	1	1	1	1	1	2	2	2	2	3	1		1	2	2
CYP65	4	3	2		7	1	5	14	9	10	8	3	2	2	4	2
CYP68	1	1			2		1	3	1	2	3	1	1	1	5	2
CYP505	2	1	3				2	3	4	3	2	1	1	1	2	4
CYP526	1	1									1		1	1	2	1
CYP527	1	2						1					2	2	1	3
CYP528	1	1	1										1	2	1	

CYP529	1	1											1	1		
CYP530		1		1		1					1	1	1	1	1	1
CYP532	1	1	1					2	1	1	1		1	1	4	3
CYP533	1	1											1	1		
CYP534	1	1	1										1	1	1	
CYP535	1								1	1			1	1		
CYP536	1	1	1										1	1		
CYP537	1	1							1	1	1		1		1	2
CYP538	1												1	1		
CYP539	1	1	1	1	2	1	2	1	1		3	2	1	1	2	1
CYP541		1					1	1	1	1	1	1	1	1		
CYP543	1	1	1										1	1		
CYP544	1	1											1	1	1	1
CYP546	1	1											1	1	1	
CYP547	1	1	1					1	1	1	2	1	1	1	1	1
CYP548	2	1	3	2	1	2	4	6	2	3	3	1	1	1	2	6
CYP549	1	1											1	1		
CYP552		1							1		1		1	1	3	3
CYP559	1	1	2										1	1		1
CYP563	1	1														
CYP570		1	2	1		1									3	5
CYP578	1	1	3	2	1	2	2	3	3	2	2	2				2
CYP586	1	1			1			2			1	1				
CYP606		1					1				1				1	1
CYP617	1	1	1	1			5	3	1	1	1	3			3	3
CYP628		1			1			1	1	1	1				1	4
CYP629		1			1										1	1
CYP631	2								2	1					1	
CYP634	1														1	1

CYP638	1	1			1										1	1
CYP639	1														1	1
CYP643	1							1	1						4	1
CYP660		1				1	1	1	2	2	1	1				
CYP663	2	1	1				1	1	1		1					
CYP682	3	2	1	4	1	6	2	2	3	2	2	1				1
CYP684	1		1					2	1	1	3					1
CYP687	1								1							
CYP5056		1														
CYP5065	1	1														1
CYP5070	1	1							1	1						
CYP5071	1															
CYP5075	1	1						1	2	2						
CYP5080		1														
CYP5104	1					1		2			1					
CYP5105	1							1	1							
CYP5129	1															
CYP5188		1	1													
CYP5249	2	1			1											
CYP5272	1						1							1		
CYP6001		1	1		1	1	2	2	3		3			1		1
Total no. of P450s in the genome	**61**	**53**	**89**	**38**	**47**	**40**	**91**	**154**	**162**	**142**	**124**	**74**	**41**	**43**	**109**	**140**

Table 1.4. Analysis of thermostability of the P450omes of *T. terrestris* and *M. thermophila*. The aliphatic index and protein melting temperature (T_m) were used to measure the thermostablity of each P450. P450s having analiphatic index of > 90, implies that the P450 is thermostable. The aliphatic index of each thermostable P450 is compared with their homologous P450s from mesophilic ascomycetes (presented as reference species). Ascomycete thermostable P450s' aliphatic index and T_m were also compared with thermostable P450s CYP119 of *S. Solfataricus* and CYP175A1 of *T. thermophilus*. Abbreviations: PN, *Phaeosphaeria nodorum*; HC, *Histoplasma capsulatum*; AN, *Aspergillus niger*; AF, *Aspergillus flavus*; AT, *Aspergillus terreus*; NC, *Neurospora crassa*; ND, *Neurospora discrete*; FG, *Fusarium graminearum*; FO, *Fusarium oxysporum*; NS, new subfamily. (Taken from Syed *et al*., 2014b).

P450	**Aliphatic index**	**Melting temperature (Tm)**		**Homolog P450 from mesophilic ascomycetes**	
		Tm Index	Predicted Tm (°C)	Aliphatic index	P450 and species name
T. terrestris					
CYP538A	92.67	-0.1	<55	87.72	CYP538A1 NC
CYP539A	91.17	0.6	55-65	86.63	CYP539A8P AT
CYP643C	93.35	1.4	>65	89.09	CYP643C2 AF
CYP684A	92.63	0.7	55-65	83.71	CYP684A4 AT
CYP5070A	90.83	0.1	55-65	83.37	CYP5070A1 AF
CYP5129A	100.43	1.21	>65	96.08	CYP5129A PN
CYP65AW	92.64	0.78	55-65	88.00	CYP65AW1 FO
CYP53A	95.1	0.8	55-65	85.13	CYP53A1 AN
CYP5249A	91.09	1.17	>65	98.91	CYP5249A HC
CYP5104B	91.33	-	<55	86.22	CYP5104B1 AN
CYP68D	92.31	0.5	55-65	88.31	CYP68D3 AN
CYP5071A	96.83	1.2	>65	88.53	CYP5071A FG
CYP547B	102.34	1.61	>65	100.8	CYP547B3 FO
CYP54A	94.57	1.1	>65	86.05	CYP54A1 ND

M. thermophila					
CYP628C	93.3	0.5	55-65	89.66	CYP628C1 AF
CYP617A	91.12	0.5	55-65	85.34	CYP617A3 FO
CYP547B	101.3	1.07	>65	78.66	CYP547A5 FO
CYP539A	92.33	0.8	55-65	86.63	CYP539A8P AT
CYP5249NS	92.78	1.5	>65	99.84	CYP5249A1P HC
CYP541A	91.88	1.0	55-65	92.89	CYP541A1 NC
CYP606A	97.49	0.7	55-65	86.23	CYP606A3P AT
CYP527A	93.25	0.7	55-65	81.03	CYP527A1 ND
CYP54C	98.56	0.96	55-65	86.88	CYP54C5 FO
CYP660NS	92.21	0.7	55-65	88.24	CYP660A4 AT
CYP61A	96.74	1.3	>65	83.01	CYP61A1 AT
T. thermophilus					
CYP175A1	97.20	1.39	>65		
S. Solfataricus					
CYP119*	96.63	0.94	55-65		

1.4. Rationale and objectives of the study

Cytochrome P450 monooxygenases (P450s) are heme-thiolate proteins distributed across biological kingdoms (Nelson, 2013). P450s perform a wide variety of reactions, such as activation of sp^3 hybridized carbon (C) atoms, epoxidation, deamination and dehalogenation, aromatic hydroxylation and N-oxidation, as well as *N*-, *O*- and *S*-dealkylation, which suggests that these enzymes are capable of accepting diverse substrates (Bernhardt, 2006).

P450s have immense catalytic activities and varied substrate recognition, which have encouraged the utilization of these enzymes as prospective biocatalysts for the production of chemicals, pharmaceutical substances, antibiotics, fragrances, food flavourings and the detoxification of carcinogens/mutagens (Guengerich, 2002; Ingelman-Sundberg, 2004; Guengerich, 2006; Urlacher and Eiben 2006; Syed and Yadav, 2012). The criteria for biotechnological exploitation of P450s on an industrial scale requires that these enzymes express resistance to extreme industrial reaction conditions such as high pressure, high temperature, pH and organic solvents (Niehaus, 1999; Zeikus *et al*., 1998; Yano and Poulos, 2003). However, due to thermal and chemical instability as well as the limited substrate specificity of P450 enzymes, usage of these enzymes for industrial and biotechnological purposes is hindered (O'Reilly *et al*., 2011). Moreover, thermostable P450s used at industrial scale should not only be stable *in vivo*, but also *in vitro*. Considering the fact, that P450s are structurally and chemically less stable at extreme conditions, research has focused on exploring and identifying novel thermostable P450s (Nishida and Ortiz de Montellano, 2005).

The rationale for carrying out this study lies in the well documented findings in the literature that to date only a few thermostable cytochrome P450 monooxygenases have been described, whereby most of them are of archaeal [CYP119, CYP119A2 (P450st) and CYP231A2] (Wright *et al*., 1996; Koo *et al*., 2000; Kawarabayasi *et al*. 2001; Ho *et al*.,

2008) and bacterial origin [CYP175A1] (Yano *et al*., 2003). The best and most studied thermostable P450 to date is CYP119 from *S. solfataricus* (Wright *et al*., 1996; Koo *et al*., 2000). Hence, it can be observed from the literature that these bacterial and archael thermostable P450s show limited substrate specificity, that is, these P450s can only bind to a few or just one substrate molecule in their respective native configuration. Moreover, archael and bacterial thermostable P450s cannot generate human (eukaryotic) valuable products that are useful for commercialization. No thermostable P450s from other biological kingdoms or the eukaryotic domain have been reported. Until, recently a study conducted by Syed *et al.* (2014), unravelled a large number of thermostable P450s of biotechnological potential from two thermophilic biomass-degrading ascomycetes, *M. thermophila* and *T. terrestris*. Among the P450s identified as exhibiting thermotolerance in *T. terrestris*, is CYP53A, where the melting temperature of CYP53A was predicted to be 55-65°C (Syed *et al*., 2014). CYP53A belongs to the CYP53 family known as benzoate *para*-hydroxylase, an enzyme capable of hydroxylating benzoate (Faber *et al*., 2001).

In this study, the distribution of CYP53 in fungal species was assessed. Genome-data mining was performed to "fish out" CYP53 from four fungal phyla (Ascomycota, Basidiomycota, Chytridiomycota and Zygomycota). Subfamilies were assigned to identified CYP53 proteins in the fungal phyla. The main goal of analysis was to identify whether CYP53 can serve as novel common anti-fungal drug target. Structural analysis of CYP53A and CPR was performed to assess whether these P450s possess characteristic signature motifs of P450 and redox enzymes. Three dimensional (3D) models for CYP53A and CPR was constructed to locate specific amino acids that are highly conserved that might be used as antifungal drug target site for future studies (in case of CYP53A) and analysis of cofactor binding properties (in case of CPR). Consequently, it could be determined, whether this

CYP53 family could be used to unravel the molecular basis for thermostability of eukaryote P450s.

REFERENCES

Baj-Rossi, C., De Micheli, G., Carrara, S. (2011). *P-450-Based nano-biosensors for personalized medicine*. In: Serra, A (ed.). *Biosensors for Health, Environment and Biosecurity*. Vienna, Austria: InTech Publisher. p.448-482.

Bernhardt, R. (1996). Cytochrome P450: structure, function, and generation of reactive oxygen species. Springer, Berlin, Heidelberg. *In: Reviews of Physiology Biochemistry and Pharmacology*, 127: 137-221.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *J.Biotechnol*., 124(1): 128-145.

Black, S.D., Martin, S.T., Smith, C.A. (1994). Membrane topology of liver microsomal cytochrome P450 2B4 determined via monoclonal antibodies directed to the halt-transfer signal. *Biochemistry*, 33(22): 6945–6951.

Carmichael, A. B., Wong, L. L. (2001). Protein engineering of *Bacillus megaterium* CYP102.The oxidation of polycyclic aromatic hydrocarbons. *European Journal of Biochemistry*, 268(10): 3117-3125.

Cojocaru, V., Balali-Mood, K., Sansom, M. S., Wade, R. C. (2011).Structure and dynamics of the membrane-bound cytochrome P450 2C9.*PLoS computational biology*, 7(8): e1002152.

Cupp-Vickery, J. R., Poulos, T. L. (1995). Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Structural & Molecular Biology*,2(2), 144-153.

David, J. P., Ismail, H. M., Chandor-Proust, A., Paine, M. J. I. (2013).Role of cytochrome P450s in insecticide resistance: impact on the control of mosquito-borne diseases and use of

insecticides on Earth. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612): 20120429.

De Lemos-Chiarandini, C., Frey, A.B., Sabatini, D.D., Kreibich, G. (1987). Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies. The *Journal of cell biology*, 104(2): 209–219.

Demirjian, D. C., Morís-Varas, F., Cassidy, C. S. (2001). Enzymes from extremophiles. *Current opinion in chemical biology*, 5(2): 144-151.

Denisov, I. G., Makris T. M., Sligar , S. G., Schlichting, I. (2005). Structure and Chemistry of Cytochrome P450. *Chem. Rev.*, 105: 2253-2277.

Faber BW, van Gorcom RFM, Duine JA. (2001). Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. *Archives of biochemistry and biophysics*, 394(2): 245–254.

Funhoff, E. G., Bauer, U., García-Rubio, I., Witholt, B., van Beilen, J. B. (2006).CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. *Journal of bacteriology*, 188(14): 5220-5227.

Grogan, G. (2011). Cytochromes P450: exploiting diversity and enabling application as biocatalysts. *Curr.Opin.Chem.Biol*., 15(2): 241-248.

Guengerich, F. P. (2002). Cytochrome P450 enzymes in the generation of commercial products. *Nature Reviews Drug Discovery*, 1(5): 359-366.

Guengerich, F. P. (2006). Cytochrome P450s and other enzymes in drug metabolism and toxicity. *The AAPS journal*, 8(1): E101-E111.

Guengerich, F. P., Martin, M. V., Sohl, C. D., Cheng, Q. (2009). Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat.Protoc.*, 4(9): 1245-1251.

Hannemann, F., Bichet, A., Ewen, K. M., Bernhardt, R. (2007). Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1770(3): 330-344.

Harford-Cross, C. F., Carmichael, A. B., Allan, F. K., England, P. A., Rouch, D. A., Wong, L. L. (2000). Protein engineering of cytochrome P450cam (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. *Protein engineering*, 13(2), 121-128.

Herper, M. (2013). Forbes, The cost of creating a new drug now $5 billion, pushing big pharma to change. Available at: http://www.forbes.com/sites/matthewherper/2013/08/11/how-the-staggering-cost-of-inventing-new-drugs-is-shaping-the-future-of-medicine/. [Accessed: 15 September 2014].

Ho, W. W., Li, H., Nishida, C. R., Ortiz de Montellano, P. R., Poulos, T. L. (2008). Crystal structure and properties of CYP231A2 from the thermoacidophilic archaeon *Picrophilus torridus*. *Biochemistry*, 47(7): 2071-2079.

Hogg, J. A. (1992). Steroids, the steroid community, and Upjohn in perspective: a profile of innovation. *Steroids*, 57(12): 593-616.

Holton, T.A., Brugliera, F., Lester, D.R., Tanaka., Y., Hyland, C.D., Menting, J.G., Lu, C.Y., Fancy, E., Stevenson, T.W., Cornish, E.C. (1993). Cloning and expression of cytochrome P450 genes controlling flower colour. *Nature*, 366: 276-279.

Ingelman-Sundberg, M. (2004). Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends in pharmacological sciences*, 25(4): 193-200.

Isin, E. M., Guengerich, F. P. (2007). Complex reactions catalyzed by cytochrome P450 enzymes. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1770(3), 314-329.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T., Kikuchi, H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7.*DNA Research* 8: 123 – 140.

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Arch. Biochem. Biophys*, 75(2): 376–386 .

Klingenberg, M. (2003).Pigments of rat liver microsomes. *Archives of biochemistry and biophysics*, 409(1): 2-6.

Koo, L. S., Tschirret-Guth, R. A., Straub, W. E., Moënne-Loccoz, P., Loehr, T. M., de Montellano, P. R. O. (2000). The active site of the thermophilic CYP119 from *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 275(19): 14112-14123.

Kushner, D.J (ed.). (1978). Microbial Life in Extreme Environments. London, New York: Academic Press. p. 465.

Lah, L., Kraševec, N., Trontelj, P., Komel, R. (2008). High diversity and complex evolution of fungal cytochrome P450 reductase: cytochrome P450 systems. *Fungal Genetics and Biology*, 45(4): 446-458.

Lederer, F., Ghrir, R., Guiard, B., Cortial, S., Ito, A. (1983). Two homologous cytochromes b5 in a single cell. *European Journal of Biochemistry*, 132(1): 95–102.

Luthra, A., Denisov, I. G., Sligar, S. G. (2011). Spectroscopic features of cytochrome P450 reaction intermediates. *Archives of biochemistry and biophysics*, 507(1): 26-35.

Mandai, T., Fujiwara, S., Imaoka, S. (2009a).A novel electron transport system for thermostable CYP175A1 from *Thermus thermophilus* HB27.*FEBS Journal*, 276(8): 2416 – 2429.

Mandai, T., Fujiwara, S., Imaoka, S. (2009b). Construction and engineering of a thermostable self-sufficient cytochrome P450. *Biochemical and Biophysical Research Communications*, 384(1): 61 – 65.

Matsumura, H., Matsuda, K., Nakamura, N., Ohtaki, A., Yoshida, H., Kamitori, S., Yohda M. Ohno, H. (2011). Monooxygenation by a thermophilic cytochrome P450 via direct electron donation from NADH. *Metallomics*, 3 (4): 389 – 395.

Matsumura, H., Wakatabi, M., Omi, S., Ohtaki, A., Nakamura, N., Yohda, M., Ohno, H. (2008). Modulation of redox potential and alteration in reactivity via the peroxide shunt pathway by mutation of cytochrome P450 around the proximal heme-ligand. *Biochemistry*, 47(16): 4834-4842.

McLean, K.J., Sabri, M., Marshall, K.R., Lawson, R.J., Lewis, D.G., Clift, D., Balding, P.R., Dunford, A.J., Warman, A.J., McVey, J.P., Quinn , A.M., Sutcliffe, M.J., Scrutton, N.S., Munro, A.W. (2005). Biodiversity of cytochrome P450 redox systems. *Biochemical Society Transactions*, 33(4): 796–801.

Meunier, B., de Visser, S.P., Shaik, S. (2004).Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev*, 104(9): 3947-3980.

Miners, J.O. (2002). Evolution of drug metabolism: hitchhiking the technology bandwagon. *Clinical and experimental pharmacology and physiology*, 29(11): 1040-1044.

Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J., McLean, K. J. (2013). What makes a P450 tick?. *Trends in biochemical sciences*, 38(3): 140-150.

Myszka, D. G., Rich, R. L. (2000). Implementing surface plasmon resonance biosensors in drug discovery. *Pharmaceutical science & technology today*, 3(9): 310-317.

Nebert, D.W., Adesnik , M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Geungerich, F.P., Gunsaius., I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., and Sato, R., Waterman, M.R. (1987). The P450 gene superfamily: recommended nomenclature. *DNA*, 6(1): 1–11.

Nebert, D. W.; Nelson, D. R.; Coon, M. J.; Estabrook, R. W.; Feyereisen, R.; Fujii-Kuriyama, Y.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Loper, J. C.; Sato, R.; Waterman, M. R.; Waxman, D. J. (1991). The P450 superfamily: Update on new sequences, gene mapping and recommended nomenclature. *DNA Cell Biol.*, 10(1): 1–14.

Nelson, D. R. (1998a). Cytochrome P450 nomenclature. *Methods Mol. Cell.Biol*, 107:15–24.

Nelson, D. R. (1998b). Metazoan cytochrome P450 evolution. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, 121(1): 15-22.

Nelson, D. R. (1999). Cytochrome P450 and the individuality of species. *Arch.Biochem. Biophys*, 369: 1–10.

Nelson, D. R. (2006). Cytochrome P450 Nomenclature, 2004.In Cytochrome P450 Protocols. *Humana Press*, 320: 1-10.

Nelson, D. R. (2009). The Cytochrome P450 Homepage. *Human Genomics*, 4(1): 59-65.

Nelson, D. R. (2011). Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(1): 14-18.

Nelson, D. R. (2013). A world of cytochrome P450s. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612): 20120430.

Niehaus, F., Bertoldo, C., Kähler, M., Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Applied microbiology and biotechnology*, 51(6), 711-729.

Nishida, C. R., Ortiz de Montellano, P. R. (2005). Thermophilic cytochrome P450 enzymes. *Biochemical and biophysical research communications*, 338(1): 437-445.

Ogata, J., Kanno, Y., Itoh, Y., Tsugawa, H., Suzuki, M. 2005. Plant biochemistry: anthocyanin biosynthesis in roses. *Nature*, 435: 757-758.

Omura, T., Sato, R. (1962). A new cytochrome in liver microsomes. *Journal of Biological Chemistry*, 237(4): PC1375-PC1376.

Omura, T., Sato, R. (1964). The carbon monoxide-binding pigment of liver microsomes: I. Evidence for its hemoprotein nature. *J. biol. Chem*, 239(7): 2370-2378.

O'Reilly, E., Köhler, V., Flitsch, S. L., Turner, N. J. (2011). Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chemical Communications*, 47(9): 2490-2501.

Park, S. Y., Yamane, K., Adachi, S. I., Shiro, Y., Weiss, K. E., Maves, S. A., Sligar, S. G. (2002). Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus*: high resolution structure and functional properties. *Journal of inorganic biochemistry*, 91(4): 491-501.

Paternolli, C., Antonini, M., Ghisellini, P., Nicolini, C. (2004).Recombinant cytochrome P450 immobilization for biosensor applications. *Langmuir*, 20(26): 11706-11712.

Peterson, D. H., Murray, H. C., Eppstein, S. H., Reineke, L. M., Weintraub, A., Meister, P. D., Leigh, H. M. (1952). Microbiological Transformations of Steroids.1 I. Introduction of Oxygen at Carbon-11 of Progesterone. *Journal of the American Chemical Society*, 74(23): 5933-5936.

Poulos, T. L., Finzel, B. C., Howard, A. J. (1987). High-resolution crystal structure of cytochrome P450cam.*Journal of molecular biology*, 195(3): 687-700.

Roberts, G. A., Grogan, G., Greter, A., Flitsch, S. L., Turner, N. J. (2002). Identification of a New Class of Cytochrome P450 from a *Rhodococcus sp*. *Journal of bacteriology*, 184(14): 3898-3908.

Rothschild, L. J., Mancinelli, R. L. (2001). Life in extreme environments. *Nature*, 409(6823): 1092-1101.

Rouse, R., Hardiman, G. (2003). Microarray technology-an intellectual property retrospective. *Pharmacogenomics*, 4(5): 623-632.

Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Applied and environmental microbiology*, 77(5): 1718-1727.

Rylott, E.L., Jackson, R.G., Sabbadin, F., Seth-Smith, H.M.B., Edwards, J., Chong, C.S., Strand, S.E., Grogan, G., Bruce, N.C. (2011a). The explosive-degrading cytochrome P450 XplA: biochemistry, structural features and prospects for bioremediation. *Biochem Biophys Acta*, 1814: 230–236.

Rylott, E. L., Lorenz, A., Bruce, N. C. (2011b). Biodegradation and biotransformation of explosives. *Current opinion in biotechnology*, 22(3): 434-440.

Sen, K., Thiel, W. (2014). Role of Two Alternate Water Networks in Compound I Formation in P450eryF. *The Journal of Physical Chemistry B*, 118(11): 2810-2820.

Sono, M., Roach, M. P., Coulter, E. D., Dawson, J. H. (1996). Heme-containing oxygenases. *Chemical Reviews*, 96(7): 2841-2888.

Syed, K., Shale, K., Pagadala, N. S., Tuszynski, J. (2014a). Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PloS one*, 9(1), e86683.

Syed, K., Shale, K., Nazir, K. N. H., Krasevec, N., Mashele, S. S., Pagadala, N. S. (2014b). 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): 321-333.

Syed, K., Yadav, J. S. (2012). P450 monooxygenases (P450ome) of the model white rot fungus Phanerochaete chrysosporium. *Critical reviews in microbiology*, 38(4): 339-363.

Urlacher, V. B., Eiben, S. (2006). Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends in biotechnology*, 24(7): 324-330.

Urlacher, V. B., Girhard, M. (2012). Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in biotechnology*, 30(1): 26-36.

Urlacher, V. B., Lutz-Wahl, S., Schmid, R. D. (2004). Microbial P450 enzymes in biotechnology. *Applied microbiology and biotechnology*, 64(3): 317-325.

Van Den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Current opinion in microbiology*, 6(3): 213-218.

van Beilen, J. B., Duetz, W. A., Schmid, A., Witholt, B. (2003). Practical issues in the application of oxygenases. *Trends in biotechnology*, 21(4): 170-177.

van Beilen, J.B., Holtackers, R., Lüscher, D., Bauer, U., Witholt, B., Duetz, W.A. (2005). Biocatalytic production of perillyl alcohol from limonene by using a novel Mycobacterium sp. cytochrome P450 alkane hydroxylase expressed in *Pseudomonas putida*. *Applied and environmental microbiology*, 71(4): 1737-1744.

Vergeres, G., Waskell, L. (1995). Cytochrome b5, its functions, structure and membrane topology. *Biochemie*, 77(7): 604–620.

Wright, R. L., Harris, K., Solow, B., White, R. H., Kennelly, P. J. (1996). Cloning of a potential cytochrome P450 from the Archaeon *Sulfolobus solfataricus*. *FEBS letters*, 384(3): 235-239.

Yano, J. K., Blasco, F., Li, H., Schmid, R. D., Henne, A., Poulos, T. L. (2003). Preliminary characterization and crystal structure of a thermostable cytochrome P450 from *Thermus thermophilus*. *Journal of Biological Chemistry*, 278(1): 608-616.

Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., Poulos, T. L. (2000). Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 275(40): 31086-31092.

Yano, J. K., Poulos, T. L. (2003). New understandings of thermostable and peizostable enzymes. *Current opinion in biotechnology*, 14(4): 360-365.

Zeikus, J. G., Vieille, C., Savchenko, A. (1998). Thermozymes: biotechnology and structure–function relationships. *Extremophiles*, 2(3): 179-183.

Zhang, F., Rodriguez, S., Keasling, J.D. (2011). Metabolic engineering of microbial pathways for advanced biofuels production. *Current opinion in biotechnology*, 22(6): 775-783.

CHAPTER 2

GENOME-WIDE IDENTIFICATION, ANNOTATION AND COMPARATIVE ANALYSIS OF CYP53A FAMILY IN FUNGI

2.1. Introduction

Among microorganisms fungi, the largest biological kingdom comprising of diverse lower eukaryotic microorganisms, acquired a special place owing to their ability to be pathogens for not only humans but also other animals and plants (Table 2.1). These lower eukaryotes developed or are constantly developing new strategies to adapt to diverse ecological niches. In order to develop novel drugs by identifying potential novel drug targets and harnessing their potentials for the production of human valuables, a large number of fungal genomes have been sequenced and many fungal genome sequencing projects are in progress. Efforts of the Broad Institute of MIT and Harvard (http://www.broadinstitute.org/), Wellcome Trust Sanger Institute (https://www.sanger.ac.uk/) and Joint Genome Institute (JGI) United States Department of Energy (US-DOE) (http://genome.jgi.doe.gov/programs/fungi/index.jsf) resulted in genome sequencing of a large number of fungal species.

Genome sequencing analysis of fungal species revealed the presence of a large number of cytochrome P450 monooxygenases (P450s) in their genomes, with some exceptions. P450s are heme-thiolate proteins ubiquitously present across the biological kingdoms (Nelson, 2013). In fungi P450s are known to be involved in both primary and secondary metabolic processes (Črešnar and Petrič, 2011; Hlavica, 2013) and in the degradation of xenobiotic compounds (Syed and Yadav, 2012). P450s have been explored as anti-fungal drug targets owing to their key role in fungal physiology through involvement in stereo- and regio-specific

oxidation of substrates (Yoshida, 1988). Among fungal P450s CYP51, also known as sterol 14α-demethylase, the highly conserved P450 across the biological kingdoms (Lepesheva and Waterman, 2004), is the primary target of conventional antifungal azole drugs (Kelly and Kelly, 2013). CYP51 performs demethylation of lanosteol, a key step in biosynthesis of cell membrane ergosterol (Lepesheva and Waterman, 2004). Studies have indicated that fungal organisms are developing resistance to azole drugs (Hof, 2001; Sanglard, 2002). Furthermore, the currently available anti-fungal drugs have limitations because of the metabolic pathways similarity between fungi and other organisms (mainly mammals) and hence researchers are in search of alternative novel fungal drug targets (Sangamwar *et al*.,2008).

Table 2.1. Genome-wide comparative analysis of CYP53 family in fungi. Twenty-three species from ascomycota and 28 species from basidiomycota were used in this study. Identification of CYP53 members in fungal species was carried out as described in the "Materials and methods" section. If no CYP53 member was found in the species, the space was left blank. The abbreviation NS indicates a new subfamily. Fungal species capable of causing diseases in humans were indicated with the word "human" in the table.

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Ascomycota								
Magnaporthe grisea	Plant pathogen	1						1
Neurospora crassa	Model organism	1						1

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Ascomycota								
Neurospora discreta	Distantly related to *Neurospora crassa*	1						1
Fusarium graminearum	Plant pathogen	3						3
Fusarium solani f. batatas (Nectria haematococca)	Plant pathogen and animal pathogen (opportunistic human pathogen)	2						2
Fusarium verticillioides	Plant pathogen and animal pathogen (opportunistic human pathogen)	2						2
Fusarium oxysporum	Plant pathogen and animal pathogen (opportunistic human pathogen)	2			1			3
Neosartorya fischeri	Animal pathogen (including human)	1						1
Aspergillus nidulans	Model organism for study of eukaryotic cell biology	1						1
Aspergillus fumigatus	Animal pathogen (opportunistic human pathogen)	1						1
Aspergillus terreus	Human, animal and plant pathogen	1						1
Aspergillus oryzae	Economically important, used for fermentation	2						2
Aspergillus flavus	Plant and animal pathogen (human pathogen)	1						1
Aspergillus niger	Plant and animal pathogen (human pathogen)	1						1

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Ascomycota								
Aspergillus clavatus	Animal pathogen (human pathogen)	1						1
Coccidioides immitis	Animal pathogen (human pathogen)	1						1
Histoplasma capsulatum	Animal pathogen (human pathogen)	0						0
Uncinocarpus reesii	Non-pathogen	1						1
Mycosphaerella fijiensis	Plant pathogen	1						1
Zymoseptoria tritici(formerly named as *Mycosphaerella graminicola*)	Plant pathogen	1						1
Thielavia terrestris	Non-pathogen	1						1
Myceliophthora thermophila	Non-pathogen	1						1
Cochliobolus lunatus	Plant and animal pathogen (human pathogen)	1						1
Total count (Ascomycota)		**28**			**1**			**29**
Basidiomycota								
Phanerochaete chrysosporium	Model white rot fungus – study of wood degradation			1				1
Postia placenta	Model brown rot fungus – study of wood degradation			1	7			8

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Basidiomycota								
Ustilago maydis	Plant pathogen			1				1
Cryptococcus neoformans	Animal pathogen (human)							0
Cryptococcus gattii	Animal pathogen (human)							0
Laccaria bicolor	Symbiotic fungus (ectomycorrhizas)							0
Malassezia globosa	Animal pathogen (human)							0
Puccinia graminis	Plant pathogen		1					1
Sporobolomyces roseus	Non-pathogen		1					1
Phanerochaete carnosa	Model white rot fungus - study of soft wood degradation			6			1	7
Bjerkandera adusta	Wood-degrading white rot fungus			1		7		8
Ceriporiopsis subvermispora	Wood-degrading white rot fungus			4				4
Ganoderma sp.	Wood-degrading white rot fungus			1				1
Ganoderma lucidum	Medicinal mushroom (wood-degrading white rot fungus)			1				1
Phlebia brevispora	Wood-degrading white rot fungus			1				1

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Basidiomycota								
Agaricus bisporus	Litter-degrading fungus			2				2
Serpula lacrymans	Model fungus known as dry rot fungus – study of dry wood degradation			1				1
Stereum hirsutum	Wood-degrading white rot fungus			1				1
Trametes versicolor	Wood-degrading white rot fungus			2				2
Wolfiporia cocos	Wood-degrading brown-rot fungus			9				9
Auricularia delicata	Wood-degrading white rot fungus	1		1				2
Coniophora puteana	Wood-degrading brown rot fungus	1		2				3
Dacryopinax sp.	Wood-degrading brown rot fungus			1				1
Dichomitus squalens	Wood-degrading white rot fungus			1				1
Fomitiporia mediterranea	Wood-degrading white rot fungus			9			1	10
Fomitopsis pinicola	Wood-degrading brown rot fungus			4				4
Gloeophyllum trabeum	Wood-degrading brown rot fungus			1				1
Punctularia strigosozonata	Wood-degrading white rot fungus	1		1				2

Total count (Basidiomycota)	**3**	**2**	**52**	**7**	**7**	**2**	**73**
Total CYP53 members in fungi	**31**	**2**	**52**	**8**	**7**	**2**	**102**

Research on fungal P450s revealed that the P450 family CYP53 could serve as a novel alternative anti-fungal drug target (Podobnik *et al*., 2008). CYP53 family members are well known as benzoate *para*-hydroxylases that are involved in the detoxification of a benzoate molecule (Faber *et al*., 2001). Benzoate is a naturally occurring anti-fungal plant material and a naturally occurring intermediate in the degradation of aromatic compounds in fungi (Durham *et al*., 1984; Jensen *et al*., 1994; Lapadatescu *et al*., 2000; Amborabe *et al*., 2002). Benzoate exhibits its toxicity by disruption of the membrane, inhibiting essential cellular processes, changing pH balance and inducing stress response in fungi (Brul and Coote, 1999; Amborabe *et al*., 2002). CYP53 mediated *para*-hydroxylation of benzoate is the only known pathway in fungi that ultimately channels this toxic compound into the *β*-ketoadipate pathway (Harwood and Parales, 1996). Furthermore, the CYP53 gene was found to be essential for fungal species' survival (Fraser *et al.*, 2002). The CYP53 gene-knock out fungal strain growth was found to be inhibited by the accumulation of toxic intermediate benzoate (Fraser *et al*., 2002). This clearly suggests that this P450 is critical in the survival of fungal species, by playing a key role in the detoxification of benzoate.

Considering the fungal resistance to the currently available drugs, especially CYP51 enzyme-based azoles and a preliminary study suggesting that CYP53 P450 family members can serve as novel alternative fungal drug targets (Sanglard, 2002; Podobnik *et al*. 2008), the present study aimed to understand the role of CYP53 members in fungal physiology, performing comparative evolutionary and structural analysis of CYP53 members to check their distribution and structural conservation in fungi. In this way, one can determine whether

this P450 family can serve as a common drug target against a broad range of fungal pathogens. Furthermore, its role in adaptation of basidiomycetes to diverse ecological niches such as colonization on wood were also assessed.

2.2. Materials and methods

2.2.1. Genome data mining and annotation of CYP53 members

Fifty-one fungal species were selected for the analysis of CYP53 member P450s. As shown in Table 2.1, 23 species from ascomycota and 28 species from basidiomycota were included in this analysis. CYP53 members of the basidiomycete species, such as *Phanerochaete chrysosporium, P. carnosa, Bjerkandera adusta, Ganoderma* sp., *Phlebia brevispora*, and *Ceriporiopsis subvermispora*, and ascomycete species, such as *Thielavia terrestris* and *Myceliophthora thermophila,* were retrieved from published and publicly available literature (Fernandez-Fueyo *et al*., 2012; Floudas *et al*., 2012; Suzuki *et al*., 2012; Syed and Yadav, 2012; Syed *et al*., 2013; Syed *et al*., 2014 (a, b)). CYP53 members in the remaining 20 ascomycetes were obtained from the Cytochrome P450 Homepage (Nelson, 2009). Two basidiomycete species, namely *Agaricus bisporus* and *Serpula lacrymans* CYP53 members, were obtained from the Fungal Cytochrome P450 Database (FCPD) (Moktali *et al*., 2012). CYP53 members belonging to *Postia placenta* were taken from published literature (Ide *et al*., 2012).

To identify CYP53 members in the basidiomycete species, such as *Wolfiporia cocos, Auricularia delicata, Coniophora puteana, Dacryopinax* sp., *Dichomitus squalens, Fomitiporia mediterranea, Fomitopsis pinicola, Gloeophyllum trabeum, Punctularia strigosozonata, Stereum hirsutum,* and *Trametes versicolor,* genome data mining was performed using the established procedure described elsewhere with slight modifications

(Syed *et al*., 2014a;Syed and Mashele, 2014). BLAST analysis was performed at the respective species' genome data base that is publicly available, using *P. chrysosporium* CYP53C2 (protein ID: 130996) (Grigoriev *et al*., 2011). Considering the presence of CYP53 members in low copies (one or two numbers) in ascomycetes and basidiomycetes, the top 20 hit proteins were selected for further analysis. The hit proteins were subjected to the NCBI Batch Web CD-Search Tool to separate proteins belonging to the P450 superfamily (Marchler-Bauer *et al*., 2011). This software groups the proteins into different super families based on the conserved domain characteristics of the protein family. The proteins that are grouped under the P450 superfamily were selected for further assignment to the P450 family and subfamily. Assigning the family and subfamily names to the P450 proteins was performed using the standard procedure established in the laboratory (Syed *et al*., 2014a; Syed and Mashele, 2014). Briefly, individual proteins were blasted against all named fungal P450s at the Cytochrome P450 Homepage (Nelson, 2009). A family and subfamily were assigned to the P450 proteins based on standard International P450 Nomenclature criteria, i.e. >40% homology for a family and >55% homology for a subfamily. Among the selected proteins, those grouped under the CYP53 family were used in the analysis (Appendix A). The Cytochrome P450 Homepage (Nelson, 2009) was visited to check for the presence of CYP53 members, if any, in the basidiomycetes *Ustilago maydis*, *Cryptococcus neoformans*, *Cryptococcus gattii*, *Laccaria bicolor*, *Malassezia globosa*, *Puccinia graminis* and *Sporobolomyces roseus*. A CYP53 member for *Cochliobolus lunatus* was obtained from the published literature (Podobnik *et al*., 2008).

2.2.2. Phylogenetic analysis

Phylogenetic analysis of CYP53 members was carried out using the Molecular Evolutionary Genetics Analysis (MEGA) software (Tamura *et al*., 2011) following the protocol described

elsewhere (Syed *et al*., 2014(a, b)). Phylogenetic analysis was inferred using the minimum evolution method (Rzhetsky and Nei, 2000). The minimum evolution method is widely used in P450 research, as it involves pairwise distance algorithms for the reconstruction of phylogenies (Chen *et al*., 2012; Syed *et al*., 2014 (a, b)). Hence, in the current study the minimum evolution method was used for phylogenetic analysis of CYP53 member P450s. The evolutionary distances were computed using the Poisson correction method and are in the units of the amino acid substitution per site (Zuckerkandl and Pauling, 1965). The minimum evolution tree was searched using the close-neighbor-interchange algorithm (Nei and Kumar, 2000). The neighbor-joining algorithm was used to generate the initial tree (Saitou and Nei, 1987).

2.2.3. Intron-exon analysis

Gene structure organization of CYP53 family members was carried using the established protocol described elsewhere (Syed *et al*., 2014b). Briefly, each CYP53 member gene was accessed at its genome data base at the JGI, US-DOE or Broad Institute of MIT and Harvard. For each P450 the size of the exons and the location of introns were recorded. A schematic diagram showing horizontal lanes representing the exons and vertical lanes representing the introns' location were drawn. The length of the horizontal lane corresponds to the gene length. CYP53 members that showed high conservation in terms of the size of exons and the location of introns were shown in a figure.

2.2.4. Analysis of homology

To identify the percentage homology between CYP53 members, ClustalW2 multiple sequence analysis was performed (Larkin *et al*., 2007). CYP53 members in FASTA format

were included in the analysis and the result summary showing the percentage identity matrix was downloaded. After the file had been downloaded, the results were converted into table format and checked for the percentage homology between CYP53 members.

2.2.5. Analysis of amino acid conservation

The number of amino acids conserved in CYP53 members across the fungi and between ascomycota and basidiomycota was determined using PROfile Multiple Alignment with predicted Local Structures and 3D constraints (PROMALS3D) (Pei *et al*., 2008). PROMALS3D aligns multiple protein sequences and/or structures, with enhanced information from database searches, secondary structure prediction, 3D structures or user-defined constraints and it will also give a conservation index (Pei and Grishin, 2001). The conservation index follows numbers above 4, where 9 is the invariantly conserved amino acid across the input sequences.

2.3. Results and discussion

2.3.1. CYP53 distribution in fungi

The CYP53 family is one of the P450 families apart from CYP51 and CYP61 that are conserved between the phyla ascomycota and basidiomycota (Črešnar and Petrič, 2011; Moktali *et al*., 2012). In this study, 51 fungal species belonging to ascomycota (23 species) and basidiomycota (28 species) were screened for analysis of CYP53 family members. Genome data mining of ascomycetes (23 species) and basidiomycetes (28 species) revealed the presence of one to nine copies of CYP53 members in their genomes (Table 2.1). The CYP53 family member count ranged from one to three in ascomycetes and one to ten in basidiomycetes. The basidiomycete species *F. mediterranea* showed the maximum number of

CYP53 members (10 CYP53 P450s) in its genome. No CYP53 member was identified in the ascomycete *Histoplasma capsulatum* and in the basidiomycetes *C. neoformans*, *C. gattii* and *M. globosa* or the symbiotic *L. bicolor* (Table 2.1). Overall, ascomycete species showed a lower number of CYP53 member P450s in their genomes compared to basidiomycete species (Table 2.1), suggesting a possible duplication of CYP53 members after the phylum divergence. Moreover, analysis revealed the complete absence of CYP53 member P450s in phyla zygomycota and chytridiomycota. Furthermore, in ascomycota only species belonging to subphyla pezizomycotina showed CYP53 members in their genomes and CYP53 member P450s were not found in species of the subphyla saccharomycotina (Kgosiemang *et al*., 2014) and taphinomycotina, which is in accordance with the smaller size of the P450ome in relation to the growth form of the fungus. Overall, contrary to the established assumption that this family is conserved in fungi, this study showed that CYP53 is not conserved across the fungal species. In future, further genome sequencing analysis of species belonging to chytridiomycota and zygomycota and the subphylum taphrinomycotina could be performed that may provide more information on the presence of this protein family in their genome. However, considering the life style and small size genomes of saccharomycotina species, the absence of CYP53 family members is expected.

Analysis of the CYP53 family suggested the dominance of specific CYP53 subfamilies in ascomycota and basidiomycota (Table 2.1). Ascomycete species showed only the CYP53A subfamily in their genomes, with the exception of *F. oxysporum*, which showed a single copy CYP53 member belonging to the CYP53D subfamily (Table 2.1). In contrast to ascomycete species, basidiomycete species showed divergence in CYP53 subfamilies. Five subfamilies were observed in basidiomycetes, i.e. CYP53A, CYP53B, CYP53C, CYP53D, and CYP53H (Table 2.1). Analysis of CYP53 members in basidiomycetes revealed the presence of two new CYP53 subfamilies in *P. carnosa* and *F. mediterranea*. Among the CYP53 subfamilies

observed for basidiomycota, the CYP53C subfamily was dominant, with 52 members, followed by CYP53D (eight members) and CYP53H (seven members). A single copy of CYP53A members was found in *A. delicata*, *P. strigosozonata*, and *C. puteana* (Table 2.1). Considering the presence of CYP53A and CYP53D subfamilies in both phyla, one can assume that after the divergence of phyla, ascomycete species might have lost CYP53 subfamilies such as CYP53B, C and H. On the other hand, basidiomycete species enhanced CYP53 numbers in their genome, possibly by genome duplication of CYP53 members in view of the possible requirement of these P450 family members to adapt to diverse ecological niches.

2.3.2. Phylogenetic analysis of CYP53 P450 family

In order to understand the evolution of the CYP53 family and its distribution in fungi, evolutionary analysis of the CYP53 family was performed using the minimum evolution method (Rzhetsky and Nei, 1992). Minimum evolution analysis of CYP53 members showed subfamily-specific and species-specific alignment/grouping of CYP53 members (Figure 2.1), suggesting that after divergence of phyla (ascomycota and basidiomycota) CYP53 members have been subjected to phylum-specific amino acid changes in their structure. The most striking feature was that CYP53 members belonging to a particular basidiomycete species were grouped together (Figure 2.1). This clearly indicates that paralogous evolution of CYP53 members, possibly *via* genome duplication, occurred in basidiomycete species. In a recently published study, the same phenomenon of genome duplication of member P450s in basidiomycete species was reported (Syed *et al*., 2014b). Furthermore, authors also showed that these P450 duplications were necessitated by the fungal species to adapt to diverse ecological niches (Syed *et al*., 2014b). Interestingly, CYP53D1 of *F. oxysporum*

(ascomycete) did not align with its counterpart present in *P. placenta* (basidiomycete) (Figure 2.1), suggesting that extensive changes specific to phyla might have occurred in their primary structure.

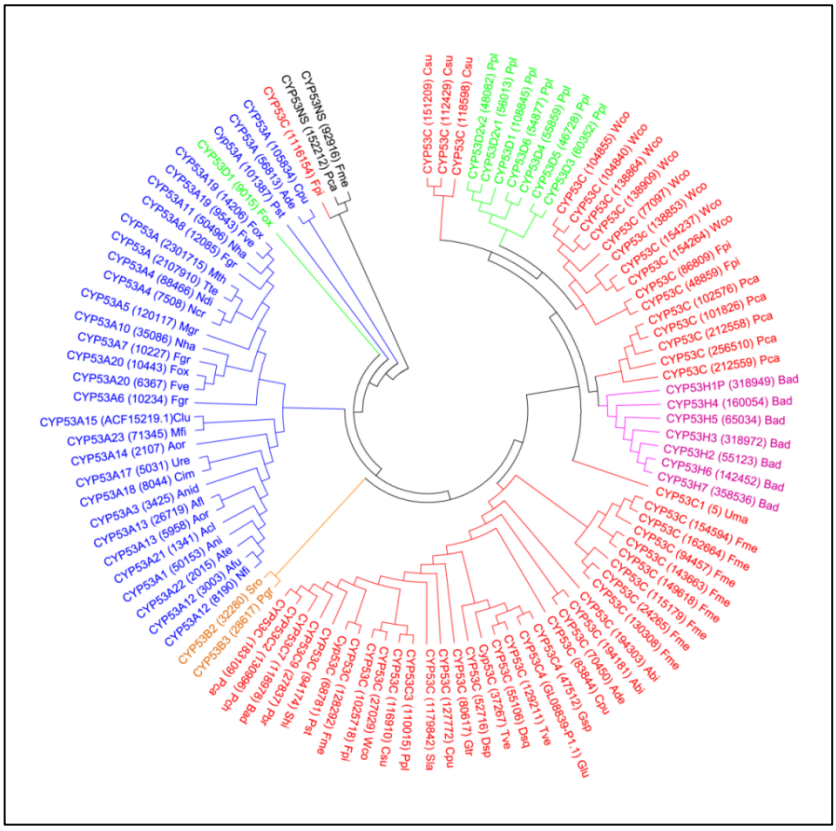

Figure 2.1. Phylogenetic analysis of CYP53 family in fungi. The tree was constructed with 101 CYP53 P450s belonging to six different CYP53 subfamilies. Phylogeny was inferred using the minimum evolution method and the tree was constructed using MEGA (5.05) software. For ease of visual identity, the tree branch color, protein name, protein ID (parenthesis) and species name were presented in unique color as per sub-family. Fungal species' names were indicated with three letters, where the first letter is taken from the genus name and the other two letters from the species name.

2.3.3. High conservation of primary structure of CYP53 members in ascomycota

From the above study, it is highly positive that after divergence of ascomycota and basidiomycota, CYP53 members have been subjected to phyla-specific changes or conservation in their primary structure. In order to understand these phyla-specific changes or conservations in CYP53 members, we followed two methods. Firstly, we analyzed the percentage homology and secondly we deduced amino acids conserved in CYP53 members in both ascomycetes and basidiomycetes.

ClustalW2 analysis of CYP53 members revealed a high percentage homology among CYP53 members (Table 2.2) in ascomycota; some of the members showed >90% homology compared to CYP53 members in basidiomycota. The observed high percentage homology in CYP53 members of ascomycota (Table 2.2) might be due to the dominance of a single CYP53A subfamily. It is noteworthy that although the CYP53C subfamily is dominant in basidiomycota (Table 2.1), most of its members seem to be subjected to major amino acid changes, as the percentage homology between CYP53C members is not high with exception of a few P450s, as observed for CYP53A members for ascomycota (Table 2.2).

Table 2.2. Analysis of homology between CYP53 members in fungi. The percentage (%) homology between CYP53 members was obtained from the Cytochrome P450 Homepage (Nelson, 2009) based on their highest hit to reference proteins and also estimated using ClustalW2 (Larkin *et al*., 2007). P450s showing more than 90% homology were selected and presented in the table. As shown in the table, a higher number of CYP53 members from ascomycota showed more than 90% homology, suggesting high conservation of the primary

structure in ascomycete species CYP53 members compared to basidiomycete species CYP53 members. For each P450 protein, IDs were shown in parenthesis.

CYP name	Species name	Homology (%)	CYP name	Species name
Ascomycota				
CYP53A4 (7508)	*Neurospora crassa*	98	CYP53A4 (88466)	*Neurospora discreta*
CYP53A (2107910)	*Thielavia terrestris*	91	CYP53A (2301715)	*Myceliophthora thermophila*
CYP53A19 (9543)	*Fusarium verticillioides*	98	CYP53A19 (14206)	*Fusarium oxysporum*
CYP53A8 (12085)	*Fusarium graminearum*	95	CYP53A19 (9543)	*Fusarium verticillioides*
CYP53A8 (12085)	*Fusarium graminearum*	95	CYP53A19 (14206)	*Fusarium oxysporum*
CYP53A20 (6367)	*Fusarium verticillioides*	99	CYP53A20 (10443)	*Fusarium oxysporum*
CYP53A7 (10227)	*Fusarium graminearum*	97	CYP53A20 (6367)	*Fusarium verticillioides*
CYP53A7 (10227)	*Fusarium graminearum*	97	CYP53A20 (10443)	*Fusarium oxysporum*
CYP53A12 (8190)	*Neosartorya fischeri*	98	CYP53A12 (3003)	*Aspergillus fumigatus*
CYP53A12 (3003)	*Aspergillus fumigatus*	94	CYP53A21 (1341)	*Aspergillus clavatus*
CYP53A13 (5958)	*Aspergillus oryzae*	99	CYP53A13 (26719)	*Aspergillus flavus*
Basidiomycota				
CYP53C4 (47512)	*Ganoderma* sp.	95	CYP53C4 (GL08839-P1.1)	*Ganoderma lucidum*

CYP53D3 (60352)	*Postia placenta*	95	CYP53D5 (46728)	*Postia placenta*
Basidiomycota				
CYP53C (104855)	*Wolfiporia cocos*	67	CYP53C (154237)	*Wolfiporia cocos*

To link the high percentage homology observed for CYP53 members of ascomycetes towards conservation of amino acid in their primary structure, amino acid conservation studies were performed using PROMALS3D (Figures 2.2). PROMALS3D analysis of CYP53 members across fungi suggested conservation of eight amino acids. Conservation of only eight amino acids in CYP53 members across fungi is understandable, considering the high diversity of CYP53 members across fungal species (five subfamilies and two new subfamilies). The most striking difference was observed in the number of amino acids conserved in the CYP53 members of ascomycota and basidiomycota (Figures 2.2). A hundred and three amino acids were found conserved in CYP53 members of ascomycota compared to CYP53 members of basidiomycota (Figure 2.2A), which showed only seven amino acids conserved in their primary structure (Figure 2.2B). This strongly suggests that the observed high percentage homology between CYP53 members of ascomycota is due to the high conservation of amino acids in their primary structure.

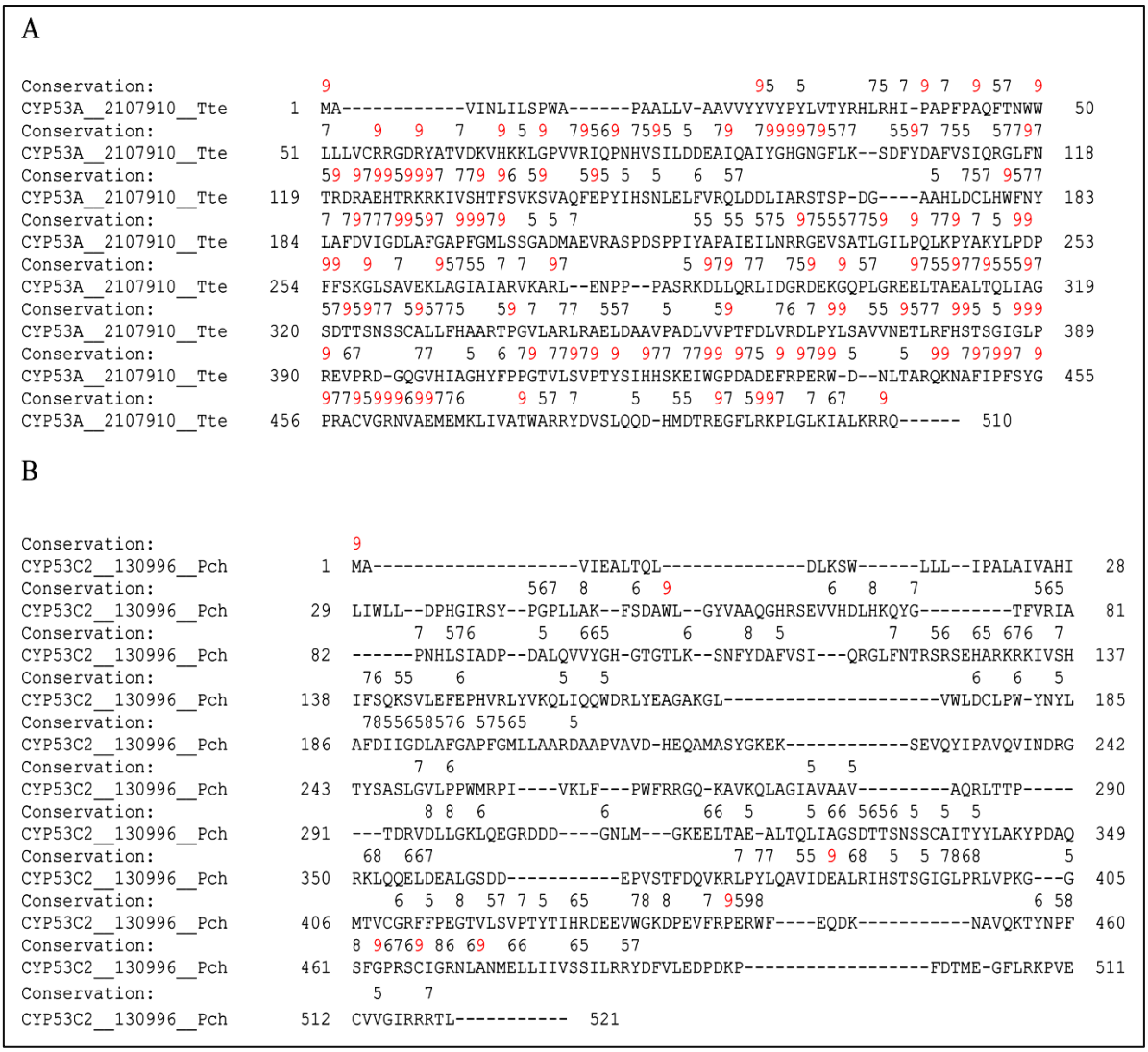

Figure 2.2. Analysis of amino acid conservations in CYP53 family members of ascomycota (A) and basidiomycota (B). Analysis of amino acid conservations was carried out using PROMALS3D (Pei *et al*., 2008). CYP53A from *T. terrestris* and CYP53C2 from *P. chrysosporium* were presented as a representative of ascomycota (A) and basidiomycota (B) CYP53 members. The residues conserved in CYP53 members of ascomycota (A) and basidiomycota(B) are shown with the conservation index on top of the amino acid residue (Pei and Grishin, 2001).

One can argue that the high conservation of amino acids (103 amino acids) in CYP53 members of ascomycota (Figure 2.2A) is due to the presence of a single CYP53A subfamily whereas five subfamilies and two new subfamilies exist in basidiomycota. To rule out this argument, two types of evidence were presented in this study. Firstly, CYP53A members from ascomycete species belonging to 11 different genera were used in this study (Table 2.1), suggesting the high diversity of host species, which should thus reflect in CYP53A primary structure as well. However, this was not true, as ascomycete CYP53 members showed high conservation in the primary structure (Figure 2.2A). Secondly, estimation of the number of amino acids conserved in the CYP53C subfamily was performed (Figure 2.2B), which is the subfamily that is the most dominant in basidiomycota. Interestingly, analysis revealed that there was conservation of only 20 amino acids in CYP53C subfamily members in basidiomycota (Figure 2.3), further consolidating the hypothesis that basidiomycota CYP53 members have been subjected to extensive primary structure changes. Further studies were carried out to map the location of conserved amino acids to extrapolate the effect of the conservation in CYP53 substrate specificity or catalytic activity, if any.

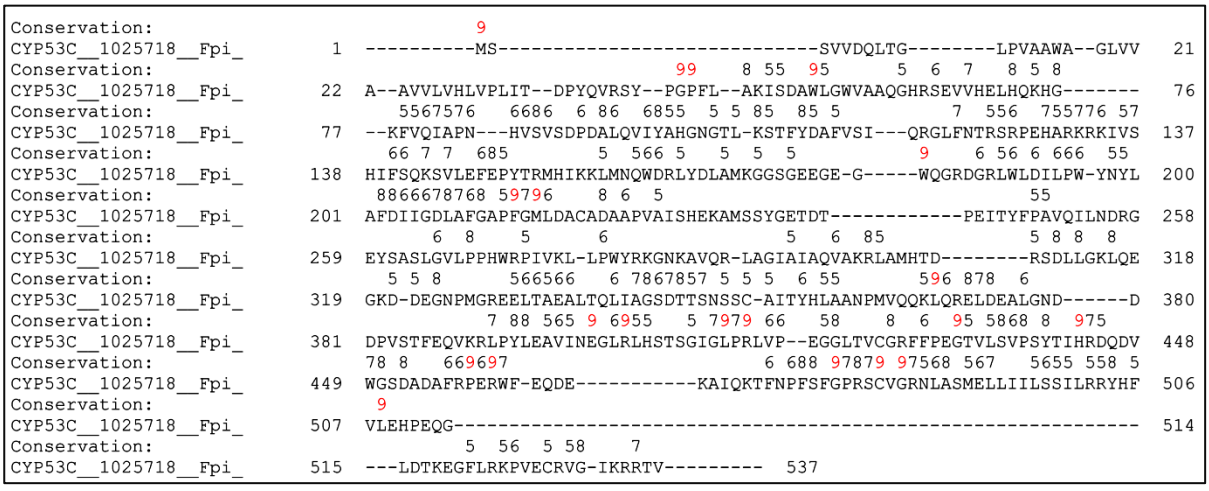

Figure 2.3. Analysis of amino acid conservations in CYP53C subfamily of basidiomycota. Analysis of amino acid conservations was carried out using PROMALS3D

(Pei *et al*., 2008). CYP53C from *F. pinicola* is presented as a representative of CYP53C members. The residues conserved in CYP53C members are shown with the conservation index (Pei and Grishin, 2001) on top of the amino acid residue.

2.3.4. Gene conservation and genome duplications of CYP53 members

The above study indicated high conservation of CYP53 members' primary structure (at a protein level) in ascomycetes compared to basidiomycetes. To gain insight into this aspect, further analysis was carried out on the gene structure of CYP53 members (Figure 2.4). Analysis of the size of exons and the location of introns indicated high conservation of the gene-structure in CYP53 members belonging to both fungal phyla ascomycota (Figure 2.4.) and basidiomycota (Figure 2.4). Gene structure analysis suggested that some ascomycete species, such as *F. oxysporum*, *F. solani f. batatas (Nectria haematococca*), *F. verticillioides*, and *F. graminearum*, contain two types of ortholog P450s in their genome. The first type contains a single intron and the second one contains three introns (Figure 2.4). Paralog P450s were found in *F. graminearum* (protein IDs: 10234 and 10227) and *A. oryzae* (protein IDs: 2107 and 5958), suggesting the genome duplication of these P450s. Overall, ascomycete species CYP53 members showed simple gene structure with single and triple introns (Figure 2.4).

It is evident from Figure 2.5, especially considering the exon sizes and location of introns, that basidiomycete species enriched CYP53 members in their genome by genome duplications (paralogous evolution). The high conservation in the size of exons and location of introns of CYP53 members of basidiomycetes strongly suggests that CYP53 members are genome-duplicated. In comparison to ascomycete species (Figure 2.4), CYP53 members of basidiomycete species showed more introns in their structure (Figure 2.5). An interesting discovery was that basidiomycete species selectively enriched a single type of CYP53

member in their genome (Figure 2.5). Some conclusion were presented in support of this argument: (i) in *P. placenta* two orthologs, of which one duplicated seven times while no duplication was observed for the second ortholog (protein ID: 110015); (ii) in *W. cocos* three orthologs were found: one ortholog duplicated seven times whereas no duplications were observed for the remaining two orthologs (protein IDs: 138909 and 27029); (iii) in *F. pinicola*, *P. carnosa* and *F. mediterranea* two orthologs were found in their genomes; in these species one ortholog was duplicated whereas no duplication was observed for the second ortholog protein ID: 1025718 (*F. pinicola*); protein ID: 183190 (*P. carnosa*); protein ID: 162664 (*F. mediterranea*); (iv) *B. adusta* showed three orthologs; two orthologs (protein IDs: 318949 and 118978) have remained the same since the divergence of this species.

From the above results it is clear that the higher number of CYP53 members in basidiomycetes is due to the genome duplication of selective CYP53 members. Despite the conservation of gene structure and the paralogous evolution of CYP53 members in basidiomycete species, the low percentage of homology among them suggests that during the genome duplication events, extensive changes in the primary structure occurred. Most of the changes might be destined to acquire novel functions to serve fungal species (basidiomycete) to adapt to diverse ecological niches.

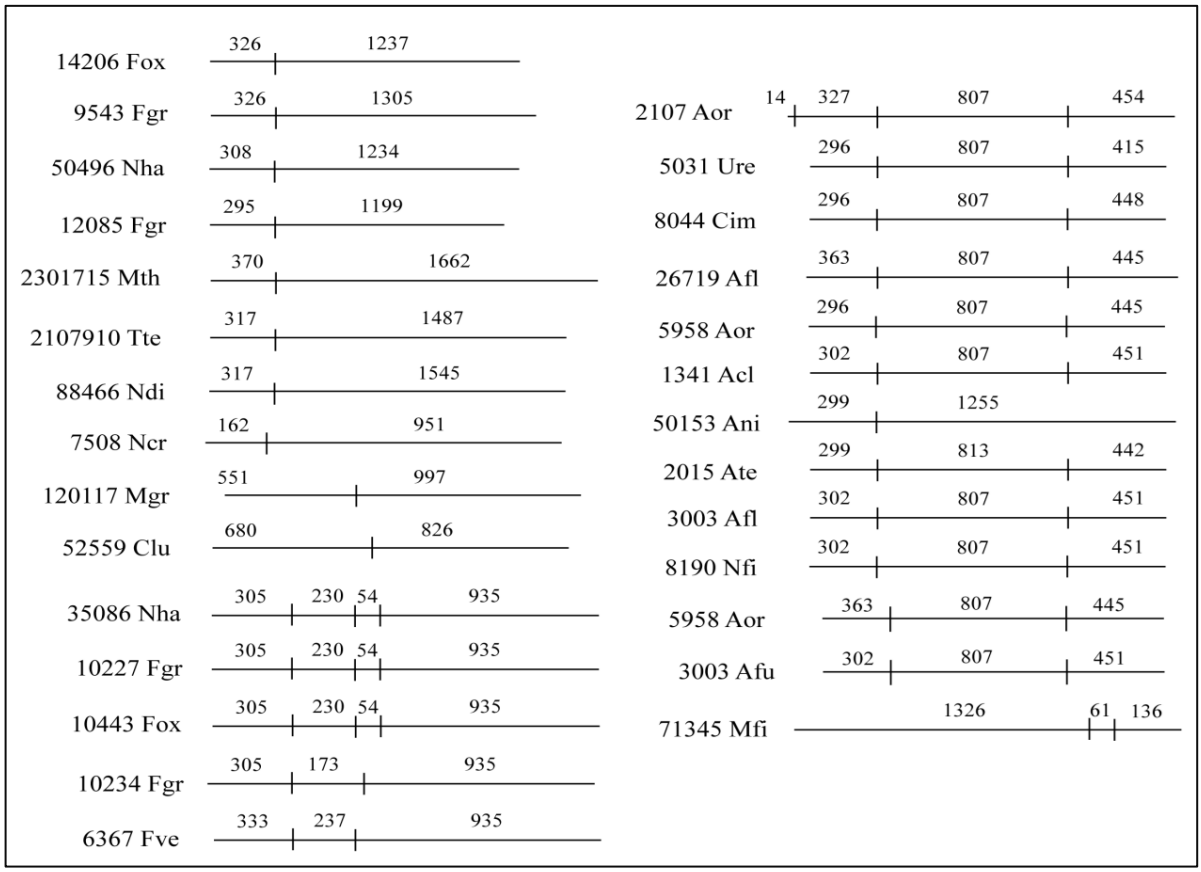

Figure 2.4. Gene-structure analysis of CYP53 family in ascomycete species. Horizontal lines indicate gene size and vertical lines indicate introns. For each CYP53 gene the size of the exons (base pairs) and protein ID from the JGI US-DOE is shown in the figure (Grigoriev *et al*., 2011).

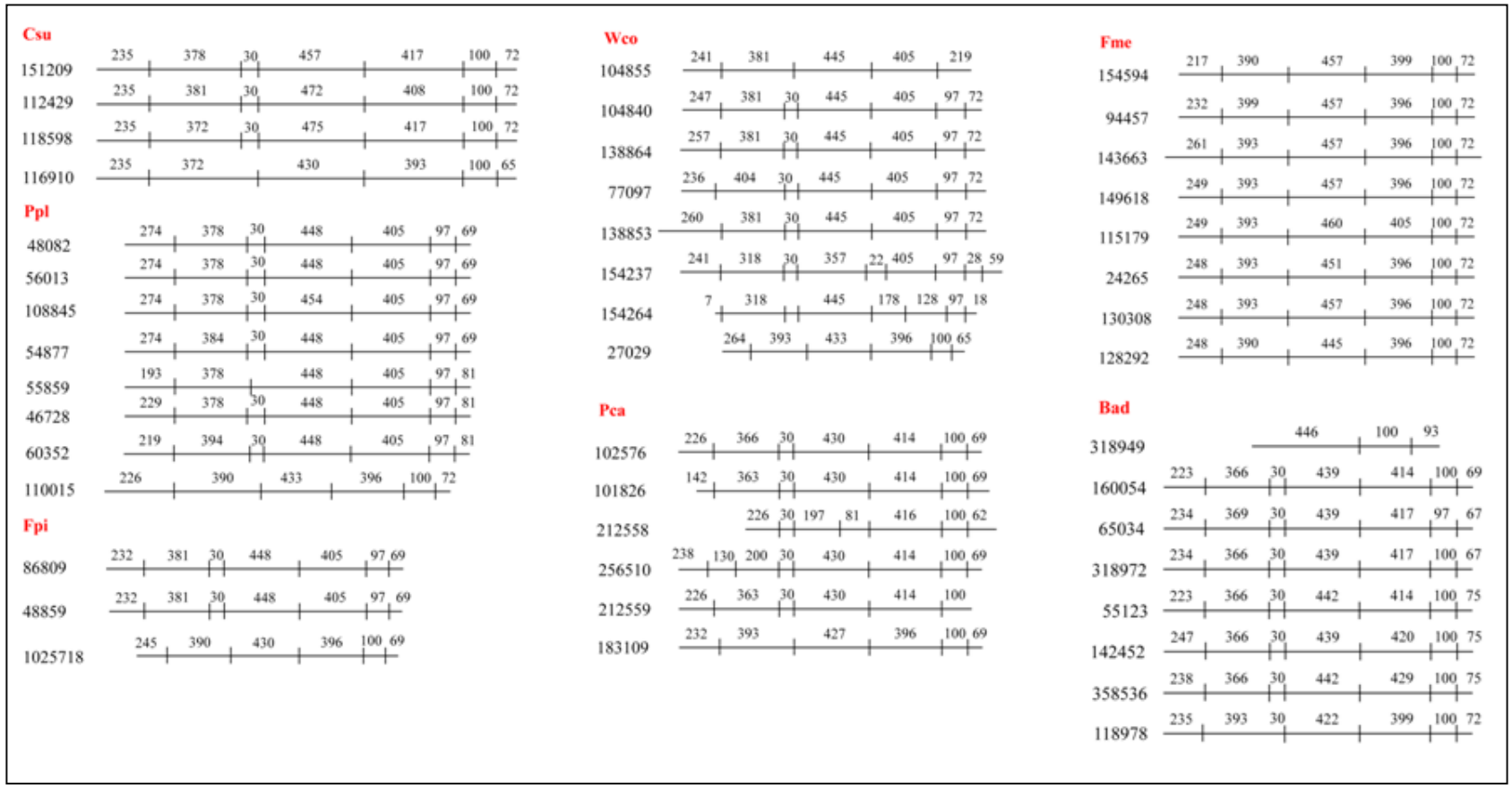

Figure 2.5. Gene-structure analysis of CYP53 family in basidiomycete species. Horizontal lines indicate gene size and vertical lines indicate introns. For each CYP53 gene, the size of the exons (base pairs) and protein ID from the JGI US-DOE is shown in the figure.

2.3.5. Functional significance of CYP53 family and its potential role as a common antifungal drug target

CYP53 family members play a key role in fungal primary metabolism, by using the *β*-ketoadipate pathway (Faber *et al*., 2001; Fraser *et al*., 2002), and secondary metabolism, which is the detoxification of phenolic compounds (Ide *et al*., 2012; Fujii *et al*., 1997). The *β*-ketoadipate pathway is a convergent pathway for aromatic compound degradation (Harwood and Parales, 1996) that is widely distributed in soil bacteria and fungi. Fungal-mediated degradation of aromatic compounds such as phenylalanine, toluene, and cinnamic acid leads to the formation of benzoate (Lapadatescu *et al*., 2000; Durham *et al*., 1984; Jensen *et al*., 1994). As part of the *β*-ketoadipate pathway CYP53 is involved in detoxification of this toxic compound and key intermediate molecule. CYP53 hydroxylates benzoate to 4-hydroxybenzoate (Faber *et al*., 2001), the prime reaction in the benzoate metabolism that subsequently leads to protocatechuate as the ring fission substrate (Wright, 1993). This reaction is critical for fungal organisms in order to detoxify the benzoate; to date this hydroxylation reaction carried out by CYP53 is the only way to detoxify this compound. Further support of CYP53's critical role in fungal primary metabolism can be obtained from a study where CYP53 deletion proved to be lethal for fungal organisms' survival (Fraser *et al*., 2002). This suggests that the CYP53 family can serve as a novel alternative drug target against fungal pathogens, especially ascomycete pathogens. Results from this work showing high conservation of the primary structure of CYP53 members (Figure 2.4) across the ascomycetes (consisting of animal and plant pathogen fungal species) indicate that any inhibitor developed against a CYP53 member could serve as a novel common drug against a large number of pathogenic ascomycete fungi. Literature data suggested that inhibitors directed at this P450 effectively inhibited CYP53 activity and also growth inhibition of

different fungal species such as *C. lunatus*, *A. niger* and *Pleurotus ostreatus* (Podobnik *et al*; 2008; Korošec *et al.*, 2014). Furthermore, this P450 family offers an advantage over the CYP51 family, the currently exploited target against fungal infections, as CYP53 does not have a homolog in higher eukaryotes. This will offer researchers the opportunity to design selective and potent inhibitors of pathogenic fungi.

Overall, the facts discussed above, such as (i) the critical role of CYP53 in fungal primary metabolism, (ii) high conservation of the primary structure of CYP53 members in ascomycetes and (iii) CYP53 not having any homolog in higher eukaryotes (advantage over CYP51 family), this strongly support the suggested hypothesis that the CYP53 family can be a potential novel alternative anti-fungal drug target and an inhibitor designed against this P450 family can serve as a common drug against pathogenic ascomycetes. The most interesting aspect of the CYP53 family's role in basidiomycete fungi extends beyond detoxification of benzoate. The current study showed that most of the ascomycetes contain a single CYP53 member in their genomes, whereas basidiomycetes showed multiple CYP53 members (Table 2.1). Results from this study (Figure 2.5) revealed that the number of CYP53 members increases in basidiomycete species' genomes, by duplication of CYP53 members after speciation (paralogous evolution) occurs.

Several lines of evidence can be found on the critical role of these CYP53 members in basidiomycetes that forced basidiomycetes to enhance this P450 family member in their genomes. First, basidiomycetes are well known for their role as bio-degraders of wood (Martinez *et al*., 2005). Wood is composed of many aromatic compounds, including benzoic acid derivatives and other phenolic compounds, among others eugenol, isoeugenol and guaiacol (Haupert *et al*., 2012). Most of these compounds are anti-fungal and toxic to fungi (Amborabe *et al*., 2002). The multi-factorial phenomenon of toxicity of these compounds, including membrane disruption, inhibition of essential metabolic reactions, changes in pH

homeostasis, and accumulation of toxic anions, has been proposed toward fungi (Brul and Coote, 1999). Basidiomycete species require an enzyme that can detoxify the benzoate molecule, if they want to colonize on wood, as this molecule is an intermediate in detoxification of wood components comprising of many aromatic compounds. Since there is an enormous need for successful wood colonization, wood-degrading basidiomycetes amplified the number of CYP53 members in their genomes.

Secondly, synthesis of aryl-metabolites, including veratryl alcohol by basidiomycete fungi, involves the formation of benzoate and *para*-hydroxybenzoic acid as intermediate molecules (Lapadatescu *et al*., 2000). Veratryl alcohol is a secondary metabolite and plays a key role in lignin-peroxidase-mediated oxidation of wood components (Ten Have *et al*., 1998). In a recent study, veratryl alcohol was shown to be the dominant extracellular ligninolytic oxidant in decaying wood (Hunt *et al*., 2013). The presence of a high number of CYP53 members and the generation of benzoate and *para*-hydroxybenzoate as an intermediate in the biosynthesis of veratryl alcohol suggest that in basidiomycete species CYP53 members also play a role in the generation of veratryl alcohol and help basidiomycete species directly in the degradation and subsequent colonization of wood.

Thirdly, demethylation of stilbene, a class of molecule found in plants, by CYP53D subfamily members from *P. placenta* (basidiomycete) indicates that CYP53 family members play a critical role in the detoxification or degradation of plant compounds and help fungi in the colonization of wood (Ide *et al*., 2012). It is noteworthy that CYP53D members are present in the highest numbers (seven P450s) in *P. placenta* and all evolved *via* paralogous evolution (Figure 2.5). This strongly indicates that *P. placenta* duplicated CYP53D members in its genome in order to colonize successfully on wood.

The above-mentioned role of CYP53 in wood-degrading basidiomycete species physiology (primary or secondary metabolism) is based on the available data and further experimentation would provide more insight into this aspect. Collectively, the above results indicate that in ascomycetes the CYP53's role is limited to the detoxification of toxic molecules, whereas in basidiomycetes CYP53 plays an additional role, in the generation of veratryl alcohol and degradation of wood-derived compounds.

2.4. Conclusion

In this advanced scientific era, understanding of animal (including human) and plant pathogenic fungal organisms in terms of controlling their causative diseases and developing effective drugs is still poorly understood. Currently available drugs and drug targets are becoming ineffective because fungal species develop resistance. Genome sequencing analysis of the fungal species gives researchers the opportunity to look for novel drug targets against these pathogens and to search for novel enzymes for the generation of human valuables. The present study is such an example where fungal genome sequencing results were explored to understand the role of a P450 family (CYP53) in serving as a common drug target against pathogenic ascomycetes and in basidiomycetes, particularly in terms of the wood-degradation process. The CYP53 family plays a key role in the detoxification of the toxic molecule benzoate and this family has proven to be essential for the organism's survival. Findings from this study suggest that this P450 family can serve as a common anti-fungal (toward pathogenic ascomycetes) drug target in view of its highly conserved primary protein structure and gene-structure organization in ascomycetes. Moreover, this study identified CYP53 P450s can play an additional role in basidiomycetes, that is, in the generation of the wood-degrading oxidant veratryl alcohol and degradation of wood-derived compounds. This additional role of basidiomycetes seems to have enriched this P450 family by extensive

duplication of CYP53 members in their genomes (paralogous evolution). During the duplication process, extensive changes in the protein primary structure occurred to enhance/acquire novel functions, such as involvement in wood degradation.

REFERENCES

Amborabe, B.E., Fleurat-Lessard, P., Chollet, J-F., Roblin, G. (2002). Anti-fungal effects of salicylic acid and other benzoic acid derivatives towards *Eutypalata*: structure-activity relationship. *Plant Physiol Biochem.*, 40 (12): 1051-1060.

Brul, S., Coote P. (1999). Preservative agents in foods. Mode of action and microbial resistance mechanisms. *Int J Food Microbiol.*, 50 (1): 1–17.

Chen, S., Xu, J., Liu, C., Zhu, Y., Nelson, D.R., Zhou, S., Li, C., Wang, L., Guo, X., Sun, Y. , Luo, H., Li, Y., Song, J., Henrissat, B., Levasseur, A., Qian, J., Li, J., Luo, X., Shi, L., He, L., Xiang, L., Xu, X., Niu, Y., Li, Q., Han, M.V., Yan, H., Zhang, J., Chen, H., Lv, A., Wang, Z., Liu, M., Schwartz, D.C., Sun, C. (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nature Commun.*, 3: 913.

Črešnar, B., Petrič, S. (2011). Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta.*, 1814 (1): 29–35.

Durham, D.R., Mcnamee, C.G., Stewart, D.B. (1984). Dissimilation of aromatic-compounds in *Rhodotorula-graminis*. Biochemical-characterization of pleiotropically negative mutants. *J Bacteriol.*, 160 (2): 771–777.

Faber, B.W., van Gorcom, R.F.M., Duine, J.A. (2001). Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. *Arch Biochem Biophys.*, 394 (2): 245–254.

Fernandez-Fueyo, E., Ruiz-Dueñas, F.J., Ferreira, P., Floudas, D., Hibbett, D.S., Canessa, P., Larrondo, L.F., James, T.Y., Seelenfreund, D., Lobos, S., Polanco, R., Tello, M., Honda, Y., Watanabe, T., Watanabe, T., Ryu, J.S., Kubicek, C.P., Schmoll, M., Gaskell, J., Hammel,

K.E., St John, F.J., Vanden Wymelenberg, A., Sabat, G., Splinter BonDurant, S., Syed, K., Yadav, J.S., Doddapaneni, H., Subramanian, V., Lavín, J.L., Oguiza, J.A., Perez, G, Pisabarro, A.G. , Ramirez, L , Santoyo, F , Master, E , Coutinho, P. M. , Henrissat, B , Lombard, V, Magnuson, J. K. , Kües,U., Hori, C, Igarashi, K., Samejima, M. , Held, B.W., Barry, K.W. , LaButti, K.M., Lapidus, A., Lindquist, E.A., Lucas, S.M., Riley, R., Salamov, A.A., Hoffmeister, D., Schwenk, D., Hadar, Y., Yarden, O., de Vries, R.P., Wiebenga, A., Stenlid, J., Eastwood, D., Grigoriev, I.V., Berka, R.M., Blanchette, R.A., Kersten, P., Martinez, A.T., Vicuna, R., Cullen, D. (2012). Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. *Proc Natl Acad Sci U S A.*, 109 (14):5458–5463.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., Aerts A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P.M., de Vries, R.P., Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Górecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J.A., Kallen, N., Kersten, P., Kohler, A., Kües, U., Kumar, T. K. A., Kuo, A., LaButti, K., Larrondo, L.F. , Lindquist, E., Ling, A., Lombard, V., Lucas, S. , Lundell, T., Martin, R. , McLaughlin, D.J., Morgenstern, I., Morin, E., Murat, C., Nagy, L.G., Nolan, M., Ohm, R.A., Patyshakuliyeva, A., Rokas, A., Ruiz-Dueñas, F.J. , Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J.C., St. John, F., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A., Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D.C., Martin, F., Cullen, D., Grigoriev, I.V., Hibbett, D.S. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, 336 (6089): 1715–1719.

Fraser, J.A., Davis, M.A., Hynes, M.J. (2002). The gene gmdA, encoding an amidase and bzuA, encoding a cytochrome P450, are required for benzamide utilization in *Aspergillus nidulans*. *Fungal Genet Biol.*, 35 (2): 135-146.

Fujii, T., Nakamura, K., Shibuya, K., Tanase, S., Gotoh, O., Ogawa, T., Fukuda, H. (1997). Structural characterization of the gene and corresponding cDNA for the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. *Mol Gen Genet.*, 256 (2): 115–120.

Grigoriev, I.V., Cullen, D., Goodwin, S.B., Hibbett, D., Jeffries, T.W., Kubicek, C.P., Kuske, C., Magnuson, J.K., Martin, F., Spatafora, J.W., Tsang, A., Baker, S.E. (2011). Fueling the future with fungal genomics. *Mycology*, 2 (3): 192-209.

Harwood, C.S., Parales, R.E. (1996). The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol.*, 50 (1): 553-590.

Haupert, L.J., Owen, B.C., Marcum, C.L., Jarrell, T.M., Pulliam, C.J. (2012). Characterization of model compounds of processed lignin and the lignome by using atmospheric pressure ionization tandem mass spectrometry. *Fuel*, 95: 634-641.

Hlavica, P. (2013). Evaluation of structural features in fungal cytochromes P450 predicted to rule catalytic diversification. *Biochim Biophys Acta.*, 1834 (1): 205–220.

Hof, H. (2001). Critical annotations to the use of azole antifungals for plant protection. *Antimicrob Agents Chemother.*, 45 (11): 2987-2990.

Hunt, C. G., Houtman, C. J., Jones, D. C., Kitin, P., Korripally, P., Hammel, K. E. (2013). Spatial mapping of extracellular oxidant production by a white rot basidiomycete on wood reveals details of ligninolytic mechanism. *Environ Microbiol.*, 15 (3): 956-966.

Ide, M., Ichinose, H., Wariishi, H. (2012).Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete *Postia placenta*. *Arch Microbiol.*, 194 (4): 243–53.

Jensen, K.A., Evans, K.M.C., Kirk, T.K., Hammel, K.E. (1994). Biosynthetic-pathway for veratryl alcohol in the ligninolytic fungus *Phanerochaete chrysosporium*. *Appl Environ Microbiol.,* 60 (2): 709–714.

Kelly, S.L., Kelly, D.E. (2013). Microbial cytochrome P450: biodiversity and biotechnology, where do cytochrome P450 come from, what do they do and what can they do for us?. *Phil Trans R Soc B Biol Sci.,* 368 (1612): 20120476.

Kgosiemang, IKR., Mashele, SS., Syed, K. (2014). Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum Saccharomycotina. *J Pure Appl Microbiol*., 8(Spl. Edn. 2): 291-302.

Korošec, B., Sova, M., Turk, S., Kraševec, N., Novak, M., Lah, L., Stojan, J., Podobnik, B., Berne, S., Zupanec, N., Bunc, M., Gobec, S., Komel, R. (2014). Antifungal activity of cinnamic acid derivatives involves inhibition of benzoate 4-hydroxylase (CYP53). *J Appl Microbiol.,* 116 (4): 955-966.

Lapadatescu, C., Ginies, C., Le Quere, J.L., Bonnarme, P. (2000). Novel scheme for biosynthesis of aryl metabolites from L-phenylalanine in the fungus *Bjerkandera adusta*. *Appl Environ Microbiol.,* 66 (4): 1517-1522.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R. ,McGettigan, P.A., McWilliam, H., Valentin, F. , Wallace, I.M., Wilm, A. , Lopez, R. , Thompson, J.D., Gibson, T.J., Higgins , D.G. (2007). Clustal W and Clustal X version 2.0.*Bioinformatics,* 23 (21): 2947-2948.

Lepesheva, G.I., Waterman, M.R. (2004). CYP51—the omnipotent P450. *Mol Cell Endocrinol.*, 215 (1): 165–170.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C. , Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson,

J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H, Mullokandov, M., Omelchenko, M.V. , Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., Bryant, S.H. (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res*., 39 (suppl 1): D225-D229.

Martínez, Á.T., Speranza, M., Ruiz-Dueñas, F.J., Ferreira, P., Camarero, S., Guillén, F. ,Martínez, M.J., Gutiérrez, A., del Río, J.C. (2005). Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *Int Microbiol*, 8 (3): 195-204.

Moktali, V., Park, J., .Fedorova-Abrams, N. D., Park, B. , Choi, J., Lee, Y.H., Kang, S. (2012). Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics*, 13 (1): 525.

Nei, M., Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nelson, D.R. (2009). The cytochrome P450 homepage. Hum Genomics, 4 (1): 59–65.

Nelson, D.R. (2013). A world of cytochrome P450s. *Phil. Trans. R. Soc. B. Biol. Sci.,* 368 (1612): 20120430.

Pei, J., Grishin, N.V. (2001).AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, 17 (8): 700-712.

Pei, J., Kim, B.H., Grishin, N.V. (2008).PROMALS3D: A tool for multiple sequence and structure alignment. *Nucleic Acids Res*, 36 (7): 2295-2300.

Podobnik, B, Stojan, J., Lah, L., Krasevec, N., Seliskar, M., Rizner, T.L., Rozman, D., Komel, R. (2008). CYP53A15 of *Cochliobolus lunatus*, a target for natural antifungal compounds. *J Med Chem*, 51(12): 3480-3486.

Rzhetsky, A., Nei, M. (1992).A simple method for estimating and testing minimum evolution trees. *Mol Biol Evol*, 9 (5): 945–967.

Saitou, N., Nei, M. (1987).The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4 (4): 406–425.

Sangamwar, A.T., Deshpande, U.D., Pekamwar, S.S. (2008). Antifungals: need to search for a new molecular target. *Indian J Pharm Sci*, 70 (4): 423-430.

Sanglard, D.(2002). Resistance of human fungal pathogens to antifungal drugs. *Curr Opin Microbiol*, 5: 379-385.

Suzuki, H., MacDonald, J., Syed, K., Salamov, A., Hori, C., Aerts, A., Henrissat, B., Wiebenga, A., van Kuyk, P.A, Barry, K., Lindquist, E., LaButti, K., Lapidus, A., Lucas, S., Coutinho, P., Gong, Y., Samejima, M., Mahadevan, R., Abou-Zaid, M., de Vries, R.P., Igarashi, K., Yadav, J.S, Grigoriev, I.V., Master, E.R. (2012). Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics*, 13 (1): 444.

Syed, K., 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): e95616.

Syed, K., Nelson, D.R., Riley, R., Yadav, J.S. (2013).Genome-wide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the Polyporale

species *Bjerkandera adusta*, *Ganoderma sp.* and *Phlebia brevispora*. *Mycologia*, 105 (6): 1445–1455.

Syed, K., Shale, K., Nazir, K.H.M.N.H., Krasevec, N., Mashele, S.S., Pagadala, N. S. (2014a).Genome-wide identification, annotation and characterization of novel thermostable cytochrome P450 monooxygenases from the thermophilic biomass-degrading fungi *Thielavia terrestris* and *Myceliophthora thermophila*. *Genes Genom*, 36 (3): 321-333.

Syed, K., Shale, K., Pagadala, N.S., Tuszynski, J. (2014b). Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 Monooxygenase families enriched in model basidiomycete fungi. *PLoS ONE*, 9(1): e86683.

Syed, K., Yadav, J.S. (2012). P450 monooxygenases (P450ome) of the model white rot fungus *Phanerochaete chrysosporium*. *Crit Rev Microbiol*, 38 (4): 339–363.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011).MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, 28 (10): 2731–2739.

TenHave, R., Rietjens, I.M., Hartmans, S., Swarts, H.J., Field, J.A. (1998). Calculated ionization potentials determine the oxidation of vanillin precursors by lignin peroxidase. *FEBS Lett*, 430 (3): 390-392.

Wright, J.D. (1993).Fungal degradation of benzoic-acid and related compounds. *World J Microbiol Biotechnol*, 9 (1): 9–16.

Yoshida, Y. (1988).Cytochrome P450 of fungi: primary target for azole antifungal agents. *Cur Top Med Mycol*, 2: 388-418.

Zuckerkandl, E., Pauling, L. (1965).Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. *Evolving Genes and Proteins*. Academic Press, New York.97: 97–166.

Appendix A

P450 Sequences for CYP53 Family:

>1CYP53A5 (120117) Mgrs

MAIVNLVFTPLGLASLGAFMLVAYYVVPYFTTFGHLRSIQPASPLAGFSNLWLLYTSRVGKRSLLVDEAHARLGPVLRVQPNHVSIADDEAINI
IYGHGNGFLKSSFYDAFVSIRRGLFNTRDRAEHTRKRKLISHTFAPKSVGQFEPYIHGNLELFAKKWDELIERTKKSDGWAPVECLQWFFGAPF
GMLNAGADIAEVRMSVDSEPIYAPAVEILNRRGEVSATLGTLPELKPYAGYLPDSFFSKGLAAVQNLAGIAIARVKSRLENPPDVNRKDLLARL
QEGRDAKGEPLGFEELTAEALTQLIAGSDTTSNSSCALLYWTARTPGVLAKLQAELDAAIPDGVFAPAFDMIRNLPYLEAVINETLRIHSTSGI
GLPRQIPADSPGVTIRGQYYPPGTVLSVPTYTIHHSKEIWGPDADEFRPERWIENGGLTDRQKNAFIPFSYGPRACVGRNVAEMEMKMIAATWA
RRYDVEVRQDVMEVREGFLRKPLALEIGLKRRS

>2CYP53A4 (7508) Ncr

MAIISLLMSPWAPVVLLAGVAFYYLVPYFVTYSALRKIPSPFPAQFTDLWLLSVCRRGNRYQRVDELHKKLGPVVRIQPNHVSICDDAAIPTIY
GHGNGFLNDFYDAFVSIRRGLFNTRDRAEHTRKRKIVSHTFSAKSVQQFEPYMHSNLELFVKQWDSMIKNSKNPDKAAHLDCLEWFNYLAFDVI
GDLSFGQPFGMLSSGADMAEIRSSPDAAPIYAPAIEILNRRGEVSATLGIHPALKPFAKYLPDPFFTKGLAAVENLAGIAIACVKSRLDNPPPV
TRKDLLQRLMEGRDEKGEPLGREELTAEALTQLIAGSDTTSNSSCALLFHAVRTPGVMQKLQAELDANIPPEVDVPTYDMVKELPYLEAVINEV
LRFHSTSGIGLPRQIPHDASQGVHIQGYYLPPGTVLSVPTYSIHHSKEIWGPDADEFKPERWERLTARQKNAFIPFSHGPRSCVGRNVAEMEMK
LIVATWARRYEVKLLQDYMDTREGFLRKPLGLKVGLKLRK

>3CYP53A4 (88466) Ndi

MAIISLLMSSWAPVVLLAGVAFYYLVPYFVTYSALRKIPSPFPAQFTDLWLLSVCRRGNRYQRVDELHKKLGPVVRIQPNHVSICDDAAIPTIY
GHGNGFLKSDFYDAFVSIRRGLFNTRDRAEHTRKRKIVSHTFSAKSVQQFEPYMHSNLELFVKQWDSMIKNTKNPDKAAHLDCLEWFNYLAFDV
IGDLSFGQPFGMLSSGADMAEIRSSPDAAPIYAPAIEILNRRGEVSATLGIHPALKPFAKYLPDPFFTKGLAAVENLAGIAIACVKSRLENPPP
VTRKDLLQRLMEGRDEKGEPLSREELTAEALTQLIAGSDTTSNSSCALLFHAVRTPGVMQKLQAELDANIPSEVDVPTYDMVKDLPYLEAIINE
VLRFHSTSGIGLPRQIPCDAAQGVHIQGYYFPPGTVLSVPTYSIHHSKEIWGPDADEFKPERWERLTPRQKNAFIPFSHGPRSCVGRNVAEMEM
KLNVATWARRYEVKLLQDYMDTSEGFLRKPLGLKVGLKLRK

>4CYP53A6 (10234) Fgr

MAMFTVLPLIWLAPLGLISLFFYYIIPYFWNYRHLRSIPGPLFARLSNWWLVYACREKSRWKYVNDAHTRYGPVVRIQPNHVSIANEEVINAIY
GHGNGMLKSSFYDASVITTYSIFTSRDRAEHSRKRKVVSHSFAPQSMRNFEPFIQQHLNVFLQKWDAMAANEAKFDGYADVESRVWLNYLVLDI
IGDLAFGAPFGVLAKGSEVVDFETEKGPSSLPVITSLSTRSEIAATVGALPELKPYLKWSPDPFFRTGFNGMINLRTLGTSRITDRLNNPPGDE
REKDLLERVREGRDHKGQPFGKGELIAEALTVLIAGTDTTSSTMAALLYHVVRTPGVLKKLQAELDEAIPADVSIPSFEMVKNLKYLGFVVNEA
LRHHSTISLGLPRLVPENGNGVTIAGYHFAPGTVLSIPIYTVHHLKEVWGPDADEFKPERWEDVTQRQKQAFIPFSHGPRACLGRNLAEMELKV
ITATWARRYDLIMRDDTMEILEGLARKPEAVNVGIRRRM

>5CYP53A7 (10227) Fgr

MAITELLVSPWAPVALVVAFVAWYILPWVSNKDLRGIPAPFPAQFSNLWLLSTCRRGKRYEIVDQVHKKLGPLVRIAPNHVSVADADAINTIYG
HGNGFLKADFYDTFVSIRRGLFNTRDRAEHSRKRKIVSHTFAPKSVLEFEPYIRQNLEIFVKQWDRISSNKERDGYGRVDCLNWFNFLAFDIIA

DLAFGKPFGMLASGADIAEVKASPTSPTIYAPAVEIMNRRGEVSATLGCLPQLKPYAKYLPDPFFSQGLQAVENLAGIAIARVSERLERGGDST
RKDLLARLMQGRDEKGEPLGRDELTAEALTQLIAGSDTTSNSSCALLYHIVRTPGVMKKVYEEISAVMPDGVDIPDFESVKHLPYLGYCINETL
RIHSPSGIGLPREVPPNHKGVTIHGRYFGPGTVLSVPTYTIHHSTEIWGPDADDFKPERWETLTDKQKSAFIPFSYGPRSCVGRNLAEMQMRMI
AATWIKRYDVVLRQDVMETREGFLRKPMGLDVGLARR

>6CYP53A8 (12085) Fgr

MAIVDLLFTWWSLPIAAGLVAASYLYSYFITYGHLRDIPAPFPAQFSNLWLLYVCRRGERYRVVDQIHKELGPVVRIQPNHTSIADAEAIATIY
GHGNGFLKSEFYDAFVSIRRGLFNTRDRAEHTRKRKLISHVFSAKSISQFEPYIHANLELFVKQLDKLVASGQMAKNGKREALMDCLPWFNYLA
FDVIGDLAFGVPFGMLASGADVAEVRDTPDSPPIYASAIEILNRRGEVSATLGCFPQLKPYAQYLPDPFFSNGLNAVKNLAGIAIARVKNRLDN
PPSIERMDLLARLMEGRDEKGEPLGREELTAEALTQLIAGSDTTSNSSCALLYHVTRTPGVLEKLQSELDNAIPSEVSVPTYDMVRDLPYLANV
INETLRYHSTSGIGLPRQIPPNSPGVTIKDHFFPPGSILSVPTYTLHHSKEIWGADADDFRPERWENPTELQKTAFNPFSHGPRACVGRNVAEM
EMKLIAATWARRYVPELRQGVMETREGFLRKPLGLDIALMMRE

>7CYP53A10 (35086) Nhe

MALINLLISPWAPVALVVLFVGWYLVPYFGANRGLRGIPAPFPAQFSNLWLMSTCRRGKRFEVVDQVHKRLGTVVRIAPNHVSIADADAINVIY
GHGNGFLKSDFYDPFVSIRRGLFNTRSRAEHSRKRKIVSHTFAPKSVLEFEPYIHQNLDLFVKQWDRASSNPEADGAGRLDCLSWFNYLAFDVI
ADLAFGKPFGMLATGADIAEVKASPTSPAIYAPAVEIMNRRGEVSATLGCMPWLKPYAKWLPDPFFSQGLQAVENLAGIAIARVSERLERGADT
TRKDLLARLMQGRDEKGEPLGRDELTAEALTQLIAGSDTTSNSSCALLYHVVKTPGVLQKLQQEIDEATADEGVIPSYESVKHLPYLGMCINET
LRHHSPSGIGLPREIPAKSKGVTLHGRYFGPGTVLSVPTYTVHHSTEIWGPDAEEFKPERWENITDKQKIAFIPFSHGPRSCVGRNLAEMQMRL
IAATWIKRYNIFLRQEKMETREGFLRKPLGVEIGVSRR

>8CYP53A11 (50496) Nhe

MAIVDLLLSWWTLPIGVAVLVGTYLYAYFVTYGYLRGIPAPFPAQFSNLWLLYVCRRGERYDVMDKIHKKMGPVVRIQPNHVSIADDEAIPIIY
GHGNGFLKSEFYDAFVSIRRGLFNTRDRAEHTRKRKLISHTFSTKSISQFEPYIHSNLELFVKQLDKLITSGTTKDNQGHQQALIDCLPWFNYL
AFDVIGDLAFGAPFGMLANGADVAEVRATPESPPIYASAIEILNRRGEVSATLGCFPQLKPYAKWLPDPFFSNGLNAVQNLAGIAIARVKARLD
NPPPEERMDLLARLMEGRDEKGEPLGREELTAEALTQLIAGSDTTSNSSCALLYHVTRTPGVLEKLQAELDASIPSHVSVPTFDMVRDLPYLNC
VINETLRYHSTSGIGLPRQVPEGSPGVTIRGHFFPAGSVLSVPTYSIHHSKEIWGPDADDFKPERWEDVTPRQKNAFIPFSHGPRACVGRNVAE
MEMKLIAATWARRYNVELKQEIMETREGFLRKPLGLDIALKIR

>9CYP53A19 (9543) Fve

MAVVDILFTWWSIPIATGVLIATYLYSYFVTYGHLRDIPAPFPAQFTNLWLLYVCRRGERYRVVDGIHKRLGPVVRIQPNHTSIADPDAIATIY
GHGNGFLKSDFYDAFVSIRRGLFNTRDRAEHTRKRKLISHVFSAKSISQFEPYIHANLELFVKQLDKLVASGQTTDKNGKRQALIDCLPWFNYL
AFDVIGDLAFGVPFGMLANGADVAEVRATPDSAPIYASAIEILNRRGEVSATLGCWPQLKPYAQWLPDPFFSNGLNAVKNLAGIAIARVKARLD
NPPSVERKDLLARLMEGRDEKGEPLGREELTAEALTQLIAGSDTTSNSSCALLYHVTRTPGVLEKLQAELDEAIPADVSVPTYDMVRDLTYLNN
VISETLRYHSTSGIGLPRQIPDNSPGVTIKGHYFPPGSVLSVPTYTLHHSKEIWGPDADDFKPERWDSLNELQKTAFNPFSHGPRACVGRNVAE
MEMKLIAATWARRYTPELKQEVMETREGFLRKPLGLDIALKMR

>10CYP53A20 (6367) Fve

MALTELLISPWAPLALAVALVAWYILPWISNSNLRGIPAPFLAQFSNLWLLSTCRRGKRYEIVDQVHKKLGVLVRIAPNHVSVADADAINTIYG
HGNGFLKADFYDTFVSIRRGLFNTRDRAEHSRKRKIVSHTFAPKSVLEFEPYIRQNLDIFINQWDRIASNKEADGYGSVDCLNWFNFLAFDIIA
DLAFGKPFGMLSTGADIAEVKVSPTSPTIYAPAVEIMNRRGEVSATLGCLPQLKPYAKYLPDPFFSQGLQAVENLAGIAIARVSERLERGGDST

RKDLLARLMQGRDEKGEPLGRDELTAEALTQLIAGSDTTSNSSCALLYHVVRTPGVMQKLYEEISAVVPEDVAIPDYESVKHLPYLGHCINETL
RIHSPSGIGLPREIPPNHKGVTLHGRYFGPGTVLSVPTYTIHHSTEIWGPDADEFKPERWESLTDKQKNAFIPFSYGPRSCVGRNLAEMQMRLI
AATWIKRYDVRLRQDIMETREGFLRKPMGLDVGLARR

>11CYP53A19 (14206) Fox

MAVVDILFTWWSIPIAACVLIATYLYSYFVTYGHLRDIPAPFPAQFTNLWLLYVCRRGGRYRVVDEIHKRLGPVVRIQPNHTSIADPDAIATIY
GHGNGFLKSDFYDAFVSIRRGLFNTRDRAEHTRKRKLISHVFSAKSISQFEPYIHANLELFVKQLDKLVASGQTTDKNGKRQALIDCLPWFNYL
AFDVIGDLAFGVPFGMLANGADVAEVRETPDSAPIYASAIEILNRRGEVSATLGCWPQLKPYAQWLPDPFFSNGLNAVKNLAGIAIARVKARLD
NPPSVERKDLLARLMEGRDEKGEPLGREELTAEALTQLIAGSDTTSNSSCALLYHVTRTPGVLEKLQAELDEAIPADVSVPTYDMVRDLTYLNN
VISETLRYHSTSGIGLPRQIPDNSPGVTIKGHYFPPGSVLSVPTYTLHHSKEIWGSDADDFKPERWDSVNNLQKTAFNPFSHGPRACVGRNVAE
MEMKLIAATWARRYTPELKQEVMETREGFLRKPLGLDIALKMR

>12CYP53A20 (10443) Fox

MAITELLISPWAPLALAVALVAWYILPWISNSNLRGIPAPFPAQFTNLWLLSTCRRGKRYEIVDQVHKKLGVLVRIAPNHVSVADADAINTIYG
HGNGFLKADFYDTFVSIRRGLFNTRNRAEHSRKRKLVSHTFAPKSVLEFEPYIRQNLDIFINQWDRIASNKDADGYGSVDCLNWFNFLAFDIIA
DLAFGKPFGMLSTGADIAEVKASPTSPTIYAPAVEIMNRRGEVSATLGCLPQLKPYAKYFPDPFFSQGLQAVENLAGIAIARVSERLERGGDST
RKDLLARLMQGRDEKGEPLGRDELTAEALTQLIAGSDTTSNSSCALLYHVVRTPGVMQKLYEEISAVVPEDVAIPDYESVKHLPYLGYCINETL
RIHSPSGIGLPREIPPNHKGVTLHGRYFGPGTILSVPTYTIHHSTEIWGPDADEFKPERWENLTDKQKTAFIPFSYGPRSCVGRNLAEMQMRLI
ATTWIKRYDVRLRQDIMETREGFLRKPMGLDVGLARR

>13CYP53D1 (9015) Fox

MSLDSLLLSPWAPLAVLLCVLLFYILPYFYTYRHLRGIPGPLLARFSDLWLLYICRQSKRSYTVYDLHERLGPVVRIQPNHVSIVDERAINLVY
GHGNGLEKSSWYDSSISLTRSIFTARKRAEHARKRRYIAHSFAPKSSRAAEGPIADKVELLVRKWDEIIDKGPQFDGFTQLECRRWFTYLSFDI
TGDLLFSEPFGMLENGSDLVKIDNKPRSYVSMMNSLAQRSAAVATLGVLPWLKPHAHHLPDPFFHRGMDGLQNLLGVTSAHVKERMAAGQDNHD
DWLSLLLRARDDQGELLKFEEIASESLTLFMAAIETVSNTLSAMMYYLATSPSSLQKLQAEVDSIDIPGTVPPFTNVRALPYLDAVLNETMCLH
SVLGIGLPREVLPGSKGVHLDSFYFPPGTVLSVPIYAIHRSRDIWGQDANNFRPERWEKLSDRQKTSFIPFGHGPAACVGRNLAEVEMKIIAAT
LIKRYNFCMMDSEIESTEGTTRKLIKVNIGIKRRQ

>14CYP53A12 (8190) Nfi

MITDLLTPQNTGFILLGLIAAYYIVPYLQKWHLHDIPSPSFAAFSNLWLLLQARRGRRFLKVDEAHKKYGKLVRIAPRQVSIADDAAIQAIYGH
GNGFLKSDFYDAFVSIRRGLFNTRDREEHTRKRKTVSHTFSMKSIGQFEQYIHQNVELFVQQWTKLAKLNGNPRSGYATIDALNWFNYLAFDII
GDLAFGAPFGMLEKSKDIAEMRKAPDSDPTYVQAVEVLNRRGEVSATLGCLPRLIPYAKYLPDRFFKDGIQAVENLAGIAVARVNERLKPEVME
KNTRVDLLSRLMEGKDSNGNKLGREELTAEALTQLIAGSDTTSNTTCAILYWCMSTPGVIPKLQKVLDEAIPDDVDVPTHAMVKDIPYLQWVIW
ETMRIHSTSAMGLPREIPPGNPPVTISGHTFYPGDVVSVPSYTIHRSKEIWGPDAEEFVPERWDPARLTARQKAAFIPFSTGPRACVGRNVAEM
ELLVMTGTIFRLFEFEMQQDGPMETREGFLRKPLGLIVGMKRRAVHASV

>15CYP53A3 (3425) Anid

MITDFLTPENITPERIALALLGLLAAYYVVPYLQTWRLSDIPAPGLAAWTNFWLLLQTRLGHRFISVDNAHKKYGKLVRIAPRHISIADDAAIQ
AVYGHGNGFLKSDFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSAKSIGQFEQYIHHNIENLVKQLTRISNLQRNPKNGYATVDALNWFNFV
AFDIIGDLAFGAPFGMLDKGQDIAEMRKSPDSPPQYVQAVEVLNRRGEVSATLGCYPALKPFAKYLPDRFFRDGLEAVENLAGIAIACVNERLK
PEVMANNTRVDLLARLMEGKDANGNKLGRAELTAEALTQLIAGSDTTSNTSCAILYYCLRTPGVIDKLHKVLDEAIPQDVEVPTHAMVKEIPYL

QWVIWETMRIHSTSAMGLPREIPEGNPPVEISGHIFKPGDILSVPTYTIHHSKEIWGADADEFIPERWAPERLTARQKAAFIPFSTGPRACVGR
NVAEMELLVICSTVFRMFDWELQQKGPMETREGFLRKPLGLTVGVKRRTIV

>16CYP53A12 (3003) Afu

MITDLLSLQNAGLILLGLIAVYYVIPYLQKWHLHDIPSPRFAAFSNLWLLLQARRGRRFLKVDEAHKKYGKLVRIAPKHVSIADDAAIQAIYGH
GNGFLKADFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSMKSIGQFEQYIHQNVELFVQQWTKLAKLNGNPRSGYATIDALNWFNYLAFDII
GDLAFGAPFGMLEKGKDIAEMRKTPDSEPTYVQAVEVLNRRGEVSATLGCLPRLIPYAKYLPDRFFKDGVQAVENLAGIAVARVNERLKPEVME
KNTRVDLLSRLMEGKDSNGNKLGREELTAEALTQLIAGSDTTSNTTCAILYWCMSTPGVIPKLQKVLDEAIPDDVDVPTHAMVKDIPYLQWVIW
ETMRIHSTSAMGLPREIPPGNPPVTISGHTFYPGDVVSVPSYTIHRSKEIWGPDAEKFVPERWDPARLTARQKAAFIPFSTGPRACVGRNVAEM
ELLVMTGTIFRLFEFEMQQDGPMETREGFLRKPLGLIVGMKRRAVHASV

>17CYP53A22 (2015) Ate

MIADLAAINPAYLLLAAVAAYYIVPYLKRWHLRSIPTPSVAGFTNLWLLIQARRGNRFEVVDNLHKKHGKLVRLAPRHVSIADDAAINAIYGHG
NGFLKALSDFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSMKSIGQFEQYIHHNIELFVKQWTRLSETQGNPRSGYATIDALNWFNFLAFDI
IGDLAFGAPFGMLEKGQDIAEMRKSADAAPTYVQAVEVLNRRGEVSATLGTLPALIPYAKYIPDRFFKDGIQAVENLAGIAIARVNERLRPEVM
ANNTRVDLLARLMEGKDANGNKLGREELTAEALTQLIAGSDTTSNTACAILYWCMSTPGVIDKLHKVLDEAIPADVDVPTHSMVKDIPYLQWVI
WETMRIHSTSAMGLPREIPAGSPPINISGHVFYPGDVVSVPSYTIHRSREIWGPDAEKFVPERWDPARLTPRQKAAFIPFSTGPRACVGRNVAE
MELLVMAGTVFRLFDFEMQQKGPMETREGFLRKPLGLIVGMKRRTPA

>18CYP53A13 (5958) Aor

MIAELLTPTGAAYVLTAAVIVYYILPYLQLWRLRDIPSPGFAAFSNLWLMLQYRKGNRFVTVDNAHKKYGKLVRIAPRHVSIADDEAIQAIYGH
GNGFLKADFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSMKSIGQFEQYIHGNAELFVKQWNRIADTQSNPKTGYATIDALNWFNYLAFDII
GDLAFGAPFGMLEKGQDIAEMRKSPNDKPSYVQAVEVLNRRGEVSATLGACPSLIPWAKYIPDRFFRDGLEAVENLAGIAVARVNERLRPEVMA
NNTRVDLLARLMEGKDSNGNKLGREELTAEALTQLIAGSDTTSNTSCAILYWCLRTPGVIEKLHKVLDESIPKDVDVPVHAMVKDIPYLQWVIW
ETMRIHSTSAMGLPREIPAGNPPVTISGHTFYPGDVVSVPTYTIHRSKEIWGPDAEQFVPERWDPKRLTARQKAAFIPFSTGPRACVGRNVAEM
ELLVIVGTVFRLFDFEIQQDGPMETREGFLRKPLGLMVGMKRRSVAV

>19CYP53A14 (2107) Aor

MDGDSCPSCCSTCTSLNVFLCPHIPILTLGLLLVYYVTGYLKRWHLRDIPGPFIAGFSRIWLIVQVRQGYRSLVVHDLHRRYGKIVRLAPNHIS
IADESAIQAIYGHGNGFLETDFYNAFLNVDWSIFTTRSRAEHTRKRKIVSHAFSARSLAQVEQHAHNNMEHLVRQWRKMIDSEEGLDDPYAVID
ARVWCNYLTFDIIGDLAFGAPFGMLERENATVSMRKAPENPEVTLDAVEVLNHRGDVSAAFGICPDLIPYAKWLPDPFFRQGAEAIANLAGVAG
AAVDRRLKMDTSMTEKRGDLLALLIDAEDQAGAKLGHRELTGEAVTLIAAGSDTSSSTLCALLYWVSSTPRVLWKLQNVLDEVIPVDIEVPYLA
MVKKITYLQWVIWEALRIHSTFGQGLPREVPPERGPVEICGHTFYPGDVLSVPGYTMHHSADIWGIDVEDFVPERWDPRRLTQRQKDSFIPFSE
GPRACIGRNLAEMELFVGCATLFRLFEFRVEGQGPLKVRERWLRKPVSLQVGIRRRYLDARSS

>20CYP53A13 (26719) Afl

MIAELLTPTGAAYVLTAAVFVYYILPYLQLWRLRDIPSPGFAAFSNLWLMLQYRKGNRFVTVDNAHKKYGKLVRIAPRHVSIADDEAIQAIYGH
GNGFLKADFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSMKSIGQFEQYIHGNAELFVKQWNRIADTQSNPKTGYATIDALNWFNYLAFDII
GDLAFGAPFGMLEKGQDIAEMRKSPNDKPSYVQAVEVLNRRGEVSATLGACPSLIPWAKYIPDRFFRDGLEAVENLAGIAVARVNERLRPEVMA
NNTRVDLLARLMEGKDSNGNKLGREELTAEALTQLIAGSDTTSNTSCAILYWCLRTPGVIEKLHKVLDESIPKDVDVPVHAMVKDIPYLQWVIW

ETMRIHSTSAMGLPREIPAGNPPVTISGHTFYPGDVVSVPTYTIHRSKEIWGPDAEQFVPERWDPKRLTARQKAAFIPFSTGPRACVGRNVAEM

ELLVIVGTVFRLFDFEIQQDGPMETREGFLRKPLGLMVGMKRRSVAV

>21CYP53A1 (50153) Ani

MLALLLSPYGAYLGLALLVLYYLLPYLKRAHLRDIPAPGLAAFTNFWLLLQTRRGHRFVVVDNAHKKYGKLVRIAPRHTSIADDGAIQAVYGHG

NGFLKSDFYDAFVSIHRGLFNTRDRAEHTRKRKTVSHTFSMKSIGQFEQYIHGNIELFVKQWNRMADTQRNPKTGFASLDALNWFNYLAFDIIG

DLAFGAPFGMLDKGKDFAEMRKTPDSPPSYVQAVEVLNRRGEVSATLGCYPALKPFAKYLPDSFFRDGIQAVEDLAGIAVARVNERLRPEVMAN

NTRVDLLARLMEGKDSNGEKLGRAELTAEALTQLIAGSDTTSNTSCAILYWCMRTPGVIEKLHKALDEAIPQDVDVPTHAMVKDIPYLQWVIWE

TMRIHSTSAMGLPREIPAGNPPVTISGHTFYPGDVVSVPSYTIHRSKEIWGPDAEQFVPERWDPARLTPRQKAAFIPFSTGPRACVGRNVAEME

LLVICGTVFRLFEFEMQQEGPMETREGFLRKPLGLQVGMKRRQPGSA

>22CYP53A21 (1341) Acl

MITEILTPQNTGYVLLGLLTAYYIIPYLQTWHLHDIPSPGFAAFSNLWLLLQARQGHRFLKVDEAHKKHGKLVRIAPGHISIADDGAIQAVYGH

GNGFLKADFYDAFVSIRRGLFNTRDRAEHTRKRKTVSHTFSTKSIGQFEQYIHHNIELFVKQWTKLSKLNGNPRSGYATIDALNWFNFLAFDII

GDLAFGAPFGMLEKGKDFAEMRKTPDSPPTYVEAIEVLNRRGEVSAALGCFPRLIPYAKWIPDRFFKDGLQAVENLAGIAVARVNERLKPEVMA

NNTRVDLLSRLMEGKDSNGNKLGREELTAEALTQLIAGSDTTSNTACAILYWCMQTPGVITKLQKVLDEAIPADVDVPTHSMVKEIPYLQWVIW

ETMRIHSTSSMGLPREIPPGNPPVTISGHVFYPGDVVSVPSYTIHRSREIWGPDAEQFVPERWDPARLTPRQKAAFIPFSTGPRACVGRNVAEM

ELLVMTATVFRLFEFEMQQDGPMETREGFLRKPLGLIVGMKRRAAHASV

>23CYP53A18 (8044) Cim

MLLSFLFNPYVLAGIFIFVFYIVPYLRLSYLRDIPSPFAAGFSNLWLLYQCRRGKRYQAVHDAHKKYGKLVRIQPDHVSVADADAIQTIYGHGN

GFLKSEYYDAFVSIRRGLFNTRSRAEHTRKRKTVSHTFSAKSVGQFEQYIHANLQLFFQQWTNISEVQRNPKSGYASIDALNWFNYLAFDIIGD

LAFGAPFGMLSKGRDVAEMRKSPDSPASYVPAIQVLNRRGEVSATLGCFPALKPFAKYLPDKFFRDGLEAVEHLAGIAVARVSERLRPEVMAKN

TRVDLLSRLMEGRDETGAKLGREELTAEALTQLIAGSDTTSNTSCAMLYWVLRTPGVIEKLQEALDEAVPAHVNVPSFSMVRDIPYLQWVIWET

MRIHSTSSLGLPREIPPNSPPVTIEGHVFHPGTILSVPAYTIHHSPEIWGPDVEEFVPTRWDPARLTPRQKAAFIPFSYGPRACVGRNVAEMEL

HCIAATVFKNFEFRLEQDGPMETSEGFLRKPLGLMVGIRRRQPNLN

>24CYP53A17 (5031) Ure

MLLAFLFNPYIIAGFTICYFYIVPYLQRWDLRDIPAPFPASLSHLWLLYQSRKGKRYQAVHNAHGKYGKLVRIQPNHVSVADADAIQTIYGHGN

GFLKSEYYDAFVSIRRGLFNTRNRAEHTRKRKTVSHTFSAKSIGQFEQYIHANLQLFLQQWTQICDLQRNPRSGYASIDALNWFNYLAFDIIGD

LAFGAPFGMLSKGRDVAEMKKSPNSPASYVPAIQVLNRRGEVSATLGCFPALKPFAKYLPDRFFRDGLEAVENLAGIAVARVAERLRPEVMAKN

TRVDLLSRLMEGRDETGAKLGREELTAEALTQLIAGSDTTSNTSCAMLYWVLRTPGVIEKLQEVLDEAIPAHVEVPTFSMVKDIPYLQWVILET

MRIHSTSSLGLPREIPQGSPPVTIQGHVFHPGTILSVPAYTIHHSSEIWGPDVEEFVPTRWDPARLTAQQKAAFIPFSHGPRACVGRNVAEMEL

HCIAATVFKNFEFQLEQNGPMETSEGFLRKPLGLLVGIKRRQLDPVVN

>25CYP53A23 (71345) Mfi

MFLSFLFTPWALLASPFLFYLLPFLRNWSIRDVPGPFLAKFTTLWYMYECRRCRRYYTVYKLHEKYGKFVRVQPNHVSIAEPEAIPIIYGHGTG

FLKSEYYDAFVSIQRGLFNTRDRAEHTRKRKTVSHTFSAKSVGQFEQYIHHNLELLAKRWDEIAKNTGAGKYTRFDALHWFNYVAFDIIGDLAF

GAPFGMLEKGADIAEVQLNPDGPVTYAPAIEVLNRRGEVSNAVGCWPAIKPYAKYLPDPFFSKGMEAIANLAGIATARVNQRLAAAERGEIDRV

DLLARLMEGKDENGNKLAKAELTAEALTQLIAGSDTTSNTSCALLFHCLKNPHVVKKLQAELDEALPSDDVPTYEQVKNLQYLDQVISETLRIH

STSSQGLPRVVPPGDGVEVAGRHFPPGVVLSVPAYVMHHSKEIWGPDADEFRPERWEKVTERQKLAFIPFSYGPRACVGRNVAEMELALIVATV

FRRYEFELYQDELETREGFLRKPLGLQVGMRKRS

>26CYP53A (2107910) Tte

MAVINLILSPWAPAALLVAAVVYYVYPYLVTYRHLRHIPAPFPAQFTNWWLLLVCRRGDRYATVDKVHKKLGPVVRIQPNHVSILDDEAIQAIY

GHGNGFLKSDFYDAFVSIQRGLFNTRDRAEHTRKRKIVSHTFSVKSVAQFEPYIHSNLELFVRQLDDLIARSTSPDGAAHLDCLHWFNYLAFDV

IGDLAFGAPFGMLSSGADMAEVRASPDSPPIYAPAIEILNRRGEVSATLGILPQLKPYAKYLPDPFFSKGLSAVEKLAGIAIARVKARLENPPP

ASRKDLLQRLIDGRDEKGQPLGREELTAEALTQLIAGSDTTSNSSCALLFHAARTPGVLARLRAELDAAVPADLVVPTFDLVRDLPYLSAVVNE

TLRFHSTSGIGLPREVPRDGQGVHIAGHYFPPGTVLSVPTYSIHHSKEIWGPDADEFRPERWDNLTARQKNAFIPFSYGPRACVGRNVAEMEMK

LIVATWARRYDVSLQQDHMDTREGFLRKPLGLKIALKRRQ

>27CYP53A (2301715) Mth

MAIVNFILSPWAPVALLAAAVVYYVYPYLVTYRHLRHIPAPFPAQFTNWWLLLVCRRGNRYETVDKLHKKLGPVVRIQPNHVSLCDDAAIQVVY

GHGNGFLKSDFYDAFVSIQRGLFNTRDRAEHTRKRKIVSHTFSVKSVAQFEPYIHSNLELFVRQLDNLIARSTNPDGAAHLDCLHWFNYLAFDV

IGDLAFGAPFGMLSSGADMAEVRASPESPPIYAPAIEILNRRGEVSATLGILPALKPYAKYFPDPFFSRGLQAVENLAGIAIARVKARLENPPP

SHRKDLLQRLIEGRDEKGEPLGRQELTAEALTQLIAGSDTTSNSSCALLYHAVRTPGVMQKLQAELDAAIPADMDVPTFDMVRDLPYLSAVVNE

TLRFHSTSGIGLPRQVPPDGQGVHFGGHYFPPGTVLSVPTYSIHHSKEIWGPDADEFRPERWERLTPRQKNAFIPFSHGPRACVGRNVAEMEMK

LIVATWARRYDVSLRQAHMDTREGFLRKPLGLEIALKRRKRD

>28CYP53C2 (130996) Pch

MAVIEALTQLDLKSWLLLIPALAIVAHILIWLLDPHGIRSYPGPLLAKFSDAWLGYVAAQGHRSEVVHDLHKQYGTFVRIAPNHLSIADPDALQ

VVYGHGTGTLKSNFYDAFVSIQRGLFNTRSRSEHARKRKIVSHIFSQKSVLEFEPHVRLYVKQLIQQWDRLYEAGAKGLVWLDCLPWYNYLAFD

IIGDLAFGAPFGMLLAARDAAPVAVDHEQAMASYGKEKSEVQYIPAVQVINDRGTYSASLGVLPPWMRPIVKLFPWFRRGQKAVKQLAGIAVAA

VAQRLTTPTDRVDLLGKLQEGRDDDGNLMGKEELTAEALTQLIAGSDTTSNSSCAITYYLAKYPDAQRKLQQELDEALGSDDEPVSTFDQVKRL

PYLQAVIDEALRIHSTSGIGLPRLVPKGGMTVCGRFFPEGTVLSVPTYTIHRDEEVWGKDPEVFRPERWFEQDKNAVQKTYNPFSFGPRSCIGR

NLANMELLIIVSSILRRYDFVLEDPDKPFDTMEGFLRKPVECVVGIRRRTL

>29CYP53C3 (110015) Ppl

MSAPAALFTSNLVYGLAVIPVAVLLVHFVPYLLDPHGIRAYPGPFLARLSDIWLGWIAAQGHRSETVHELHKQYGTFVRIAPNHVSISDPEAIQ

YVYAHGNGTTKSNFYDAFVSIRRGLFNTRSRPEHARKRKIVSHIFSMKSVMEFEPYTRMHVAQLLKQWDRLYELGIKGASGEEGEGWKGRDGRV

WLDCLPWYNYLAFDIIGDLAFGAPFGMLHACADAAPVATEHKDAMASYGADNAPKVTYFPAVQVLNDRGEYSASMGVLPPHWRPLVVRFIPWYR

NGGKAVKRLAGIAIAAVSKRLTAPTDRADLLGKLQEGKDDEGNPMGREELTAEALTQLIAGSDTTSNSSCALTYHLAANPRVQQKLQRELDEAL

GSDDDPVATYEQVKRLPYLEAVVNEALRVHSTSGIGLPRVVPEGGLSVCGRFFPAGTVLSVPTYTVHRDAETWGADVDAFRPERWEERDKNAVQ

KAFNPFSFGPRSCVGRNLASMELLIIIASILRRYHFVLEEPKHKLETKEGFLRKPVACKVGLRRRSA

>30CYP53D1 (108845) Ppl

MDSSTPPLPPLGSIILAREGLTGLLLPGLSVLACLIATVIASSILLPYFNDPYKLRAYPGPFFAKFTSAWLSWIIGHNRWSETVYHLHRQHGPI

VRLGPDNVSISDPSALAAIYGHSSGALKSTFYDAISSIRIRNLFNTRDRAEHSRKRRIEAHMFSPRGIRALEDTARVHFQVLVRQWDTLCAHTD

KAIRGSAEGTIGTVHWKVHSGRVWFDCMPWFTFWSFDTISDLAFGHPFGMLEAAKDTAKISKSNIKGMQAISQGNSHSDEAELELEEIPAIEML

AEHLDIIVSLAFLPAWLQPIVGRLPSVRYGYDAAPKLAGLAVAAVANRFASKTKIDRADMLSELLRGRDEDGKPYGPEELSAEAELLIIAGGDT

TANSSCATTYHLARNPRIRAKLQAELDAALEGIDSDVAPYDAVKDLPYLDAVINEGLRLHSTIGAGLPRVVPPGGLTVLGQHLKEGTVVSSPLY

CLHRNEVVWGENAHEFYPERWLEASADAKKEMMRSFAPFSVGPRACLGRNLALQQMHIVLATIFHRYNLVLESDAPLPLRDGFVRKPKNCVVGV

QRRK

>31CYP53D2v1 (56013) Ppl

MDSSTSLLPPLGSIILACEGLTSLVPLILSVMVCLIATVTISPTLLAYFNDPFELRAYPGPFLARFTSAWISWIISQNRWSETVDLMHRQHGPI

VRLSPDHVSVASPAAFAAVYGHSSGALKAPFYNAFANFKIRSIFNTRDRAEHSRKRRVEAHMFSPRSIRALEDTARVHFQVLVRQWDALCAPTG

KTVRGSAEGTLGTISWKVHGDRVWFDCMPWFNFWSFDTISDLAFGRPFGMLEAAKGSAHVSKSNTKSVQAVSQDTSHSNEAQSELLEIPAMEVL

SELLDFTVALAYLPAWVQPVFGRLPMFRDGYDAAPKLANLSLTAVANRVASQTDRADMLSELLRGRDEEGKPYGLEELSTEAELLIIAGGDTTA

NTSCATAYYIARDLQIQAKLQAELDVALDGVESDVAPYDAVKDLPYLDAVINEGLRLHSTIGAGLPRVVPSGGMTVLGQHLKEGTVVSSPIYTL

HRNEAVWGKNAYEFYPERWLEASADAKKEMMQSFAPFSVGPRACLGRSLALQQLHILLATIFHRYSLVLENNAPAQLPLRDGFARKPMKCIVGV

QRRK

>32CYP53D2v2 (48082) Ppl

MDSSTSLLPPLGSIILACEGLTSLVPLILSVLVCLVATVTISPTLLAYFNDPFELRAYPGPFLAKFTSAWISWIISQNRWSETVDLMHRQHGPI

VRLSPDHVSVASPAAFAAVYGHSSGALKAPFYNAFANFKTRSIFNTRDRAEHSRKRRVEAHMFSPRSIRALEDTARVHFQVLVRQWDALCAPTG

KTGRGSAEGTLGTISWKVHGDRVWFDCMPWFNFWSFDTISDLAFGRPFGMLEAAKGSAHVSKSNTKSVQAVSQDTSHSDEAQSELLEIPAMEVL

SELLDFTVALAYLPAWVQPVFGRLPMFRDGYDAAPKLANLSLTAVANRVASQTDRADMLSELLRGRDEEGKPYGLEELSTEAELLIIAGGDTTA

NTSCATAYYIARDLQIQAKLQAELDVALDGVESDVAPYDAVKDLPYLDAVINEGLRLHSTIGAGLPRVVPSGGMTVLGQHLKEGTVVSSPIYTL

HRNEAVWGKNACEFYPERWLEASADAKKEMMQSFAPFSMGPRACLGRSLALQQLHILLATIFHRYSLVLENNAPAQLPLRDGFARKPMKCIVGV

QRRK

>33CYP53D3 (60352) Ppl

MLTYDEALNFSTVLSALTCLSAGVITLSLLLPYFIDQLQLREYPGALLAKFTSGWISWIISRNQWSETVDRLHSAHSMGSFVRLAPNHVSVSGP

SAFEAIYSHPSSALKAPFYDIFSAGGAANIFTTRDRAEHARKRRVEAHMFSPQSIRTLESTVSVHFHALEDQWDALCAHIQKAGSGGAEGIIGS

VSWKVHDSRVWFDWFMFWSFDTIADLSFGRPFGMLVSAKDVVRIPKSNASGIQAIAEAASHSKKTELEMVEVPLIEVLVKRGKTIAALAYLPAW

AQPIIGRLPGFREGYGAIPKLNGIAIAAVADRLRSPNGRADMLTKLLEGRDGEGYRYGPQELSAEAKTLIAAGGDTTASASCAITYYIARDPRI

QAKLQAELDAALDGIGSEIAPYGTVKVLPYLEAVVNEGLRLHSGVGAGLPRVVPAGGMTILGHHLMEGTVVSSPIYTLHRSKAVWGANADEFYP

ERWIDASADTKKEMMSSFAPFSIGLRACIGRNLAMQQLQIVTATIFRRYSIVLQDDAPLEIQDSFGRKPKKCIIGVKRRDLKVL

>34CYP53D4 (55859) Ppl

MLSVLVCLIPTVIASTFLIPYFSDPYKLRAYPGPFLAKFTSVWASWVINHNRWSETVDLLHRKYGPIVRLGPDSVSIADPSAFAVIYGHSSGAL

KAPFYDAFANHRIRDLFNTRDRAEHSRKRRVEAHIFAPQSIRALEDTARVHFEILVRQWDAMCAHAEKAGRGSAEGAIGTVPWKVHDGRVFFNC

MLWFSFWSFDTIGDLAFGHPFGMLETGKDTAQTVKSDVRGMEAIAQATSNSEKTKLELVDIPAIEALTARADTLFVVAYLPAWLQPIVGHLPSL

QSGYNAAPKLAGLAVAAVANKFASKTDRADMLSKLLEGRDKNGNLYGPEELSAETWLLIIAGGDTTANTSCATTYYLARNPRIQAKLQAELDTA

LDGIDSDVASYDAVKDLPYLGAVINEGLRLHATVGVGLPCVVPPGGLTVLGHHLKEGSVVSSPIYSLQRSEAVWGENAREFYPERWLEASADAK

KEMMRSFAPFSVGPRACLGRNLALQQLYIMFATLFRRYDFVLENDAPLAVQDRLVQKSKECIIGVQRRKGLVSRQ

>35CYP53D5 (46728) Ppl

MLTFDEALNFSTILSALACLSAGVITLLLFLPYFIDQLQLREYPGALLAKFTSGWISWIISRNQWSETVDRLHVQHGSFVRLAPNHVSVSGPSA

FEAIYGHPSSAAKAPFYDIFSAGGAANIFTTRDRAEHARKRRVEAHMFSPQSIRTLESTVSVHFHALVDQWDALCAHIQKAGSGGAEGIIGSVS

WKVHESRVWFDCMPWFMFWSFDSIADLSFGRPFGMLVSAKDVVRIPKSNASGIQAIAEAASHSEKTELEMADVPLIEVLEIRGKTVAALAYLPA

WAQPIIGRLPGFREGYGAIPKLNGIAIAAVANRLRSPNGRADMLTKLLEGRDGDGHSYSPQELSAEARTLIAAGGDTTASASCAITYYIARDPR
IQAKLQAELDAALDGTGSEIAPYGAVKVLPYLEAVVNEGLRLHSGVGAGLPRVVPAGGMTILGHHLMEGTVVSSPIYTLHRSKAVWGANADEFY
PERWIDASADTKKEMMSSFAPFSIGLRACIGRNLAMQQLQIVTATIFRRYSIVLQDGAPLEIQDSFGRKPKKCIIGVKRRDLKVL

>36CYP53D6 (54877) Ppl

MDSSTPPLPLLGGIILTREGLTGLLLPGLSVLACLIATVIASLILIPYFNDPYKLRAYPGPFLAKFTSAWISWTISHNPISEIVDHLHRQYGPI
VRLGPNNVSIADPSGSAFSVIYSHSSGVTKSAFYDTFANFRIRNIFNTRDRAEHSRKRRVEAHMFAPQSIRALEETARVHFQVLLRQWDAMCAH
AEKAGRGRADGAIGAVPWKVHGGRVWFNCMIWFSYWSFDTIGDLAFGHPFGMLETGKDVAQIAKSNARGMQAIAQGTSDSEKATLELVDIPAIE
VLTARADALFVVAYLPPWAQKIVGRLPSFRSGYAAAPKLAGMAVAAVANRLASQYDRADMLSKLLQGRDEDGKPYSPEELSAEAWVLIIAGGDT
TANSSCALTYHLARNPRVQAKLQAELDAALDGIDSDVAPYDAVKDLPYLDAVINEGLRLHSTVGGGLPRVVPAGGLTVLGQHLKEGTVVSSPIY
TLHTNEVVWGENAHEFYPERWLEASADAKKEMMRSFVPFSAGPRACLGRNLALQQLHVMFATIFRRYSFALENDAQLTIQESFVRKPKNCFVGV
QRRK

>37CYP53C1 (5) Uma

MVETDLVPRIGAAIQWSVESPAHVVITLLGAVVLFHVVPYITNTACIKYPGPFFAKFTDFWLLRTALIGHRFEEVHKQHQKYGKFVRIAPNHVS
IADHEALQPIYGHGTGTLKPAYYDAFVPPRPFPRGLFNTRDRAEHTRKRKIVSHTFAPKTIVAFEPFIRREVQLLLERWDEFCDKATKDNTEGP
RGIKGRAWLDSLMWLNYFAFDTIGALAFGKTFGMLENGVDQAKVEYEDANGNKQVDYCSAVQIINERGEFSGTMGLAPVWMRPYLIKLPWFSSR
LKSVKKLTGIALARVNDRLQNGSEREDLLAKLQAAKDDRGEPMGKMELTAEALTQLIAGSDTTSNTSCAIVYHLATHPDKMRKLQAELDRELEH
AEEVPLHADVQELPYLQAVLSESLRYHSTSAIGLPRVIPAGGATVCGQQFPSGTILSVPAYTLHRDKSVFGADAEEYNPDRWLAPNAKRDFEKA
FIPFSVGPRACVGRNVAMMELSILIAAIFRRYDIVLAEPDKPLDTFEGFLRKPVKLEVGLKRRN

>38CYP53B3 (28617) Pgr

MLVIGLIAAFLEYSILLGLVGISCYYLTGYLRNKHQLNRYPGPFLAKFSRLWLGYATRFGNRYQIIHQLHQKHGRFVRIAPNELSIADPDAVHI
VLGHGTGTTKSKFYDAFVAIHRGLFNTRDRADHTRKRKIISSTFSQKSILEFEPYIADTLACFLRKIDQVASEPNLVQLPSHSWDSKHLNERWR
IIDILPWFNYLAFDIIGDLAFGERFGMIERGADIAAVEKEGKVIYLPAIQILNERGEFSATQGQFDHIKFSKKKKEKKKYIDPWFSRGAASVQN
LTGIATNQVNLRISQTGQSRRDLLARLQTGQDADGNPMGKDELIAEALTQLIAGSDTTSNSSCAILWWVVKHPEVHKRLMEELDEHLGTEEGVI
SYADCKELKYLNACINETLRIHSTSSIGLPRILPQTVSFKGHILPKGLVCSVPTFEIHHDPDVWGDPFTFRPERWLEPNAKDREKAFMPFSCGP
RSCIGRNLAMMELYMITSTIFKRYEFALVDPDLAELETREGFLRKVGFSLQQKT

>39CYP53B2 (32280)Sro

MTQSHVFTDFFEPTTLAVYFLAAPLGAVFLYLFVPHFTSLAPLRRFPGPFWAGYTRLWLARTARVGKRSELVHREHLKHGKFVRIGPNEVSIAD
PAALPIVYAHGSGSIKADFYDAFVASPVRGLFNTRNRAEHTRKRKIVSHTFAPKSVREFEPYIASTVNLLLKKWDQLAAKAQKSPPSGTGGERM
KGYAVIDSLDWFNALAFDVIGELAFGTPFGMVERDAADIVTITKEDGTVIHAGGVQILNMRGEYSATLGCLPPWSRKYMKYIDPWFARGLESVK
NLTGIARTRVNDRLEKGALDRKDILSHLQAGRDENGQPMSKDELTMEALTQLIAGSDTTSNSSCAILFQIVSTPHAHKKLQQELDEAFSGKGMS
GVLEYEDVKALPYLGACINEALRRHSTSGIGLPRIMMDDTEVLGEVFPKGTVLSVPSYTIHWSTEFWGPDAAEFKPERWLESEEKTRQLEKQLN
VFSFGPRSCVGRNVAMIELFCFMATLVYRYDFKLVDPNQKELEVVEGFLRKPTGCQIGFKLRDATQ

>40CYP53C (212558) Pca

MEQYHFILPQIDLKSIAVIRAAALLAAWVIPFFVDKYGLRGYPGPLLAEFSGFWLASQAYKGTTTSAVHALHQKYGLNFGAFDVIGDLAFGAPF
GMLEAEKDTVPVPVSEEQAMKSYGQKDMDLEWSTLPAIKLLNETIPRTSFLGCLPPQGKPLSEQDLTSEALNLIVAGSDTTSRSFDWCYRLPCC

AESGCPKRLQKELDDVLRVPNSTFNTDEVVASFDRVKNVAYLQDAINEGLRLHSTVGVGLPREVPQGGLTIAGKALLPGTHVSCPSYTLHRLKS
IWGDDADEFNPDRWTRRGDRNAILKYLAPFSIGPRACIGRNLVMMEMTICVATISHPYRLVLANPDQQLECSEGFVRKPKNVHVGSMQRRL

>41CYP53C (212559) Pca

MEQTHFLLPQLDLRLTAGLLAAALLAAWVVPFLIDRYRLKGIPGPLLAKFSCVWLASKAYKGTMSSAVHVLHEKYGPFVRISPKYVSIADPEAL
QAIYGYSSGALKTELYDAFVFFRPTMFSTRSRLEHSRKRKYTAHAMSMKYIMEFEPNVREHHHMLVKQLDTLCAAGAQGKDGILGTRPWTARDG
WAWFDCMPWFNYETFDIIGDLAFGAPFGMVEAGKDTASVPVSEKQAMKFYGQKGAEIEWSTAPAIKIVNEAVPWVFFLGCLPPHVRPLLSKLRS
FTTKASARALVKIAVAAVSKRLATGVTRRDFLSHLIAVRDDQGRPLTEQELTSEAISLIVAGSDTTSSSIAAIAYHVARNQDVQAKLQAELDDV
LGAPGSDSSTDDVVAPFDRVKNLTYLQDVINEGLRVHSTLGAGLPREVPEGGLTVAGKTLLAGTHVSCPSYTLHRLKSIWGDDADQFNPDRWTL
GDRNAMLKYFSPFSVGPRACIGRNLAMMELTICIATIFHRYRVVLASPDQRVSQLFNVLDKDVG

>42CYP53C (102576) Pca

MEQSHFILPQLDLKLITAILAASLLAAWIIPFLVDKYRLRGYPGPLLAKFSGFWLASKAYKGTTTSAVYALHQKYEGPFVRISPKYVSIADPEA
LQAIYGHTSGTLKTDFYDAFVTFRRNIFTSRSRLEHSRKRKYTSHAMSMKGITEFEPNVREYQHMLLKQLDTLCAVGAQGIDGVLGSCPWTTRD
GWVLFDCMPWLNFDTFDVIGDLAFGKPFGMLEAGKDAALVPVSEEQAMKSFGRQDTDLKWATIPAIKLLNETVPWTFFLGCLPPQARFLMSKLP
SFNAGTSRKLFVKLAVATVSKRLASEATRRDFLSHLVAARNDEGKPLSAQELTAEALNLIVGGSDTTSSSIGVVIYHVARNRDVQERLQKELDD
VLGVPNSTFSTDEVVAPFELVKNLTYLQDVINEGLRLHSTVGVGLPREVPEGGMTVAGKALLPGTHVSCPTYTLHRLKSIWGDDADEFNPDRWT
RGDRNMMLKYFAPFSIGPRACIGRNLAMMEMTICIATIVHRYRLVLANPDQQLECSEGFVRKPKNVHVGMQRRL

>43CYP53NS (152212) Pca

MSGTIVNLGLDLREWLSPARVVLCASALLIARVVLFVYAYIKARRQFPGPPVTNIWKGNLDETMTEDVHDKWRRWHRQYGPIYQTVRWNGLFSR
VIYVGDPRLIRKIANENWPKFPAQYAGFRPLSGSALFAQMDQARWKTQRRGLAPAFQPRTVHAQYPALHKYLLQFADTIDRSAAARGRAVDLAQ
LHVLLTLDFVGEVAFGAELRAVRDGAACRILQIFHAVLPELMKCGLFPLRAKVPVFESTRAMHRAIAELRGMARAAVEDARRAHEDSQEKDCGA
EKGKKIFEILAHHVPRRRRRSHGTHDDFCRGHRQLRNPAIHRKLRDELDALLPADCVVPSIEQVSRLPYLRLVIKETLRYNGPGFGTFRYTPAD
VEIEGVVLPANTTLALWNPQVHRCPNVWGADADTFRPERWMSTQGEGNEKAALPGSYFPFSYGPRKCMGEGLAMLEMSLTLATLFKRYDLELEP
GFEMDFQPSFTLCSRNGLPVYARLRK

>44CYP53C (183109) Pca

MAILEVLAQLNLTSWLVLIPALAIAAHVVHWLLDPHGIRSYPGPLLARFSDAWLGYVAAQGHRSEVVHDLHKKHGTFVRLAPNHVSISDPDALQ
VVYGHGTGTLKSDFYDAFVSIQRGLFNTRSRPEHARKRKIVSHIFSQKSVLEFEPHVRLYVNQLIRQWDRLYEAGAKGLSGDDGESGWTGRNGR
VWLDCLPWYNYLAFDIIGDLAFGAPFGMLLAARDAAPVAVNHEQAMASYGKEKSEVQYIPAVQVINDRGMYSASLGVLAPWMRPIVKLFPWFRQ
GQQAVKLLAGIAVAAVSQRLTTPTDRVDLLGKLQGGRDDDGNLMGKEELTAEALTQLIAGSDTTSNSSCAITYYLAKYPDVQRKLQQELDEVLG
YDDEPVSTYDQVKKLTYLPAVIDEALRVHSTSGVGLPRVVPEGGMTVCGRTFPEGTILSVPTYTIHRDEEVWGKDVEVFRPERWFSQDKNEVQK
TFNPFSFGPRSCVGRNLASMELLIIISSILRRYDFVLEEPDKPFDTMEGFLRKPVECLVGIKRRSL

>45CYP53C (101826) Pca

MIPFLVDKYGLRGYPGPLVAKFSSLWLASKAHKGKTTSAVHALHQKYGPFVRISSKHVSIADPEALQAIYGHNSGALKTDFYDAFVAFRHNIFT
SRSRLEHSRKRKYTAYAMSMKGIMEFEPNVREYQHMLVRQLDTLCAVGAQGIDGVLGSCPWTARDGWVLFDCMPWLNFATFDVIGDLAVGAPFG
MLEAGKDTALVPVSEEQAMKSFGQQDTDLEWATIPTIKLLNETVPWIFFLGCLPPQARFLMSKLQSFNAGASRKLFQKLGVAAVSKRLSSEATR
RDFLSQLVAARDDEGKPLSAQELTSEALNLIIAGSDTTSSSIGAIIYHIARNRDVQERLQKALDDVLGVPNSMFSTDEVVASFDLVKNLTYLQD

VINEGLRLHSTVGVGLPREVPEGGMTVAGKTLLAGTHVSCPTYTLHRLKSIWGDDADEFNPDRWTRGDRNMMLKYFAPFSIGPRACIGRNLAMM

EMTICIATMFHRYRLVLASPDQQLECSEGFVSKPKDVYVGMQRRV

>46CYP53C (256510) Pca

MEQVRSLLPQLDLKLASVLLVAAPLFAWAVPFLVDKHGLMAFPGPLLAKFSSLWFALKAYKGTTSLTVHALHERYGPFVRLSPQHVSIADPEAL

RAIYGHSSGTLKTELYDAFVTFFLARKRKYTAHAMSVKGIMQFEPNVREHQQMLVKRLDTLCTVGAQGVDGVLGSCPWAARDGWVWFDCMPWFN

FETFDIIGDLAFGASFGMLEAGKDTAPVPVYTDQAMKSYGQKDTDLEWSTAPAVQILNEAIPWFFFLGCLPPQARRLVSTLQSFNAGGSRNLIG

KIAVAAVSKRLTSEVTRRDFLSHLVAAHDDQGRPLSQQELTSEAISLIVAGSDTTSTSIAAITYHVARTQDVQAKLQEELDDALGVPDASSNAD

NVVAPFDLVKNLAYLQDVINEGLRLHSTIGVGLPREVPEEGLTVAGKALLPGTHVSCPLYTLHRLKSIWGDDADEFNPDRWARGDRKAMLKYFA

PFSTGPRACVGRNLATMEMTICIATIFHRYRVVLASPDQQLECHEGLVRKPNSVPVGMRRRV

>47CYP53C7 (118978) Bad

MAVLDYVLSPLESLGLASSLLIIPCLVLAGHFIYFVIDPHRIRSYPGPLLAKLTDAWLGYVAAHGHRSEVVHGLHQKYGKFVRIAPNHVSIADP

DALPIVYGHGNGTLKSNFYDAFVSIERGLFNTRSRHEHARKRKIVSHIFSQKSVLEFEPNVRTYVGQLIGKWDRLYENGAKGLSGDEGEGGWTG

RNGRVWLDCLPWYNYLAFDIIGDLAFGSPFGMLQACRDAAPVAVSQEDAMAGYGGKQCDVVYIPAVQILNDRGNFSASLGVLPPWMRPIVKQLP

WFKKGQKAVKDLAGIAIAAVAKRLTTPSDRTDLLGKLQQGRDDEGNPMGRPELTAEALTQLIAGSDTTSNSSCAITYHLAKNPEVQRRLQKELD

DALGAHADEPVVTFEDVKRLPYLQAVIDEALRIHSTSGVGLPRLVPEGGLTVCGQYFQEGTVLSVPTYTIHRDKEIWGEDCEAFRPERWFEQDK

NGIQKTFNPFSFGPRSCVGRNLANMELLIIVASILHRYDFVLADPEKPFDTAEGFLRKPVDCQVGIKKRAN

>48CYP53H1P (318949) Bad

MSSPEHLLPKADPKQFTSLATNLGVGTVSFGGVVWPGSHELLHFLHHLLRLDVHLIPYAHHRHQLWAMEQAPSRRQRPSSPQPDRWAKGDRATL

MKAFTPFSIGPRSCIGRNLASMELLIVVANVFHRFEAELAYPGQIMEVREGFLQHPLSVAVSLKRRGML

>49CYP53H2 (55123) Bad

MDAAWGQIVTSGVKAVLLLISFVLLVHLVPYALDRLSLSSHPGPFLASLSDLWLGWYAARGKMVRAVWDAHRVYGPVVRIAPNHVSIADVAALR

TVYGHSAGILKADLYDVFNPFGRTTLFTTRSREEHARKRKIVAHTFARKTVLEFEPIVQRYIHAVVRQWEHMCAVAADGKGGVIGNASWTSQDR

RAVFNTLKWYNFMVFDIVGDLVFRNPFGMTERGSDMALIAKHPDQVMASYNSITEEKIQYDAVNAVDAVTALNVVLTAVGMLRPFWRPFILKLP

MPWIRAGNEGVAKIAMMAVAAVAQCSQHPASRNDILAKYFDAIDDRGEKMHDLELSAEAVGLLIAGTDTTSNSLGALTFYLAQHPSAQSKLQAE

LDGALGAPHAFGNDDVESVSYENIKNLQYLQDVVNEGLRLFSIAGFGLPRIVPEGGLTILGHHFAPGAVISVPLYVVHRDKSVWGDDAEVFNPD

RWAAGDRVAMTKAFAPFSIGPRACIGRNLAFMELLLVVANIFHRFEAKIVSAEQVMEVRDGFLRQPLSSVVSVKRRGVF

>50CYP53H3 (318972) Bad

MDENFAIPLSSGLKALVFIVLLTHVVPFLRDEHDVRSHPGPALAKLSDAWLAWCAACGKINRSIYEAHKVYGPIVRIAPNHISIADVSALQPIY

GHGSGILKAESYDTFVAFDTPSLFTTRSRDEHARKRKVVSNAFAQKSVLEFEPVVKQYVGAVLKHWSHMCNAASSGDGGIIGDMKWTAQGGRAT

FNVIKWYNFMVFDIIGDLVFRAPFGMTEHGTDIARIAKNRDHAMASYDSGEVKLEYDTVNAVQALLSQNKFLTVVAQLPAYLRPLLQSVPLPWV

RSGAEAFQKFSTLSVTAVANSQQYPAPRNDILGRYFEATDEKGQKMGNHELSSEAVSLLIAGTDTTSNSAAALTFYLAHNPAAQAKLQVELDKA

LGPPCALDGDNDPAIVSYDQVKNLSYLHDVVNEGLRLFSAVGLGLPRVVPESGLTVLGRTFAPGTVVSVPTHVLHHDKTIWGDDAESFNPDRWT

EGDKTAMMEAFAPFSVGPRACIGRNLALMELLIVVANVFYYFDAAPTSSDQPMEIWDGFLRQAFKCDVSLKSRGVA

>51CYP53H4 (160054) Bad

MDPAVVLILSTGLKAAVALLAFAFLAHLVPFALDKFGVRQYPGPLLAKLSDIWLGWHAAHGKINQAVWNAHRAYGPIVRIAPNHVSVADASALH
QIYGHSTGILKADLYDAFVSFNRASIFSTRSREEHARKRKILAHTFSQKTTLEFEPVVRQYIGDMFKQWDRMCTAAVMGKGGVIGEMPWKDQDG
RAEFDTLKWYNFMVFDIIGDLVFRAPFSMTERGTDMARIVKRPEKAMSSYESVDTKLEYDTINAIEAVNARSAFLITLGVLPALWRPLIKKLPL
PWVRAGTEAFQKITKLTVTVVADRSHHSATRNDILAKYFDATDENGQKLDAQELSSEAITLLIAGTDTTSNSVAAMTFYIAHNPVVQARLQAEL
DEALGCPVCESREEPVLATYEQIKGLSYLQDVVNEGLRVFSTVGLGLPRIVPDGGLTVLGRTLSPGTVVSVPTYVLNRDKSIWGDDAEYFNPDR
WANGNKAAMSKAFAPFSTGPRACIGRNFAFMELLLVIANVFHRYEVVPGQSEPTMEVREGFLRKPLASTVAFKRRGAAL

>52CYP53H5 (65034) Bad

MDAIRDVLLPFGTVVAISTLLAYLLFSGFHRSYLSSFPGPLLARLSDAWIGWHTARGTVNRAVYEAHKAYGPFVRIAPNHISISHSSALQPIY
GHSVGILKSEFYDIFTSFNGTKSVFTTRSREEHARKRKILSHTFSQKSTLEFEPMMGQHIGDLIRQWDYMCAMAAEGKGGVVGETAWTSHEGR
AVVDTLDWYSFVIFDVIGELVFRIPFGMTNRGSDETLIVKHPNQTMALDESSSTKTEHDSVRAVQMMNERSALMVSLGMLPKYWRPLIERLPM
EWIRRGNAGVEKMSSLVMSAIADGSKNSSQRNDILSKYFNATDEDGRKMGISELYTEALVLLAAGADTTAHSLAALTFYLAQCPAAQTKLQVE
LDEALGWPNAVDDGNRPTIASYDSVKNLPYLQDVVNEGLRLFSAVGVGLPRVVPEGGLTILGRTLPPGTVVSVPAYVVHRDQAAWGDDVESFN
PDRWAKGDKTGMMRAFAPFSIGPRACIARNLASMEMLLFVANVFHRYEAKLASQLQMDIHDSFTRRPLASIVALKRRGL

>53CYP53H6 (142452) Bad

MALSFGAPDPTTLAVISLPLVFAILVHVLPYLADKLHLRRYPGPPLARFSDVWLAWHCARGSINRAVLEAHRTYGPVVRISPTQISVADVSALQ
PIYGHGSGAPKAESYDAFSGLGRPSIFTTRSREEHTRKRKSLAHTFALKTVLEFEPVVKEYVGSIIKKWDRMCAAATKGEGGTIGEMTWISQGG
RAVFDTLKWYNFMVFDIIGDLVFRAPFGMTERATEIAVIAKRRDKAIESYETSEQKLEYTTLPAVQAINKRGTFLATMGMLPKYWWPLLRKLPL
SWFSAGHAAAEMMLTLAITAVAHRSQHNAIRSDILGKYMEARDDRGQQLDNRELSSEALTLLIGGTDTTSNSAAALTFYLAQNTIVQARLQKEL
DEALGEPVYGDEDAERPVLVAYELIKNLPYLQDAVNEGLRLFSTIGLGLPRVVPESGLVVLGEAFTPGTVVSVPTYVTHHDEAIWGEDSWAFNP
ERWQTGDKAVMAKAFAPFSVGPRACLGRNLALMELLLVVANVFHRFEVRLADPAQTMEITDGFLRKPVSSVVALKRRVSA

>54CYP53H7 (358536) Bad

MTLNVPLPTAVWVMFLLSSLPVLAILFHVVSYLKDTLHLRRFPGPPLARLSDVWLAWHCGRGTINRAVLAAHRTYGPVVRIAPNHVSIADVSAL
HLVYGHGSGALKADFYDAFVGRATSSVFTTRSRHDHTRKRKALAHTFSPKSVLEFEPIIWEYLGAVIRRWDQMCTAASQGRGGVVGEMTWSSEG
GCAVFDAVKWYNFMVFDIIGHLVFRHPFGMTERATEMTLIVKQGRDDAMELQDKPGRELEYTSIPAVQSINTRSTFLATMGMLPKSWRPLLRKL
PLSWFNGGHLASEFMLSLAMTAVSHRLQHQTTFDDILGKYLEATDDRGQKMNDEELIAEALTLLIGGTDTTSSTVAALTFYLAQNPVVQAKLQK
ELDEAFGNPSAVAANDDDDGQPNLVKYERIKNLTHLQDVVNEGLRLFSTIGLGLPRVVPDGGLTVLGNTLAPGTVVSVPTYIIHHDEEIWGNDA
WSFNPNRWQTRDKDTMSKAFAPFSLGPRACIGRNLAFMELLLVVANIFHRFEASSPYPGREAHIHDGLPSKPLSSVIALRRRSGV

>55CYP53C (118598) Csu

MDWRMMIDGFTTLTSRETYTFSGFVSILLPLLFLCLSITLLLFPCVTYLLDFYGLRSYPGPFLAKFTDLWLAYKVWEGNRSPDIHLLHKKHGPF
MRIGPNHISVVSPAAISTIYSHIDPLPKSAFYDGLATFSVPDIFTTRDRVTHGRKQRMVSHLFAPKTVRLFEEDVQKYVGQLVAQWDDMCARAK
DGVVLTGHNGAMEWSTREGKVWFDCMPWFNLLAFDTISDLAFGSPFGMLIAGRDTARVARSVDIAMKNLGVAQTAQESDRIYEEEDIPAISANN
TRSDFLMFLAYLPEWLRPIVVRLPLFDDGVAAAGKIMSMAVTTVLRRLHALSSDSGDEKKNYEDFLIKLLQGHNDDGNRMGPEELTSEAQVLLI
AGSDTISNSTCATVYWVARHLNVQRNLQSELDGALADVSSDEDSFVAPIDKIDNLPYLNAVVDEGLRVHSAVGANLPRTVGPEGATVLGHSFQE
GTILSVPAYSAHRDEQVWGLDCEEFRPERWLEADREQQELMKKAFIPFSVGPRACVGRNLAKMQLLINIATIFRLYKVVLADQDKELEVFDNFV
RKPLNCRVGMQRRDLSSKREWSI

>56CYP53C (116910) Csu

MEFLHLSFDWTTALLVLAVGVALVHVVPWLLDPHGIRSYPGPLLAKFSDAWLGWVAAQGHRSEVVHELHKKYGTFVRLAPNHVSISDPDAIQIV
YAHGNGSLKSNFYDAFVSIQRGLFNTRSRPEHARKRKIVSHIFSQKNVLEFEPHVREHIRTLISQWDRLYELGKKGLSGTEGEGGWQGKNGRVW
LDCLPWWNYLAFDIIGDLAFGAPFGMLHACADSAPVAISHEAAMKNYGDDAAPEVEHFPAVQVLNSRGEYSASMGVLPPHWRPLAKRIPWFRRG
NQAVQRLAGIAVAAVAKRLSAPSDRTDLLSKLQEGKDDEGKLMGKPELTAEALTQLIAGSDTTSNSSCAITYHLAANPHVQEKLQAELDAALGD
GDPVATFDQVKRLPYLEAVINEALRIHSTSGIGLPRIVPQGGLTAAGQYFPEGTVLSVPTYTVHRDKEAWGEDADLFRPERWFEHDEKTLQRAF
NPFSFGPRSCVGRNLANLELLIIIASILHRYHFVLEDPEKHFDTREGFLRKPVECRVGIKRRHD

>57CYP53C (112429) Csu

MILDVVHSVLERPLLLVVLLPSVLLAVALVAHLRDPHHLRSYPGPFLASLTDLWLAYKVWVGDRSPGVHELHKKHGTFLRIGPNHISIASPAAL
GVIYSHSHPLLKSDFYDGLATFSAPGTFTVRDRVAHARKRRVVAHLFAPKTVRMFEGALHKYIGQLVQQWDGMYKNVETALPGTATAGKAGDMS
WIVRDGRVWFDCMPWLNFLAFDTIGDLAFGSPFGMLVSGKDTARIAKSLKAAMQTLGSTPSATEKPSTIEEEDIPAISSINRRSEFLIAFASLP
AWIRPIVKRLPMASDGMAATREIMSMAVTTVSRRVRTLYDGDAEERQRPDFLTKLLEGRDEEGSPLSPDELSSEAQTLLIAGSDTISNSTCAIV
YWIARNPDVQKKLQAELDAALADAGEGPIAPVEKTERLLYLNAVIDEGLRVHSTVGANLPRVVGPEGVTILGHTFTEGTVVSVPAYSTHRDENI
WGHDAEFFRPERWLEADKEKRDAMNKAFVPFSVGPRACVGRNLAKMQLLVNIATVFRLYDIVLENPDLPLPVHDNFIRKPLNCHVGVKRRS

>58CYP53C (151209) Csu

MLPAELVDCLPGSLPLLFGLFLTALLAVFFAPYILDRHRLRSYPGPFLARFSDLWLASQVWKSHRSEEVHRLHKKYGKCRFLRIGPNHVSVADP
AAIPILYSHSNPLMKSDFYDGFTTFRTPGIFVERDRVAHARKRRVVSHLFAPKTVKAFEPAVQNYVGQLVRQWDRLCKNADAQSVSVMSGVLGS
MTWRAYDGCVWFDCMPWLNFLSFDTIGDLAFGKAFGMVESGKDIARVAKDYTDAMRTYNAKQELPEWTPAYEEEEIPAISSLKERSKYMIFVGT
LPKSLRGIMRFLRSRDHQVMALRRIMSMATSSVARRIRAGREQDRDDFLARLLQARDDDGNPLSPDELSSEAQTLLTAGADTISNSTCATVFWI
ARAPPVKARLQAELDATLGISSTDLGSPVAPIDKIEHLPYLNAVIDEALRIHSTVAAGLPREVGPGGLNVLGHFFPEGAVLSVPTYSAHRDESI
WAPDPDAYRPERWIEADKEKREAMNKAFIPFSVGPRACVGRNLAKIQLLINVATIFRCYDVVLEKPHDPMPVHDDFTRKVIECFIGLKRREI

>59CYP53C4 (47512) Gsp

MSFLNPLLALDFATWAALGLAAIVLFHLVPYLVDSHHIRGYPGPLLAKFSDVWLGYVAAQGHRSEKVHELHEQYGTFVRIAPNHLSISDPDALQ
IVYGHGTGTLKSTFYDAFVSIQRGLFNTRSRVQHARKRKIVSNIFAQKNVLEFEPHVRQHLGTLFQQWDKLCDGGKKGLSGTEGEGGWHGSDGR
VWYDCLPWYNYLAFDIIGDLAFGAPFGMLIACKDSAPVAVSCEAAMASYGSAASSKEIQIEHFPAVQVLNDRGEYSASMGVLPPHWRPLAKRIP
WYARGNQAVQRLAGIAVAAVAQRMASPSDRVDLLAKLQEGRDDNGDPMGREELTAEALTQLIAGSDTTSNSSCAITYWLARNPAAQRKLQAELD
GALGSNDDPIASFEDVKRLPYLDAVINEALRIHATSGIGLPRLVPEGGLTVCGRFFPEGTVLSVPTYTIHRDREVWGEDVDAFRPERWFERDKN
LIQKTFNPFSFGPRSCVGRNLANLELLVIVASIFRRYEFVLEDPAAELDTREGFLRKPVECKVGMKRRNA

>60CYP53C4 (GL08839-P1.1) Glu

MSLLDPLLAFDFATWAVIGLSAIVLFHLVPYLVDSHHIRGYPGPLLAKFSDVWLGYVAAQGHRSEQVHELHKQYGTFVRIAPNHLSISDPDALQ
VVYGHGTGTLKSTFYDAFVSIQRGLFNTRSRVQHARKRKIVSNIFAQKNVLEFEPHVREHLGTLFQQWDKLCDGGKKGLSGTEGEGGWHGGEGR
VWYDCLPWYNYLAFDIIGDLAFGAPFGMLLACKDSAPVAVSCEAAMASYGSASSSKEIQIEHFPAVQVLNDRGEYSASMGVLPPHWRPLAKRIP
WFAKGNQAVQRLAGIAVAAVAQRLGSPSDRVDLLAKLQEGRDDNGDPMGREELTAEALTQLIAGSDTTSNSSCAITYWLARNQAAQRKLQAELD
GALGSASDDDSIASFEDVKRLPYLEAVINEALRIHATSGIGLPRLVPEGGLTVCGRFFPEGTVLSVPTYTIHRDREVWGEDVDAFRPERWFERD
KNLVQQAFNPFSFGPRSCVGRNLANLELLVIVASIFRRYEFVLEDPTAELDTREGFLRKPVECKVGMKRRNV

>61CYP53C9 (27837) Pbr

MSSLLTSFSLDNVTNLLLVIPGLLVLGHVVFYLVDPYKIRSYPGPLLARFSDLWLGRVAAEGHRSEIVHKLHQKYGTFVRLAPNHVSVSDPDAL
QVVYAHGNGTLKANFYDAFVSIQRGLFNTRSRPEHARKRKIVSHIFSQKSVLEFEPNTRLYVRQLIAQWDRLCELGAKGLSGDEGEGGWKGRNG
RVWLDCLPWYNYLAFDIIGDLAFGHPFGMLQACQDAAPVAVSQEAAMAAYGEGKQFEVTNIPAVRILNDRGMFSASLGVLPPWMRPLAKQLPWF
KKGNAAVKTLAGIAVAAVARRLATPVDRVDLLGKLQDGRDDEGNPMGREELTAEALTQLIAGSDTTSNSSCAITYHLAKNPDVQKKLQAELDEV
LGNDDDPVSTYEEVKKLAYLQAVIDEALRIHSTSGIGLPRVVPEGGLTVCGQFFPEGTVLSVPTYTIHRDTHVWGDDVETFRPERWFEQDDKLI
QKTYNPFSYGPRSCVGRNLASMELLIIISSILRRYEFVLENPSKPLETLEGFLRKPVDCVVGIRRRSL

>62CYP53C (194181) Abi

MILDRLSDTFLELYDKLDLQIVAFAVPGTFLAFHLLPWLWDPHGLRAYPGPFIAKFSDIWLTCVSKGAHRSELVHEAHLKYGPVVRIAPNHLSI
ANPEALQIVYAHGNGALKSIFYDAFVSIRRGLFNVRDRNEHTRKRKIVSHIFSQKNVLEFEPHIRMYVAQLQNQWDRLYDMAVKGMSGNDGEGG
WEGRDGRLWLDCLPWANYLAFDIIGDLAFGEPFGMLQAAKDSAVVPKDQKSMMKSYGKEDASIEVMEIPAVQILNGRGEFSLTMGTLPPYWRPI
ARRLPGFRQGAQDVKNLAGIAIAAVAKRLATPTDRNDLLSNLQAGRDSEGKPLGREELTAEALTLLIAGSDTTSNSTCAILYYLARNRGAQEKL
QKELDEHLGTENEFTATEAQVKNLPYLDACINEGLRLHSTSSVGLPREVPEGGMMVCGQFFAEGTVLSVPSYTIHRDRGVWGEDFEAYRPERWF
ERDQTLMQKTFNPFSYGPRACVGRNLASMELLIILASIMRRYDIVLEDPDLILDTREGFLRKPLACRVGIKRRDI

>63CYP53C (194303) Abi

MIIETFKNFVAQADTTTLACAIPAAVVLFHVIPWLTDSHSLRKYPGPFFAKFSDFWLAFTSRGGRRSEIIHDYHKKFGPVVRIAPNHVSISDPD
ALNAVYGHGTGTLKSEFYDAFVAMDRGLFNVRDRHDHTRKRKVISHIFAQKSVVAFEPKIAIYVTQLLNQWDRLYDMAVKGGSGNEGEGGWKGK
DGKLYLDILPWMNYLAFDTIGDLAFGEPFGMLAAAKDMAVVPKDQQSAMNSYGKETKEEDILTVPVIEAFNNRGEFNLVMGSLPLHWRPLARRL
PGLAQGSRDFKTVAGIAVAAASKRLSSSTDRIDLMSKLQNSRDSNGNPMSREEMTAEALTLLVAGSDTSSNACAAFLYHVAANPSVQDKLHQEL
DEQLGTEDELVATAEQIKRLTYLEACINEALRIQSVSGIGLPRVVPEGGLEVLGNFFPEGTVLSVPSYSVHRDTKSWGDDTETYRPERWFERDQ
AAMNKAFNPYSVGPRSCVGRNLAAMELSIILASIMRRYEFVLKDEDKPLVISEGFLRKPLSVDLGIKRRDV

>64CYP53C (1179842) Sla

MNTILQLNPFQDFHFSFTTALAGVPIIFILVHVFPYLADPFKQRAIPGPLLAKFSDAWLGWVSSQGHRSEVVHKMHLKYGTFVRIAPNHVSVAD
PDALQVVYAHGNGSLKANFYDAFVSIQRGLFNTRNRNEHARKRKIVSHIFSQKNVLEFEPNVRLYVGQLISQWDRLYDSAAKGASGTEGEGGWF
GKDGRLWLDSLPWYNYLAFDIIGDLAFGSPFGMILNAKDSAPVAVSQKDAMKSYGSESTYEVIEIPAVQILNDRGEFSASMGVLPPHWRPLVRL
LPWYRKGGKAVKNLAGLAVAAVAKRLTTPTDRVDLLSKLQEGRDDEGKLMGREELTAEALTQLIAGSDTTSNSSCAITYYLAQNPDAQEKLQKE
LDEALGDDDHPVSTFEQVKRLPYLEAVINEALRVHSTSSIGLPRIVPEGGLIVQGQHFPQGAVLSVPSYTIHRDTTVWGADPDQFRPERWFECD
HAAIQKTFNPFSFGPRACVGRNLASMELLIIISSILRRYHFVLADPEKPFDTREGFLRKPQECRVGIKRRQT

>65CYP53C (94174) Shi

MLASLVNAVTNVDGKTCLVAVPVLLVLAHVIPWLVDVHGIRSYPGPFWARFTDLWLGWVAAQGHRSEVVHEMHNKYGPIVRIAPNHISISDPEA
LQIVYAHGNGSLKSNFYDAFVSIHRGLFNTRDRAAHARKRKIVSHIFSQKNVLEFEPHVREYVKSLIAQWDRLYDLAVNGESGTEGEGGWVGRE
GRLWLDCLPWYNYLAFDIIGDLAFGSPFGMLQAAKDSAPVAKSAKDAIAAYGQDEAKVEVVHIPAVQILNDRGEFSASMGVLPPWLRPYVKRYI
PWFSKGDQAVKNLAGLAIAAVSKRLNQPTYRVDLLSKLQEGKDDEGRPMGREELTAEALTQLIAGSDTTSNTSCAITYYLAANPAVQEKLHVEL
DAALGNEDDPASTFEQTKNLKYLQAVIDESIRLHSTSGIGLPRIAPEGGLTVCGKYFPEGTILSVPSYTIHRDVDVWGYDVEAFRPERWFERDA
EMIQKAYNPFSFGPRACVGRNLASMELLIIISSILRRYDFVLEDPTKPFATKEGFLRKPVDCKIGIKRRNL

>66Cyp53C (37267) Tve

MAIFELLASLDLPSWAAIVLAAVVLVHLVPYVLDPHGIRAYPGPFWAKLTDLWLGKIAADGHRSERVHDLHKKYGPFVRIAPNHLSISDPDALP
VVYGHGTGTLKSDFYDAFVSVQRGLFSTRSRPEHTRKRKIVAHSFSQKSVHEFEPHVRENLSKLFKQWDTLCEGGAKGLSGNEGEGGWQGREGR
VWYDCLPWYNYLAFDIIGDLAFGAPFGMLTSGKDSAPIAVSQVDAMAAYGQGGTLKVKHVPAIQVINDRGEYAASVGVLPPHWRPFVKRLPWYN
TGDKAVQNLTGMAIAAVARRMEESDSDHRDLLAKLREARDEDGNPMGREELTAEALAQLVAGSDTTSNSSCAITYYLAKHQRVQEKLQQELDEA
LASEEDEVALFERVKHLPYLEAVINEALRIHSTAGVGLPRVVPAGGLEVCGKFFPEGTVLSVPGYTIHRDKAVWGDDADEFRPERWFGKDKAAL
QKAFAPFSVGPRSCVGRNLAHLELLTFVASIFRRYSFVLENPDEPLPTNEGFLRKPLKCNVGMQRRNV

>67CYP53C (129211) Tve

MALLGILSSLDGPSWAALVFAAVVLVHVVPYLLDPHGFRSYPGPFLAKLSDFWLGKVAADGHRSERVHELHEIYGNWTFVRIAPNHLSIADPDA
LQIVYGHGTGTLKSDFYDAFVSIQRGLFNTRSRTDHARKRKIVSHIFSQKNVLEFEPHVRVHLIQLFKQWDRLCAGGARGEAGDEGEGGWRGRD
GRVWYDCLPWYNYLAFDIIGDLAFGAPFGMLTSCKDSAPVAVSQDDAMATYGKDAAYKVEHFPAVQVLNDRGEYSASMGVVPPRWRPLVKRLPW
YNKGNQAVQRLAGIAIAAVARRLSVPESDRHDLLEKLQEGRDDNGDPMGRAELTAEALTQLIAGSDTTSNSSCAITYYLAKYQHVQEKLQKELD
DALGGEDDSVASYEQVKRLPYLDAVINEALRIHATSGIGLPRLVPAGGLEVCGRWFPEGAVLSVPTYTIHRDKAVWGDDVEEFRPERWFEQDKV
AVQKTFNPFSFGPRSCVGRNLANLELLVIVASIFRRYHFVLEDPSAPLATNEGFLRKPLKCIVGMKRRNV

>68CYP53C (104855) Wco

MSLLGDIAPQTLDYVGKLTFWLASIPIVIIAAHIVSYLLDPLGLRAFPGPIFARFTSGWLPWIISQNRWSVTVDRLHQKYGTFVRLSPNHVSIA
HPAALPAVYGHGSGAPKAPYYDGFVNFKSRNMFNTLSRSEHARKRRIESHMFSPQSVRALEGTASVHHGNLVSQWDKLYSYVKQAEGGGAKEGQ
LGASVWKVEGGRVWFDCMPCDLAFGAPFGMILAAKDTARFAKSVMAGMAAFGTSSKTSEYAFETDETPVTKLMAERADLVATIGWLPEYWQPIV
RMLPAFRGGRKSTPQLAGLAVAAVAKRLSKPDAREDMLNRLLNARDEDDEPLSREELSAEAAMLIAAGADTVANTSCATTYYLARDQRVQAKLQ
AELDEALKSVDSVAAPYDAIKHLPYLDAVVNEGLRLHATIGAGLPRVVPEGGLTMLGHTFKEGTWVSVPVYHLHRDESIWGENASEFYPERWIE
ASGERKKAMLDAFAPFSVGPRACIGRNLALMQLHKTLATLFYRFSFVLESSDPVRLRSCIRFCSTADSVRNQLPVQDGFARKPRRCMIGIRSRK
L

>69CYP53C (154237) Wco

MYGIFVCLSPNHISIVHPVALPAVYGHSSGALKAPFYDAFASFKTRNMFNTISRTGHTRKRRIESQIFSQQSVRELEGTTRVHHIDLVSQWDKL
YSYVKRAESDGVREGFNYWSFDIIGDLAFGDPFGMILAAKDTARSAKSVNASLATFSTSSNTKKFAFEMEELPVTEIIGQRGELITMLGWLPKY
VRPIILMMPGFRSNLQAIPKVAGSAVTAVAKRMNDLDAHADMLNRLLDARDENGEPMSPEELSSEASLIIVAGAITVANTSCAITYYLARDQRV
QAKLQAELDDALKADDSIVAPHDAIKHLPYLDAVVNEGLRLDSPAPGSAYLRGENASEFYPERWVEASGDQKKAMLDAFVPFSIGPRACIGCNL
ASMQLHTVLATLFHRFNVVLESDDPVRSQV

>70CYP53c (27029) Wco

MSGLLAPLNLHPAALLLLIPAAVLAVHFVPYLLDPHGIRSYPGPFLAKLSDAWLGWVAAKGHRSEVVHQLHQRYGTFVRIAPNHVSISDPDALS
EVYAHGNGTMKSNFYDAFVSIQRGLFNTRSRPEHARKRKIVSHIFSQKSVLEFEPYTRQHVGALFKQWDRMCELGTKGLFGEEGEGGWHGRDGR
VWFDCLPWYNYLAFDIIGDLAFGAPFGMLQACADAAPVVVSHADAMASYGKGDAPEVAYFPAVQVLNDRGEFSASLGVLPPHWRPLVVRFIPWY
RNGNKAVKRLAGIAIAAVAKRLTAPTDRSDLLAKLQEGKDDDGNPMGREELTAEALTQLIAGSDTTSNSSCAITYYLAANPLVQQKLQRELDEA
LGNDDDPVAMYEQVKRLPYLEAVINEGLRLHSTSGIGLPRIVPEGGLTVRGQFFPEGTVLSVPSYTIHRDREVWGADVDAFRPERWMELDKNAV
QKTFNPFSFGPRSCVGRNLASMELLIIIGSILRRYHFVLEDADKKFDTREGFLRKPVECKVGMKRRSL

>71CYP53C (138864) Wco

MSLLRDITVQALSFVDKLIFWLALIPIIGIAMYIVPYFLDPLGLRAFPGPILAKFTIGWLPWVVSQNRWSLTVNRLHQKYGIFVRLSPNHVSIA
HPAALSAVYGYSSGVTKAPYYDVFGDFRAKNLFNIISRTEHTRKRRIVSPMFSQQSVRALEINARVHHNNLVSQWERLYSYVKQAEKGGIREDK
LGASAWRVEDGRVWFDCMPWFSYWSFDTIGDLAFGSPFGMLSAANDTVRVAKSVKASLATFGTSSNAGEFGFETEEMAIEKMLSERGHLISILG
WLPEYWRRIVQMLPAYRFGTEAAPKMAGLAVAAVGKRLNNPAAREDMLNRLLDARDENGKPMSPEELSADAFQLIVAGADTTANTSCATTYYLA
RDQRVQAKLQAELDEALKSIDSAVAPYDAIQNLPYLDAVVNEGLRLYTTVGAGLPRVVPEGGLTVLGHTFKEGTWVSVPIYRLHQDESIWGENV
NEFYPERWIEASGDRKKAMLDAFAPFSMGPRSCIGRNLALMQLHIVLSTLFHRFDFVLESDDPLHVQDSLIRKPKWCMIGIKSRKL

>72CYP53C (138909) Wco

MSLLRDITVQALPFVDKLIFWLALIPIISIAVYTVPYLLDPLGLRAFPGPILAKFTIGWLPWIVSQNRWSLTVNRLHQKYGIFVRLSPNYVSIA
HPAALSAVYGHSSGATKAPYYEVFGDFRARNLFNILSRPEHARRRRLEAHMFSLQSVRALEINARVHHNNLVNQWERLYSYVKQAEKGGIREDK
LGASAWRVEDGRVWFDCMPWFSYWSFDTIGDLAFGSPFGMLSAASDTVRVAKSVKASLATFGTSSNAEEFGFETEEMAIEKVLNERGHLITILG
WLPEYWRRIVQMLPAYRFGMEAAPKMAGLAVAAVGKRLNNPAAREDMLNRLLDARDENGKPMSPEELSAEAFVLIIAGADTTANTSCATTYYLA
RDQRVQAKLQAELDEALKSIDSAVAPYDAVKNLPYLDAVINEGLRLHATIGAGLPRVVPEGGLTVLGHTFKEGTWVSVPVYHLHRDESIWGENA
SEFYPERWIEASGDRKKAMLDAFAPFSMGPRSCIGRNLALMQLHIVLSTLFHRFDFVLESDDPLPVQDSFVRKPKRCMIGIKSRKL

>73CYP53c (138853) Wco

MPLLDRITTHALDYVCMPTFWLALIPAVLTTVYIVPYLLDPLGLRTFPGPIFAKFTSGWLPWVVSQNRWSVAVDSLHRKYGIFVRLSPNYVSIA
HPVALPAVYGHSSGALKAPFYDALSGFKTRNMFNTLSRTEHARKRRIESHMFSPQSVRALEGIARMHHNDLVSQWDKLYSYVKRAESGAAREGS
LGECAWKVKDGRVWLDCMPWFNYWSFDTIGDLAFGAPFGMILAAKDTARVAKSVKASLATFGTPSQTGKFAFETEELPVIKIVGQRAELVAMFG
WLPEYVWPIILMMPGFRGSRRAIPQVSGLAVAAVAKRMNNPDARADMLNKLLDARDENGEPMSPEELSSEAFLLIVAGSDTVSNTSCATTYYLA
RDQRVQAKLQAELDDALKAVDSVVAPHDAIKHLPYLDAVVNEGLRLHSAVGAGLPRVVPEGGMTVLGHTFKEGTWVSVPIYHLHRDENIWGENA
SVFYPERWIEASGDQKKAMLDAFVPFSIGPRACIGRNLALMQLHIVLATLFRRFNVVLESDDPLPVQDSFGRRPKQCMVGIKPRNL

>74CYP53C (77097) Wco

MSLVDDTTAHVPSYPAFWIALIAITVFAVQFVPYLLDPLGLSSFPGPVLAKFSNVWLPWIVSQNRWSVTVDQLHRKYGKTLIYCGAGTFVRLAP
NHVSIAHPAALPAVYGHSSGTLKAPLYDVFGPFRARSIFSTRSRTEHARKRRIESPMFSPQSVRALEGTTRVHHSDLASQWDNLYSYVKQAENG
GSREGMLGESAWKVEDGRVWFDCMPWFNYWSFDTIGDLAFGAPFGMLLAAKDTARVAKSVKAGLATFGTVSRTGEFAFETEEIPVTKLLNKRAE
LVAILGWLPKYWQSIVGTLAVFSGGSNASPKLAGLAVASVAKRLSNPQAREDMLNRLLDARDDNGEPMSPEELSAEAMTLIIAGADTVANTSCA
TTYYLARNQRVQAKLQAELDEALKAVDSEVALYDAVKYLPYLDAVVNEGLRLHATIGAGLPRVVPEGGITVLGHTFKEGTCVSVPIYYLHRDES
IWGANATEFYPERWLDATGERKKAMLDAFTPFSVGPRACIGRNLALMQLHIVIATLFRRFNVVLESDYSLPVQDSFVRKPRWCAIGIKLRKL

>75CYP53C (154264) Wco

MPLLASITAYALDYMCMPTFWLALIPAVLTTVYIVPYLLDPLGLRTFPGPIFAKFTSGWLPWVISQNRWSAIVDSLHRKDGIFVRLSPNHISIA
HPVALPAVYGHSSGALKAPFYDAFASFKTRNMFNTLSRTEHTRKRRIESQIFSQQSVRELEGTARVHHIDLVSQWDKLYSYVKRAESGAAREGF
NYWSFDTIGDLAFGDPFGMILAAKDTARSAKSVKASLETFSTSSNTEKLAFETEELPVTKIMGQRGELVAMLGWLPEYVRPIILMMPGFRSNLQ
AIPKVAGLAVAAVAKRMNNPDARADMLNRLLDAPGAITVANTSCATTYYLARDQRVQAKLQTELDDALKAVDSVVAPHGAIKHLPYLDAVVNEG
LRLHSPVGAGLPRVVPEGGMTVLGHTFKEGTWVSVPTYHLHRDENIWGENASEFYPERWIEASGDQKKAMLDAFVPFSIGPRACIGRNLALMQL
HIVLATLFHRFNVVLESDDPVRSQAYALYVAGHAAMGIEAVMLRQKEF

>76CYP53C (104840) Wco

MPLIGDITVQTLPFVGISTIWLALISITGIVTVICVVPYLLDSLGLRAFPGPTLAKFASGWLPWVISQNRWSVTVGRLHEKYGTFVRLAPNHVS
IAHPAALSAVYGHSSGALKAPFYDASGNFKARNMFNTRSRSEHARKRRSESHMFSPQSVRALEGTARVHHGNFVNQWGKLYSYVAQAKSGEAKE
GKLGACAWKVEDGRVWFNCMPWFNYWSFDTIGDLAFGAPFGMLLSAKDTVRVATSVKAGMAAFGTSSTTGKFTLETEEIPATKLLNKRAKLVTT
LGWFPKYWQPIIELLPPFRAGREATPKLAGLAVAAVAKRLSNPDAREDMLNRLLDARDENGEPLSPEELSAEAWLLIIAGADTVANTSCATTYY
LARDQRVQTKLQAELDEALKSIDSAVAPYDAIKHLPYLDAVVNEGLRLHATVGAGLPRVVPEGGLTVLGHTFKEGTWVSVPVYHLHRDESIWGE
NANEFYPERWIEASGDRKKAMLDAFAPFSVGPRACIGRNLALMQLHIVLSTLFHRFDFVLESHDPLPVQDSFVRKPKRCVIGIKSRKL

>77CYP53A (56813) Ade

MSSFWLCLAAAALLVLWRWRADPLRNVPGPWLARWTPFWLMYHARRGERYLAVHDAHKKYGPAVRIAPWHVSFASPDAPARVYAQGSAALDKSP
FYRAFYVQGAESLFSTQNRALHAAKRRLLSQPFSYQSIRGFEGFMRESLGRFVRRLDAVCAGECFGDAVRPGGAIDALLWFNYLAFDIISDLAF
GEPLGMVNKGSDLLPAERKDGTIFEEHAAALIDQRGRTAAVVGLMPSIEDITKKLPIPFITAGYKSTESLSRIAVRCVKHRIQSGVTRDDMLER
LIDGVREKQGGEVSEEEVVTEAMLLLTAGADTTANSLTAILYFILTRPDVYKKLMAELDSINAPTAELDTGTTIDGLPTHDQVKNLPYLNAVIE
EGLRLFATNAFGLPRVSSREGFELDGWTIPAGVEVSAPAYTIQRDPRIWGPDADDYRPERWIDETDSGLKKHMLTFGMGPRACIGKNLAYVQMQ
LALATSLLRYEFLLSPKAQLRSIEGFMHKPVELWVGIRRRHQPTA

>78CYP53C (70450) Ade

MEDLLRLAVYLALAGSIGTTLFVLVPYFKDEHGVRDIPGPLAAHLSNLWLAYWSSQGKRSEMVHEQHLKCGKLVRIAPNHISVNDPDALPIVYG
HGTGTLKSEFYDAFVSIQRGLFNTRSRTQHTRKRKIVSHVFSQKNVLGFEPNLHSALSRFVSQWDRMCAAGVKGGRGNEQDGWHGDGQRVWMDC
LPCAYNYLAFDIIGDLAFGSPFGMLDACADSANAAVGGVNALKDGKPMQTVSVPAIRILNERGEFSATMGVLAPWMRPLVLKLPWFARGLSAVR
ALAGLAIAAVGRRLAEPSDRNDLLAKLQDAKDDDGMPMGPEELTAEALTQLIAGSDTTSNSSCAIAYYVARYPRVQLKLQQELDAALPNDGVTT
YEQVKRLPYLTAVINEGLRLHSTSAMGLPRIVPEGGLTVAGRFFTEGSILSVPSYTIHRDPEVWGEDFDKFRPERWSEGDQTLIQKTFNPFSWG
PRACVGRNLAMMELLLIVSTTFRRYHLVLESDDEASQTREGFLRKPVLCRIGLKRRSPV

>79CYP53A (105834) Cpu

MKPDPLADVPGPLLARWTPLWLGYYARIGQRFTAVHKLHMEYGPIVRIAPNHISVADKDALDLVYAQGSNAFDKSTFYHAFVSDKASVFSTTDR
HDHAQKRRLVSNIFAAKSLQDCTPFIRDIVDSFVVQLDRLAAKNEELNLLYWFHFLAFDVLSDLAFGQRIGMVEKGSDAVTVQKRDGSVSTENA
IALVDEREHLGAVLGVHPSFKFWSKFLPDPFFIQGRKSSDGLVDFARRQVSRRIDNRLQRNDILDKLIRARVADDQEIVGENFADLVAETVTLL
IAGSDTTSNSETAIMHLLFTNPRVYNKLIGILEEAVDEELPTADHVRDIPYLDAVINEGLRYHATTAIGLHRAVHEKGAMFGGKYFPPGTEMSV
PAWTIQHDPEIWGDPEVFRPERWIENPDLKKYLMTFGKGPRACLGRHLAYMEMRLVLSTVLLRYDLQLKSQVMETTEGFMHKPNEMFVRLSRRE
RKGAAVQAQAQA

>80CYP53C (127772) Cpu

MDSVIATLKDLVPINLNFDFDALLDRLRSISPSQLAAGVPVCLLLYFLVPWLWDPYHQRSIPGPFLAKFSNAWLGWVSAHGHRSEIVHELHKKY
GPVVRIAPNHVSVADPEALQVVYAHGNGSLKSDFYDAFVSIHRGLFNTRDRQQHARKRKIVSGIFSQKNVLEFEPHVRLYVGQLMEQWDRLCAR
AEKGESGDEGEGGWQGRGGKLWLDCLPWYNYLAFDIIGDLAFGSPFGMIHSAKDSAPVAVSHADAMSAYGSSASNIKVVHIPAVQILNDRGEYS
AAMGVLPPAIRPFMQRFVPWYRKGGKAVRNLAGIAVAAVAKRLNEPSDRVDLLRRLQEAKDDEGNPMGREELTAEALTQLIAGSDTTSNSSCAI
TYYLALHPEIQTKLQRELDDALGTDDDPVSTFDAVKRLPYLDSVINEALRLHSTSSIGLPRIAPEGGLALRGLWFPPGAILSVPSYTIHRDAGV
WGADTEAFRPERWAEEERRDAVQRAFNPFSFGPRACVGRNLASMELLVIVSSILRRYTFVLEDAAKPFDTREGFLRKPVECRVGIRRRV

>81CYP53C (83844) Cpu

MDVQQLIQHLPQNVDLTSAAAASFAAFLAVHLGPYAWDRYHLKSIPGPFWAKFSDAWLAWVAANGHRSEEVHKLHEKLGPVVRIAPNHISIADP
DALQIIYAHGSNTLKKSNFYDAFVSIRRAIFNTREKADHARKRKIVANTFSQKNVIEFEPRVRIYVGQIIDQWDRLSKLAAADGSGDEGESGWY
GKDERLWLDVLPWMNYLAFDIIGDLAFGQPFGMILKAKDSAPVAVSQDAAMDSYGKECKVIEVPAVKILNDRGDYNATLGTMPPWVRPYVRKLP
WFSQGSEAAASVAGMAVAAVSRRLTTPTDRVDILSKLQQGKDENGEIMGPEELTAEALTHLVAGSDTTANSSCAIIYYLAAYPHVQEKLQKELD
EALGSEDEPVTTYEQVKRLTYLEVVILEVLRLHSTIGLGLPRMAPEGGLTVHGTYFPEGTILSVPTYTLHRDKRVWGDDPEIFRPERWFEENSA
KMHKAFNTFSFGPRACVGRNLANLELLIIVSSLLRRYDFVLKNPGDALGTCEGFLRKPTDCWVGLRRRAL

>82CYP53C (52716) Dsp

MSSLLDYFPTPFSSNALLLYGGLAVLLLVASHLVPWVADPFGYRKKHIPGPFLAQLSDVWLARVAAQGHRSEIIHGLHQKYGKVVRIAPNHISL
SEPGALQIVYAHGNGALKSDFYEAFVSIRKNIFSTRDRAEHTRKRKIVSHIFSQKSVLEFEPYIRQALGKLVKQWDSLLSDDRKLASHRLRPNE
NGTAWFDCLNWYNYLAFDIIGDLAFGEPFGMINSGADSASVAIHGDDPTHLASGEKKLEIVRVPAVKILNDRGEYSASMGCLPIWIRPYAKKIP
WYAKGNQAVKNLAGIAIAAVDKRLATPTDRVDLLARLQQGKDEQGNLMARSELTAEALAQLIAGSDTTSNSSCAITYWLAKYPDAQRKLQKELD
EALGDDEDVPTYEQLKRLRYLDAVVNEGLRIHSTSSLGLPRIVPEGGLEVSGIHFPAGSVLSVPSYTIHRDTAIWGPDPDIYRPERWFEQDAEG
IQATFNAFSFGPRACVGKNLASMELLLIVATIFHRYEFALLSQDQPLETREGFLRKPVSCYVGMKRRST

>83CYP53C (55106) Dsq

MSLVDRLLNSEPATWAVVGFTAVLLIHFVPYLADPHHIREYPGPLLAKLSDIWLGYVAAQGHRSERVHELHKQYGTFVRIAPNHLSISDPEALQ
VVYGHGTGTLKSDFYDAFVSIQRGLFNTRSRVQHARKRKIVSNIFAQKNVLDFEPHVRQHLANLFRQWDKLCEGGKNGLSGDEGEGGWQGRDGR
VWYDCLPWYNYLAFDIIGDLAFGAPFGMLDACKDSAPVAVSHKAAMAAYGSSDSSKEIQIEHFPAVQVLNDRGEYSAAMGVLPPHWRPLAKKIP
WYSKGNQAVQKLAGIAVAAVAQRFANPSDRADLLSKLQEGRDDNGDPMGREELTAEALTQLIAGSDTTSNSSCALTYWLAKNQAAQRKLQQELD
AALGSDDDPVASYEQVKRLPYLEAVINEALRIHATSGIGLPRLVPEGGLTVCGKFFPEGTVLSVPTYTIHRDKAVWGEDVDDFRPERWFEQDKN
LVQKTFNPFSFGPRSCVGRNLANLELLVIVASIFRRYHFVLENPNAQLETREGFLRKPVECKVGMKQRHT

>84CYP53C (154594) Fme

MQMISNISPWYFCLLPAAVLVLLVVPYFIDPYWIRRNEVRGPFLASLTSLWFGWNATRGHLSQVVHDLHKKFGTFVRLSPNHVSISDPDALQTI
YGHGKGLMKSDYYDAFKGLRPSIFSTRDRAFHAWKRKAISHIFSPKSVIDFEPYIHLHLTELFEQWDKLYDGGKRGLSGVEGEGWNGRQGRVWF
NIMPWFNYLTFDIIGDLAFGAPFGMIRKGKDAAPVAVDLKAAIAQYGQAGIDGQDLEKPAIQVKEVPAVQILNDRTFLASHQAAFPKPLRPLLA
LLPQYAEMAKHSDEFIGFAVAAVAKRLVFPTERVDILSKLQQSKDENGNPQSREDLTTDGITQLVAGSDTVANTSCGITYHIASNPCVQAKLQA
ELDDALGKDMEDPVVTYAQIKNLPYLEAVLNEGQRVYSTAALGLQRIVPEGGLTISGKWFPEGTIVSVPTYTIHRDPKVWGEDVDVFRPERWLE
GDHSAMSKTLNTFSIGPRACVGRNLALLELHIFIASIFYRYELVLEEPDKPLETHEGFIRKPLTCRVGMKRRDV

>85CYP53C (143663) Fme

MILIILRVFVDIGYRIWLFLPWIFLVVHLIPYLVDRYHIRRNGITGPSLARFSDAWLGWVVANGRQSEVVHEMHKKFGPVVRLAPNHVSISDPG
ALHVIYGHGSGLLKSGYYEPFTAVRPSIFSTRSREVHSKKRKIISHVFSQKSVLEFEPFVHLHLAELFEHWDKMCDGGKEGLSGTESEGGWKRR
GGQAWFDIMPWFNYLAFDIIGDLAFGSPFGMVRNAKDAAPIAVDRKSAMAQYGPVITDNRGLEKPVIDVREVHAISVLENRMRLSAQMGVLPAW
WRPIVRQLPRFAQGVQNSKDLVDLAVAAVAKRMAYPTQRDDILSKLQQSRDEYGRPLTQEDLTTDAITQLVAGSDTISISSCGIAYHLAANPDV
QSKLQKEIDDALGGFDDPMVTYAQIKHLQYLEAVINEGLRVHPTPGLGLPRVVPEGGLNVCGKWFPEGTILSVPTYTIHRNTGVWGEDANVFRP
ERWFEGDQAAMQKVFNAFSFGPRACIGRNLAMMELYIIIASIFHRYELILEEPNKPLEIHEAFMRKPVACHVGLKRRGA

>86CYP53C (149618) Fme

MSMSSGIVQELLNVTFWLWFFLLAIIFAVHLDAYFIDSHRIRRNGISGPFLARFSGAWLGWVVFQGRQSEVVHSLHKKFGTFVRLSPNHVSISD
PDALRLVYGRGNGALKSDYYDAFLAVRPSIFTTRSKEEHARKRTAIAHAFSQKSVLEFEPYIRLHVAELFNQWDRMCNRGKNGLSGTEGEGGWI
GQGGRVWFDIMPWFHYLAFDVMSDLAFGASFGMVRNAKDAAPIAVDQRAAMAQYQGTRVDSLDLEKPSIDVKEVPAVTMLNAHIKVSARMAAVP
PRWRPILQRLPCFARDMRASEDLVALAVAAVARRLVFPAERIDVLSKLQETEDEHGRVSNMEDLTTDAFTQLVAGSDTVSSTACGIAHCVAANS
RVRAKLQQELDVVFGGSYDPVATYAQIKRLPYLEAVIIEGLRVHSTSGLGLPRTVPNGGLIVCGKWFSEGTVLSVPTYTIHRDPIVWGEDADAF
RPDRWFERDQTILQKAFSPFSFGPRACIGRELAIMELCIFVSSIFHRYDLELEAPDKPLTIREDFIRKPVACRVGMKRRNI

>87CYP53C (94457) Fme

MVPNIAHNFLLGYSWILLLLPAIILVVHLVSYFVDSKHIRRNDIPGPTLAKVSGSWLGRVALEGRQSEVVHELHKKFGTFVRLSPNHVSISDPE
ALQVVYGHGNGMLKSEYYDAFAAPNLRRSVFDTRSREEHARKRKAISHIFSQKSVLEFEPYIHTHLTDFFKQWDKLCDGGRKGFSGIEGEGGWK
GHDGRVWFNAMPWYNYLSFDIISDLAFGTPFGMIRKARDAVPVAIDHKAAMAQYGQIDTEYRDVKKLVIDTREVPAIQVVNEQGEGVAQIAAFP
PLWVPFLRCLPRFAKGMRRVEDFIGLVVLAVANRLAFPTERVDILSKLQQGKGEDGVPLTKEELTSEALVQLIAGSDTTSNTTCAITYYVAANP
HVQTKLQKELDNALGHSENHVATYSQIKQLSYLDAVVNEGLRVHSTVGIGLPREVPEGGITVLGKSFPEGTVLSVPIYTIHRDPKVWGKDVDSF
RPERWIEGDKAAMQKVFSPFSVGPRACTGRNLALMSLHIFIASIFRRYDIVLEQPDKPLEVHDAFARKPNSCRIGLKRRDV

>88CYP53C (115179) Fme

MLPNILDALAQLNTSQLCASALAVAAVVYLQPYLVDSHFIRRNGITGPFFARFSDAWLGWVAAHGNRSVVVHKLHKKYGLFVRLAPNHVSISDP
EALHIVYGHGSGTLKSDYYDAFLAIRHTVLTTRDREDHSMKRKLVAPIFSQKSVLGFEPCVHSHVTELFEQWDKLCDGGKQGLTGNAGKGGWKG
RDGRVWFDALPWLYYMCFDIIGDLVLGAPFNMVHKGTDTVPVALEPSAVIAQYGQSSITGSHDTEKPICAVKEAPAMELMNGRSAVIASLGVLP
PWWRPIASLFPWYARGNRDVGDLAGFATLAISKRLARPTEPLGLLSALLELKDDEGKPLSKEQLSADGLLLLIAGSDMVANPTCAVLYQIIANP
PVQAKLQKELDDALGAPSPSDDSVSTYSQINHLPYLEAVINEALRVHPMVGLGLPRVVPASGLTVCGKHFPEGTVLSVPTYTIHRDKEVWGEDA
DTFRPERWFEGDKSTMQKVFNAYSYGLRACAGRVLANVELQIFISSIFRRYEFVLEEPENPVEVFEGFIMRPKSCRVGMKRRAI

>89CYP53C (24265) Fme

MILKILYSVFFTNYLWLCTGLTAAFFLVHVVPYLFDKHHIRRNDISGPLLARFSDAWLGWVAAQGPRSEVVHQIHKKYGKFVRLAPNHVSIAYP
EAIGEIYGHGNGTLKTDFYDAFLSIDRTIFTTRSREEHTRKRKVIAHGFSQKSISQLEPYIRLHVAELFEKWDKLYDGGRKGLTGVEGHNSWEG
HDGRVWFNAMPWLNYLAFDIIGDLAFGAPFGMLRNAKDAAPTAVDQKAAMSENGQVNIQDLEKPVVAVREVPAVKVLNGRSEYSASMGVLPPRW
RPIARLLPWYAEGSKDVEDLAGLAVAAVAKRLAIPTDRADILSKLQQGRHEDGSPMSREELTADALTVLIAGSDTTSNSTCALMYYITSNPRVQ
AKLQKELDEALVSFDDPVTSYDLVNHLPYLDAVIHEGLRVHSTLGVGLPRLVPEGGLTVCGKWFPEGTVLSAPTYTIHRDPKVWGEDADVFRPE
RWLERDHATLLKVFNTFSYGPRACIGRNVATMELFIFISSIFRRYDLVLEEQNKPLEVHEGFIRKPMACRVGMKRRNV

>90CYP53C (130308) Fme

MIASIFYSTFFTKYLWICAGLTAGFLVVHVVPYLLDKHHIRRNGITGPFLAMFSDAWLGWVAAQGRRAEVVHEKHKKYGKFVRLAPNHVSIADP
EAIGDIYGHGNGTLKTDFYDAFISIGVTVFTTRSREEHTRKRKVISHGFSQKSVSEFEPYIRLHVSELFEQWDELYDGGRKGLTGVEGEGGWKG
HDGRVWFNAMPWCNYLAFDIISDLAFALPFGMLRNAKDAALTAVDQKAAMSENGQVNTDMQDIEKPVVAVREVPAVKVLNGRSEYSASMGVLPP
WWRPIVRLLPWYADGSQDVEDLAGLAVAAVAKRLAIPTDRTDLLSKLQQGRHEDGRPLNREELTADALTQLIAGSDTTANSSCAVLYHIISSPR
VQAKLQKELDEALASLDDPVASYDLVKHLPYLDAVIHEGLRVHSTSGNGLPRLVPEGGLTVCGKWFPAGTVLSAPTYTIHRDPKVWGEDADVFR
PERWLERGQATLLKAFNTFSYGPRACIGRNVATMELFIIVSSIFRRYEFVLEEPHKPLEVHEGFIRKPMACRVGMKRRNV

>91CYP53C (162664) Fme

MTAVILSQSTMLPNILGITVHSVFLFFTLLGVLVLVVHLVPYFIDPYHIRSNDIKGPSLAGFSGTWLGWVAVSGHQSIVVHQLHKKFGTFLRLA
PNHVSISDVEALQAVYGQGSGTLKSNYYDAFVAFRPSIFETRSKAEHSRKRKAIAHVFSQKSVAEFEPFVRLHLAELFEQWDKMCDGGKKGISG
AEGEDGWNGHDGRVWFNVAPWFNYLGFDITGDLAFGFPFGMVCNASDTVQITVDQKASMDKYGQKDADQGSGKSAIETETKEIQAVKVLNNRTN
FDAHAGSFPPRFRPFMRRLPRFSECLQSSEDLAGFTVTAVARRLAFPSYRTDILSKLQQSKDEHGKPFSKEEVTSDAQTMLIAGSDTVSNTSSA
TAYYIAKNPGVQIRLQKELDEAFAASDDAVATYAQIKDLPYLAAVVNEGLRLHAPVGVGLQRVVPKGGLTVCGKWFPEGTILSVPTYTIHRDPG
VWGEDADAFRPERWLECDQSVMQKAFNPFSIGPRACIGRNLALMELQLFASSIFHRYEIVLEGVDKPEYACMPPNHVSISDPEALHVVYGHGSR
TLKSDYYEVFNSVRPSIFSTRSRTEHARKRKAISHAFSLKSVLEFEPYIHIHITFFLGQLDKLCDGKLNIPLDTEDDMCWEGQGGRVSFDIMPW
LNYLAFDIIGACVWISIRYVRNASNTAPVAVDQKACLARRSKIGADTDSFESEHIEVKEIERFSGRTPGEFTVAAVAKRLAFPTERANVLSNLF
LTKDEQGRLPSREDLTQDAITQLVAGSDTDSSKSVKACQAPDGAERCKGLPRVIGESGLTVCGKRFPEGTCVSVPSYTIHRDPLVWGEDVDVFR
PERWFERDPSIMQNAFNAFSFGPRFYDIILEEPDKSLEIHKTFIRRPIACHIGLKRQYD

>92CYP53C (128292) Fme

MGFTGVSTLTNALARVDLISLAVAVPLVLILGLIVPYFVDPHCIRNNGITGPLSARFSDAWLGWVAAQGHRSEVVHEMHKKYGTFVRLAPNHVS
ISDPAALQIVYAHGNGTLKSSFYDAFVSIRRGLFNTRSRPEHTRKRKIVSHIFSQKSVLEFEPHIRLHVGELFTQWDKLCDGGKRGLKGTEGDG
WEGHDGWVWFDCLPWFNYLAFDIIGDLAFGSPFGMILKGKDAAPVAKDQKAAIAGYGRESASEKSACDVTELPAVQVLNDRGEYSASMGVLPPW
WRPFVRRIPWYANGNRAVKNLAGLAVAAVAKRLANPTDRTDLLSKLQEGKDDEGRPMGREELTAEALTQLIAGSDTTSNSSCAITYHLAHNPHV
LKRLQQELDTALAGEDDPVATFQQVKSLPYLDAVINEVLRIHSTSGIGLPRLVPEGGLTVCGKTFPEGTVLSVPTYTIHRDKEVWGEDVEAMRP
ERWLEGDQAAIQKTFNPFSFGPRACVGRNLASMELLIIIASIFRRYEFVLEKPDEQFDTREGFLRKPLRCRVGMRRREL

>93CYP53NS (92916) Fme

MSDIIAPFLPVKRFVASLPFAACILCLALLKIVLFFIAYFKARGQFPGPPVSSLWSGNLSESMADDVHDKWRTWHRKYGPVFQTWNGLFSRVVY
VGDPRIISKIGNSNWPKFHAQYSGFKPLSGSALFAQMDQERWKQQRKGLAPAFQPITVNDQYPMLQRYLTEFIEVIDAAARSGSVIDLSTLHVL
LTLDFVGEVAFGAELNALRDGASCRILQIFHDILPELMKCGLFPLRSQIPIMESTRRMHRSIKELRGMAHVAVKNARSSEEKSAVQPGSKRIYE
ILAQLSVNLKLLHVLVTFLPLLFQLRNPEILAKVRTELDEVLPPDSEIPTVEQASRLRYLHLVIKETLRYNGPGFGTFRYTSKDVEINGVTLPA
NTTLALWNPQVHRDPKLWGPDSDEFRPERWLVSTGSTSSEFSRFIPPPGSYFPFSYGPRKCLGEGLAILEMSLTLATLFKRYDLKLQEGFVMEF
LPSFTLCSKNGLPVTARVRAQV

>94CYP53C (86809) Fpi

MSIVQQLIDFTRGNPLLVLAALPVVLVVAKVVHYLADSSDLRSYPGPFLAKFTDAWIFWTVSSNRWSHSVEDAHIKYGPIVRIAPNHISIADPK
ALATVYGHSTGFTKANWYNAFSEFAAKNIFNTRSRSEHARKRRMEAHMFAPQSIRAVEPISHAHVNELLRQWDGLISRVAKAQQGGPNGGHIGA
TTWNVQDGRVWIDCMPWLNFWSFDTIGDLAFGLPFGMLKSGRDTAKVAKSAEEGFKAIDAMSKGGDALVVEEEEIPYIEYLSARAERNACLAWL
PPVWARVVLTLPAFSGYALTGRKLAALSIMAVARRLANPNPREDMLQKLLEARDEEGKPLSPQEMSSEAFLLIIAGSDTIANTTCGTTYYLARD
KRVQAKLQAELDGALASVDSEVVPYDAVKDLPYLDAVIHEGQRLHSTVGAGLPREVPTGGATILGHHFREGITLSVPIYRLHRDESIWGPDAAE
FRPERWIEASPERKKLMMDAFAPFSVGPRACIGRSLAIMQLHIIVATLLRRYDFALQSDEPLRVRDSFARRPQECMVGITRRK

>95CYP53C (48859) Fpi

MSIVQQLTNFTRGNPLLVLAALPVVLVVAKVVHYLVDSSDLRSYPGPFLAKFTDAWIFWTVSRNRWSRSVEDAHIKYGPIVRIAPNHISIDDPK
ALAMVYGHSTGFMKSNWYDIFAAFSVSNIFDTRSRSEHARKRRMEAHMFAPQSIRAVEPISHSHVNELLRQWDGLISRVAKAQQGGPNGGHIGA
TTWNVQDGRVWIDCMPWLNFWSFDTIGDLAFGLPFGMLKSGRDTAKVAKSAEDALKAIDTVSKGGDVLAIEEEEIPYIEYQSARAETDACLAWL
PPIWVRILGKLPMFSVHALTGQKLAALSIMAVARRIADPNPREDMLQKLLEARDEEGKPLSPQEMSAEAFVLIIAGSDTIANTTCGTTYYLARD

KRVQAKLQAELDGALASVDSEVAPYDTVKDLPYLDAVIHEGQRLYSTIGAGLPREVPAGGATILGHHFKEGTTISVPIYRLHRDESTWGPDAAE

FRPERWIEASPERKKLMMDAFAPFSVGPRACIGRSLAIMQLHIIVATLFHRYDFTLQSDEPLRVRDSLARRPLECMVGITRRKGELGYGSFEL

>96CYP53C (1025718) Fpi

MSSVVDQLTGLPVAAWAGLVVAAVVLVHLVPLITDPYQVRSYPGPFLAKISDAWLGWVAAQGHRSEVVHELHQKHGKFVQIAPNHVSVSDPDAL

QVIYAHGNGTLKSTFYDAFVSIQRGLFNTRSRPEHARKRKIVSHIFSQKSVLEFEPYTRMHIKKLMNQWDRLYDLAMKGGSGEEGEGWQGRDGR

LWLDILPWYNYLAFDIIGDLAFGAPFGMLDACADAAPVAISHEKAMSSYGETDTPEITYFPAVQILNDRGEYSASLGVLPPHWRPIVKLLPWYR

KGNKAVQRLAGIAIAQVAKRLAMHTDRSDLLGKLQEGKDDEGNPMGREELTAEALTQLIAGSDTTSNSSCAITYHLAANPMVQQKLQRELDEAL

GNDDDPVSTFEQVKRLPYLEAVINEGLRLHSTSGIGLPRLVPEGGLTVCGRFFPEGTVLSVPSYTIHRDQDVWGSDADAFRPERWFEQDEKAIQ

KTFNPFSFGPRSCVGRNLASMELLIILSSILRRYHFVLEHPEQGLDTKEGFLRKPVECRVGIKRRTV

>97CYP53 (1116154) Fpi

MPSLPLGLKIPQLSALEICGIVTALFLVSYLRRRSDPIHAIPTVGPSWPLLSYLGAWRYFRDAKGMILEGCSKYEVFKIPLSDQWLVVVSGRDM

NDELRKYPDDTMSALEAQKWVVQTEYTLGNNNPDATAKKCISGPLTHKLGHVLPDVVDEMIHSFNDIMPDAEHDWQTVPALETMIKIIARVTNR

VFVGMPFCRNEMLLETAVEFAKDVMKTKFIVNLFPDVLKPIVGHRLPWTTARRRKMAEILGDTVRERRRQMLEYGTDYEGKPDDYLTWVVEEDL

KNRGKGESIDGVMEVIAASNFAAIHTSSMAMAHALYYLCAMPQYIKPLKQEAEEKIKEHGWTKTAMDAMWKTDSFFKESLRLNGVNHLSLFRKS

MKDVVLSNGTVIPAGTIVVATSTGTHLQEALYKDAAEFRPFRFSDVREKGGADAQKQQFHIPTAEYIAFGHGKHACSGRWFAAAEVKAILAYIL

LNYDFKLEKPGGRPENMNLGPSILPHPRAKVMFRKRKASRA

>98CYP53C (80617) Gtr

MLSGILNADVSNILILLVAVVAVAHIVPYFTDPHAIKSYPGPWLAKFSDAWLGKVSAQGHRSEVVHDLHKKYGTFVRLAPNHLSIADPEALQTV

YAHGNGSLKSDFYDAFVSIRRGLFNTRDRAEHARKRKIVSHIFSQKSVLEFEPHVRLYVRQFIEQWDRLCGLAAKGERGEEGNGWEGREGRLWL

DCLPWYNYLAFDIIGDLAFGSPFGMLKACKDSAPVAVSHADAMAAYGKDDSAVQVRSLPAVQILNDRGEYSASMGVLPVWFRPVVQRLHPWYRN

GNKAVKDLAGLAVAAVAKRLRNPTDRVDLLSKLQEGKDDEGKPMGREELTAEALTQLIAGSDTTSNSSCAITYYLALHPRVQEKLQAELDEALG

NDDDPVSTFEQVKRLKYLEAVINEALRVHSTSGIGLPRVVPEGGLTVLGRTFPEGTIMSVPTYTIHRYEEVWGPDVDEFRPERWFEIDQAQINK

AFNPFSYGPRACVGRNLASMELMIIVSSIFRRYHFVLEEPEKKFETREGFLRKPVECRVGIKRRN

>99Cyp53A (101387) Pst

MQNSTLFSSATHNLPEAMQDYTWLVFLPAVVVGYIVAVSRDPLQKVPGPLLARWSNLWQAFYTRFGIRYKAIHAVHKTYGPVVRISPNHVSIAD

MSLLPSIYGQGMAAFNKSPFYDAFLSEKPSIFSTRDKQEHAQKRRNYSGAFAPKTIRSYTTTVHRFLEELLVKLDKRAALPGEPDKAPIDMLIW

SNYLVFDIMSTLAFGTPLGMLEKESDVLQAGSPKGAIENTDVREHYLTVIGWAPALAYIARLIPDPFFQKGSKSSDELRDIARMCIKQRLASGS

DDAKSDILGHLIAAHMEYKNHLDVEELTSEALTLLIAGTDATSNAITAIIHALSVNPRPLAKLREELDEALSPGGLQGPTSDLNDLPYLNACIH

EAIRLHSPTGMGLPRIVPEGGLTYRDYYFPPGTDVSVPTWTMSRDRAAWGEDADVFRPERWIEDPSLTKYFMAFSSGTRGCLGKSLAILELKMI

VATLLQRYDITPQSLVLQTTEKLMHKPTHDWVRLRRRNVD

>100Cyp53C (68781) Pst

MVLSDPLTLAGLGLAAVVVAHIGAYLLDPHNIRDIPGPTLAKFSDAWLGWVAAKGHRSEVVHEMHAQYGPVVRIAPNHVSIAEPQALQIVYAHG

NGSLKSNFYDAFVSIQRGLFNTRNRADHARKRKIVSAIFSMKNVLEFEPHVREYVGLLIKQWDRLCAEAVKGGSGDEGEGGWRGESGRLWLDCL

PWYNYLAFDIIGDLAFGQSFGMLHACKDSAPVALSQDEAMKAYGSASGYKVVSIPAVQILNDRGEFSASIGVLPPAWRPFVKNLIPWYRNGSKA

VKNLAGLAVAAVAKRLDRDTLNGGSDRVDLLAKLQQGKDDEGKPMGREELTAEALTQLIAGSDTTSNSSCAITYHLAANPNVQAKLHAELDEAL

GTDDDPVAIFDQVKRLTYLQAVIDETLRIHSTSGIGLPRIVPAGSGGMHVAGHFFPEGTVLSVPTYTIHRDKEVWGEDVEVFRPERFLEGDQAV

IQKTFNPFSFGPRACVGRNLANMELLIIIASILRRYHFVLEHPEKPFDTREGFLRKPVECKVGIKRRSA

>101CYP53A15 (ACF15219.1) Clu

MFLTSLLLTPYTILLLPVLFYLLPYLRNWRIRDIPAPFPAAWTNLWLLYQCRRGRRFLAVHEAHQKLGKLVRIQPNHVSIADADAITQVYGHGN

GFLKSEYYDAFVSIRRGLFNTRDRAEHTRKRKTVAHTFSAKSVLQFEQYIHHNLQELQNQWDRRAESVKGGWYEMDALNWFNYLAFDVIGDLAF

GEPFGMLKKGRDEAEVARGGKITYAPAIEVLNRRGEVSGTVGIFPAIKPYAKYFPDPFFSQGMKAVENLAGIAIARVNARLEKPSDRVDLLARL

MEGRDENGNKLGREELTAEALTQLIAGSDTTSNTSCALLYHCLQHPEVVQKLQNELDAALPNPDAVPSYAQVKDLPYVDAVIKETMRIHSTSSL

GLPRVIPPGPGVTILGRHFPQGTVLSVPAYTIHHSTEIWGPDADTFRPERWEKVTEQQKAAFIPFSYGPRACVGRNVAEMELALIVATVFRRYE

FELRQGEMETREGFLRKPLALQVGMRKRSFA

CHAPTER 3

STRUCTURAL ANALYSIS OF CYTOCHROME P450 MONOOXYGENASE CYP53A FROM THE THERMOPHILIC FUNGUS *THIELAVIA TERRESTRIS*

3.1. Introduction

Fungi consist of large number of lower eukaryotes that are adapted to diverse ecological niches. Recent studies revealed that cytochrome P450 monooxygenases (CYPs/P450s) play a key role in their adaptation to diverse ecological niches (Syed *et al*., 2014). Genome sequencing analysis of fungal organisms revealed presence of a large number of P450s/P450 families in their genomes (Nelson, 2011). Among the different families CYP53 family is distributed in the phyla ascomycota and basidiomycota and is also one of the CYP families that display evolutionary conservation (Yoshida *et al*., 1997; Ichinose, 2012). CYP53 members are involved in *para*-hydroxylation of benzoic acid and its derived phenolic compounds (Faber *et al*., 2001). The resultant phenolic compounds are then further metabolized and degraded *via* the *β*-ketoadipate pathway (Harwood and Parales, 1996). Benzoic acid and its derived phenolic compounds naturally possess antifungal inhibitory properties (Amborabe *et al*., 2002; Podobnik *et al*., 2008). It is of great importance to study the three dimensional (3D) structure of a protein/enzyme, whereby information pertaining to highly conserved amino acid residues which interact with the substrate, often indicating structural and functional importance, is gained (Ashkenazy *et al*., 2010). Among different methods employed in understanding the protein structure, is 3D modelling also known as homology modelling. 3D modelling is gaining momentum due to its simplicity compared to

nuclear magnetic resonance (NMR) or X-ray crystallography. In this method, a 3D model of a protein is generated experimentally using a template usually a homologous protein whose structure has been elucidated *via* by crystallography. The 3D model of the protein generated helps us to understand the structure and function of the protein, its dynamics and its interaction with a ligand or substrate, especially in drug discovery and design, a core aspect of pharmaceutical research. There are no crystal structures available for CYP53 family members. Moreover, homology models for CYP53 members are few in number with the only available 3D model being that of CYP53A15 of *Cochliobolus lunatus* (Podobnik *et al*., 2008).

In Chapter 2 it was found that the ascomycete CYP53 members are highly conserved in terms of their primary structure and they can serve as common alternative drug target against pathogenic ascomycetes. In order to validate CYP53 as a common alternative drug target one should locate the positions of the conserved amino acid residues in the structure, as P450s have a typical three-dimensional fold where substrate recognition sites (SRS) play a key role in identification and catalysis of substrates (Gotoh, 1992; Podust *et al*., 2001). Furthermore, no thermostable P450 or its structure from the domain eukaryote has been reported to date. Biotechnologically, identification and characterization of thermostable eukaryoticP450s will have great potential. Especially, if the 3D model of CYP53 from a thermostable eukaryote could be constructed, whereby this will not only serve to validate the CYP53's ability to serve as a common alternative drug target, but also provide useful information on factors contributing to thermostablity for eukaryotic P450s. Furthermore, analysis of the CYP53 model from a thermostable eukaryote will help in engineering this P450 for the production of human valuables at an industrial scale. Considering the potential biotechnological importance as mentioned above, in this chapter , the main focus was 3D modelling of CYP53A from the thermophilic fungus *Thielavia terrestris* (an ascomycete) (Syed *et al*., 2014), followed by

identification of SRS regions and the active site cavity to map the conserved amino acid residues identified in Chapter 2.

3.2. Materials and methods

3.2.1. Homology modelling

3D modelling of a CYP53 member P450 namely; CYP53A (protein ID: 2107910) from *T. terrestris* (*Tter*) was carried out using the methodology as previously described (Syed *et al*., 2013), with minor modifications. The Basic Local Alignment Search Tool (BLAST) was used for selecting the closest homologues (template) available in the Protein Data Bank (PDB) (www.pdb.org). The best and closest homologue hit observed from the PDB, was that of the recently crystallized full-length P450 protein CYP51 from the species *Saccharomyces cerevisiae* (Monk *et al*., 2014). Therefore, the protein structure of the P450 from *S. cerevisiae*, CYP51 was used as a template (abbreviated as *Scer*). The coordinates of the crystal structures of CYP51 (PDB ID: 4LXJ) were used as templates to construct the models of CYP53A (Monk *et al*., 2014).

The 3D model of CYP53A of *Tter* was generated using the homology modelling program Modeller 9v11 (Sali and Blundell, 1993). The modelling was performed with default parameters using the "allHmodel" protocol to include hydrogen atoms and the "HETATM" protocol to include prosthetic group HEM (heme). The 3D model accuracy was then validated using DFire (Zhou and Zhou, 2002), QMEAN (Benkert *et al*., 2011), and Verify3D (Lüthy *et al*., 1992). Heme-binding residues were identified using 3DLigandSite (Wass *et al*., 2010). Structure alignment between the template and CYP53A of *T. terrestris* was performed using PROMALS3D (Pei *et al*., 2008). PROMALS3D aligns multiple protein sequences and/or structures, gathered from advanced database searches, secondary structure prediction,

3D structures or user-defined constraints and will give a conservation index (Pei and Grishin, 2001). The conservation index follows numbers from 4 and above, where 9 is the most conserved amino acid across the sequences obtained (Pei and Grishin, 2001). P450 characteristic secondary structure annotations and SRS in the modelled P450 were identified according to their alignment with the template P450 and standard SRS localization regions, as described in the literature (Gotoh, 1992; Podust *et al*., 2001). Protein models were visualized using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/). The models of CYP53A from *Tter* and the template of CYP51 from *S. cerevisiae* (4LXJ) were then presented as a figure.

3.2.2. Active site residue mapping

CASTp programme was used to identify the amino acid residues lining the active site cavity of CYP53A *Tter* (Dundas *et al*., 2006). CYP53A *Tter* in this study was used as a representative of all ascomycetes. The protein active site cavity in the 3D-model of CYP53A of *Tter*, generated in this study, was also predicted using CASTp programme (Dundas *et al*., 2006). The cavity showing the highest volume and covering the most SRS regions was selected, whereby the program automatically listed the residues lining the active site. The active site cavity structure and the residues lining the cavity were then presented as a figure using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/).

3.3. Results and discussion

3.3.1. Homology modelling

To date P450s belonging to CYP53 members have not been crystallized. A study from Podobnik *et al*. 2008, resulted in the modelling of CYP53A15 from *C. lunatus* and further identification of amino acid residues involved in substrate binding (Podobnik *et al*., 2008). Since the CYP53A15 model was constructed, many other P450 crystal structures have been resolved. Therefore, the best template could be selected for, given the appropriate opportunity. In the present study a CYP53 member was selected, CYP53A from *T. terrestris* that was recently identified and characterized as a representative of ascomycete CYP53 P450s (Syed *et al*., 2014).

To identify the role of amino acids conserved as observed for ascomycete CYP53 members (Chapter 2), CYP53 homology modelling studies was performed to map these conserved/variant residue locations in the protein structure. In the present study, CYP53A from *T. terrestris* was selected as a representative of ascomycetes CYP53 P450s. The 3D-model of the CYP53 P450 (Figure 3.1) was constructed and validated as described in materials and methods. As shown in Table 3.1 all the parameters employed in assessing the quality of the model were favourable suggesting that the model of the CYP53 P450 was of good quality.

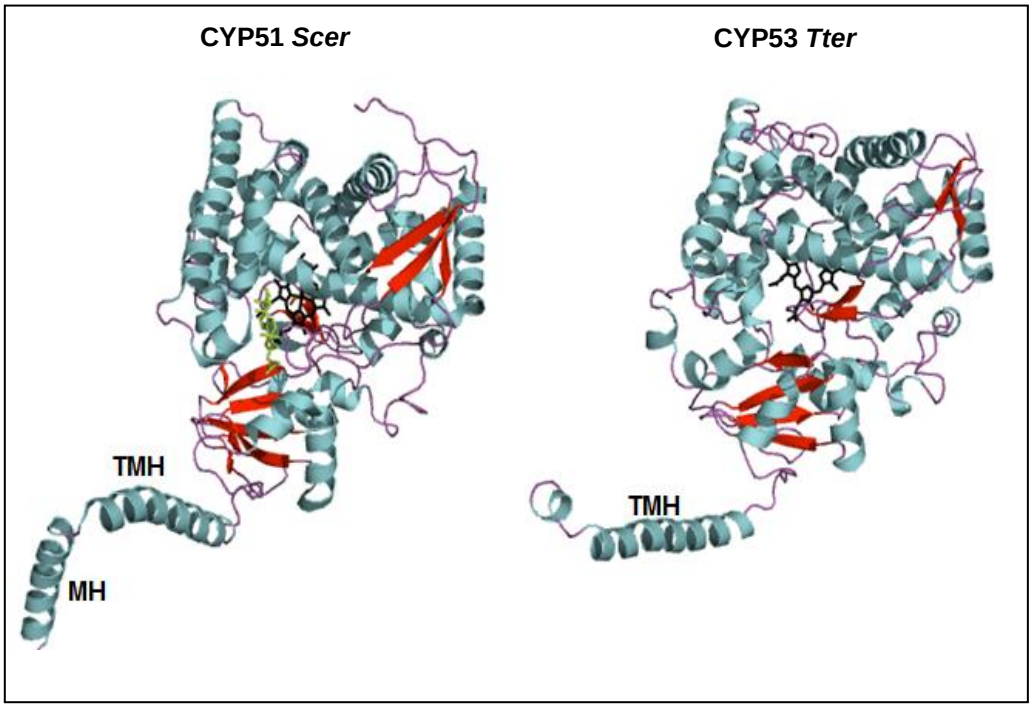

Figure 3.1. Structural analysis of CYP53A *Tter*. The 3D-model for CYP53A from *T. terrestris* (CYP53 *Tter*) was constructed based on the template CYP51 from *S. cerevisiae* (CYP51 *Scer*) (PDB ID: 4LXJ) from PDB Data Bank (www.pdb.org). The heme prosthetic group is shown in black colour and the bound substrate for CYP51 *Scer* in green colour. Alpha-helices and beta-strands are shown with blue and red. The membrane helix (MH) and the trans-membrane helix (TMH) are indicated in the models.

Table 3.1. Validation of CYP53A *Tter* 3D-model. The CYP53A *Tter* model was based on the template CYP51 (PDB ID: 4LXJ) from *S. cerevisiae* and was generated using Modeller.

CYP name[a]	Sequence Identity (%)[b]	Length[c]	Coverage[c] (%)	dDFIRE[d]	DFIRE2d[d]	QMEAN6 score[e]	Average Verify3D score ± SD[f]
CYP53A *Tter*	16	510	100%	-1101.04	-849.785	0.57	0.23 ± 0.22

Table notes:

[a] CYP53A *Tter* was based on the template CYP51 (lanosterol 14α-demethylase) (PDB ID: 4LXJ) from *Saccharomyces cerevisiae* and was generated using Modeller (Monk *et al*., 2014).

[b] Sequence identity between CYP53A *Tter* and the template CYP51 (PDB ID: 4LXJ).

[c] Number of P450s amino acids modelled and their percentage compared to the full-length P450s.

[d] dDFire and DFIRE2 pseudo-energy (lower values signify a better model) (Zhou and Zhou, 2002).

[e] QMEAN6 composite score ranging from 0 to 1 (higher values signify a better model) (Benkert *et al*., 2011).

[f] Verify3D scores ranges from -1 (bad score) to +1 (good score) (Lüthy *et al*., 1992). This program analyses the compatibility of an atomic model (3D) with its own amino acid sequence (1D).

3.3.2. Secondary structure analysis of CYP53A *Tter*

As shown in Figure 3.1, CYP53A *Tter* showed all P450 motifs in the same way as CYP51 of *S. cerevisiae* (Monk *et al*., 2014). An interesting observation was that although CYP51 possessed both a membrane helix (MH) and transmembrane helix (TMH), only the transmembrane helix was seen in CYP53A *Tter* (Figure 3.1). This suggests that CYP53A *Tter* is a biotopic membrane protein with one transmembrane helix. Structural alignment between CYP53A *Tter* and CYP51 (4LXJ) *Scer* was performed using PROMALS3D. A

detailed secondary structure analysis including heme-binding residues, substrate binding residues and substrate recognition sites (SRS) namely, SRS1 to SRS6 is shown in Figure 3.2.

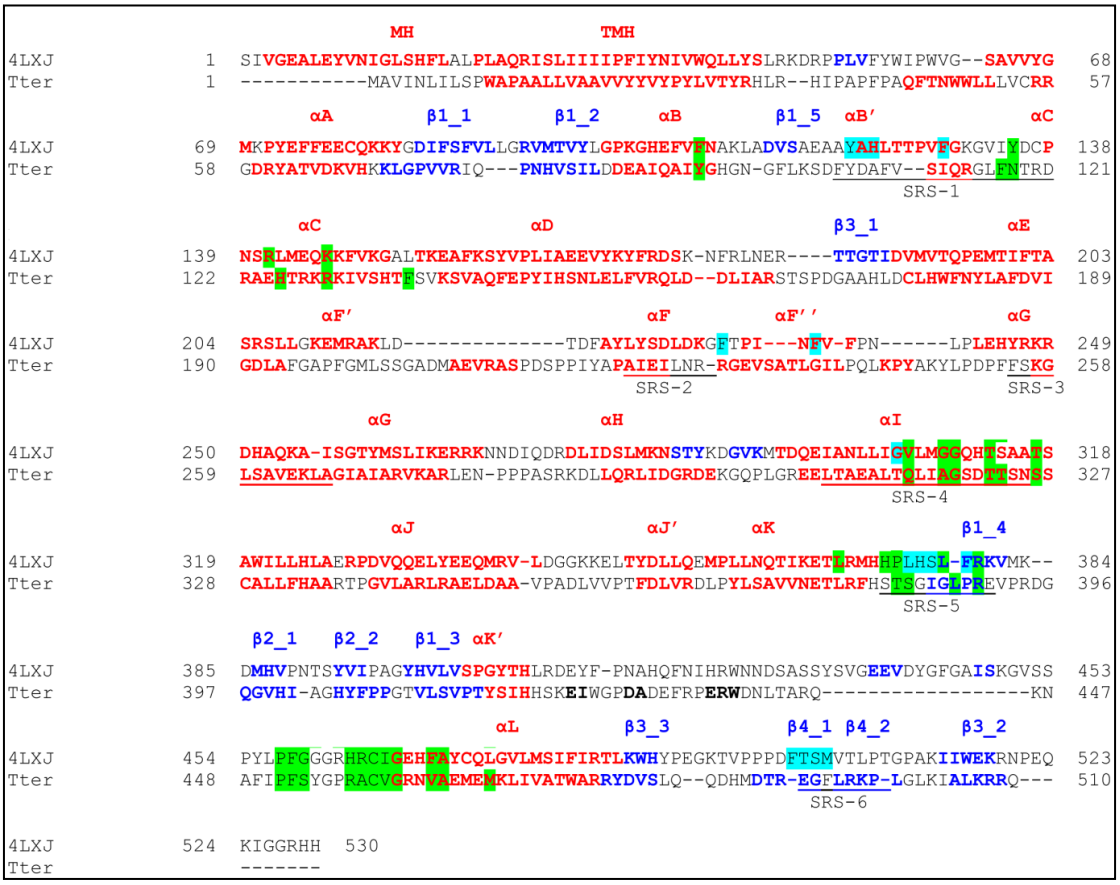

Figure 3.2. Secondary structure analysis of CYP53A *Tter* and CYP51 (4LXJ). P450 characteristic notations for alpha helices (shown in red font) and beta-strands (shown in blue font) with SRS regions mapped as per the template (4LXJ) (Gotoh, 1992; Podust *et al.*, 2001; Monk *et al.*, 2014). Highlighted amino acid residues in green and turquoise appear in contact with the heme and substrate.

3.3.3. Active site cavity and its amino acid residues identification

Active site cavity and the identification of amino acids lining the active site were performed using the model generated (CYP53A *Tter*) in this study. The aim of active site cavity mapping of conserved amino acid residues is to investigate how many of these conserved residues fall in active site cavity. The active site cavity and its amino acids were shown in Figure 3.3. Furthermore, the number of conserved amino acids present in the active site cavity is also estimated and their percent is shown in Table 3.2. As shown in Figure 3.2, CYP53A *Tter* a member of ascomycetes possesses conserved amino acid residues throughout the protein structure, at P450 signature motifs such as EXXR and CXG. CASTp was used to understand how many of these conserved amino acids are part of the active site cavity of CYP53A *Tter* (Figure 3.3 and Table 3.2) (Dundas *et al*., 2006).

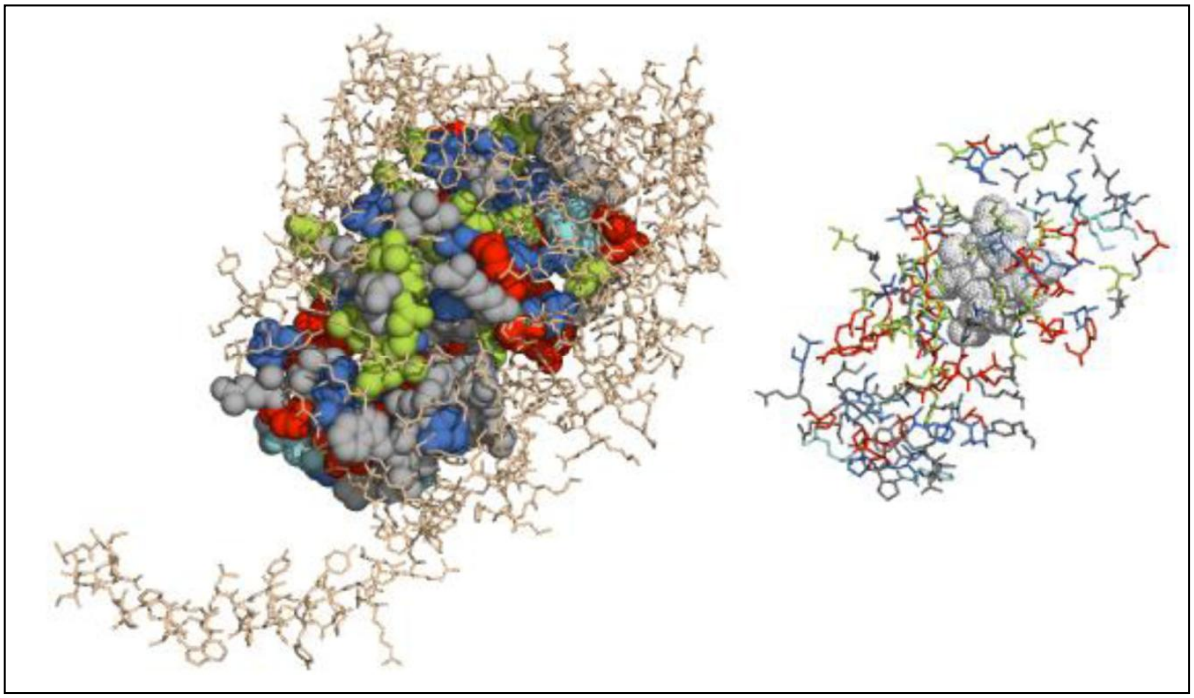

Figure 3.3. CYP53A *Tter* active site cavity mapping and analysis of the nature of the active site amino acids. Left-hand side: Mapped active site cavity (space filled) with the

protein backbone presented in line style; Right-hand side: A figure of the conserved residues lining the cavity mapping. Different colours correspond to their conservation index, 9 (conserved residues) – red, conservation index 7 – blue, conservation index 6 – cyan, conservation index 5 – green and residues with no conservation index – grey. The heme group is shown in grey dots.

Table 3.2. Analysis of amino acid conservation in CYP53A *Tter* active site cavity. Active site cavity residues were identified using CASTp (Figure 3.3) (Dundas *et al*., 2006). Conservation of amino acid residues in ascomycete species CYP53 P450s was identified using PROMALS3D (Pei *et al*., 2008). Conservation index 9 means amino acid is conserved. CYP53A *Tter* P450 amino acids and their numbering are presented in the table.

Conservation index	Amino acids	Number of Amino acids	Percentage (%) contribution in active site cavity
9	R76, P79, G99, L101, K102, Y106, F117, R120, R122, H125, R129, F136, L192, P241, E311, T314, A318, T322, L370, S382, G387, L388, P389, R390,G409, S413, F449, P451, F452, C459, G461, R462, A465, E466, K498	35	28
7	L51, H81, N98, L116, N118, V132,V188, I189, L227, A235, T236, L243, L288, L313, G319, D321, N325, S326, V364, L378, P393, G398, V399, P408, V411, L412, S453, R457, A458, M467, E468, G494	32	26
6	Q78, R365, V392, F406, V464, M469	6	5
5	F100, D104, F105, S111, T119, F144, I148, A193, A221, A223, I224, I226, N228, E232, A310, Q315, I317, S320, T323, V374, T383, G385, V460, L496	24	19

None	N48, W49, L52, R57, Y61, V74, S103, I112, L152, A185, I219, E225, I239, L240, Q242, S371, N375, S384, I386, E391, R394, Q397, P407, T410,K470, L471, R492, F495	28	22
Total		**125**	**100**

As shown in Figure 3.3 and Table 3.2, there are 125 amino acids lining the active site cavity, with 35 (28%) that are conserved (conservation index 9) and 62 amino acid residues (50%) that are moderately conserved (conservation index 5-7) across the CYP53 members of ascomycetes. Overall, the high conservation of amino acids (78%) in the active site cavity and in the rest of the protein structure (Figure 3.3) strongly suggests that the active site cavity and overall structure of CYP53 members of ascomycete species are highly conserved. Considering the structural conservation, any inhibitor developed against one of the CYP53 members could possibly act as common inhibitor against CYP53 members of ascomycete species and hence could act as a common anti-fungal (towards pathogenic ascomycetes) agent. Future studies include use of the 3D model of CYP53 *Tter* to identify factors responsible for thermostability of this eukaryotic P450.

3.4. Conclusion

In this study, structural analysis of CYP53A from *Thielavia terrestris* (*Tter*) was performed. The 3D model of CYP53A *Tter* generated in this study was found to be of high-quality after it was validated using appropriate parameters. Secondary structure analysis revealed that CYP53A *Tter* is a biotopic membrane protein, whereby one transmembrane helix transverses the phospholipid bilayer joining two domains within the membrane (Zviling *et al*., 2007). It was also found that 125 amino acid residues lining the active site cavity of CYP53A *Tter*

protein, whereby 78% of these amino acids are highly conserved. Due to the highly conserved nature of the amino acids lining the active site cavity of CYP53A *Tter*, results from active site residue mapping suggest that any inhibitor developed for this family, could act against a wide range of animal and plant pathogenic ascomycetes. Therefore, based on the 3D model generated, CYP53 family members can act as a potential anti-fungal drug target. CYP53A *Tter* model developed in this study will be used to identify factors responsible for thermostability of eukaryotic P450.

Currently, immune-compromised individuals with HIV (Human Immunodeficiency Virus) or those who are on treatment with immuno-suppressant drugs such as transplant recipients and cancer patients, are at a high risk of invasive fungal infections (Berne *et al*., 2012). Unfortunately, most azole inhibitors developed have been derived from fungal CYP51, which not only target fungal CYP51 enzymes but also those from mammals that catalyze xenobiotic compounds leading to toxicity and adverse side effects in patients (Berne *et al*., 2012). Therefore, CYP53 could be a promising novel therapeutic drug target due to its potential high substrate specificity and also because no homologs present in mammals (Berne *et al*., 2012).

REFERENCES

Amborabe, B.E., Fleurat-Lessard, P., Chollet, J-F., Roblin, G. (2002). Anti-fungal effects of salicylic acid and other benzoic acid derivatives towards *Eutypalata*: structure-activity relationship. *Plant Physiol Biochem*, 40 (12): 1051-1060.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*, 38:W529–W533.

Benkert, P; Biasini, M; Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27 (3): 343–350.

Berne, S., Podobnik, B., Zupanec, N., Novak, M., Kraševec, N., Turk, S., Korošec, B., Lah, L., Šuligoj, E., Stojan, J., Gobec, S., Komel, R. (2012).Virtual screening yields inhibitors of novel antifungal drug target, benzoate 4-monooxygenase.*J ChemInf Model*,52 (11): 3053–3063.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*, 34 (suppl2): W116-W118.

Faber, B.W., van Gorcom, R.F.M., Duine, J.A. (2001). Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. *Arch Biochem Biophys*, 394 (2): 245–254.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem*, 267 (1): 83-90.

Harwood, C.S., Parales, R.E. (1996).The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol*, 50 (1): 553-590.

Ichinose, H. (2012). Molecular and functional diversity of fungal cytochrome P450s. *Biological and Pharmaceutical Bulletin*, 35 (6): 833-837.

Lüthy, R., Bowie, J.U., Eisenberg, D.R. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356 (6364): 83– 85.

Monk, B.C., Tomasiak, T.M., Keniya, M.V., Huschmann, F.U., Tyndall, J.D., O'Connell, J.D., Cannon, R.D., McDonald, J.G., Rodriguez, A., Finer-Moore, J.S., Stroud, R.M. (2014) Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proc Natl Acad Sci USA*, 111 (10): 3865-3870.

Nelson, D. R. (2011). Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(1): 14-18.

Pei, J., Grishin, N.V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, 17 (8): 700-712.

Pei, J., Kim, B.H., Grishin, N.V. (2008). PROMALS3D: A tool for multiple sequence and structure alignment. *Nucleic Acids Res*, 36 (7): 2295-2300.

Podobnik, B., Stojan, J., Lah, L., Krasevec, N., Seliskar, M., Rizner, T.L., Rozman, D., Komel, R. (2008). CYP53A15 of *Cochliobolus lunatus*, a target for natural antifungal compounds. *J Med Chem*, 51 (12): 3480-3486.

Podust, L.M., Stojan, J., Poulos, T.L., Waterman, M.R. (2001). Substrate recognition sites in 14α-sterol demethylase from comparative analysis of amino acid sequences and X-ray structure of *Mycobacterium tuberculosis* CYP51. *J Inorganic Biochemistry*, 87 (4): 227-235.

Sali, A., Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234 (3): 779-815.

Syed, K., Porollo, A., Lam, Y.W., Grimmett, P.E., Yadav, J.S. (2013) CYP63A2, a catalytically versatile fungal P450 monooxygenase capable of oxidizing higher-molecular-weight polycyclic aromatic hydrocarbons, alkylphenols, and alkanes. *Appl Environ Microbiol*, 79 (8): 2692–702.

Syed, K., Shale, K., Nazir, K. N. H., Krasevec, N., Mashele, S. S., 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): 321-333.

Wass, M.N., Kelly, L.A., Sternberg, M.J. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res*, 38:W469-W473.

Yoshida, Y., Noshiro, M., Aoyama, Y., Kawamoto, T., Horiuchi, T., 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. *J. Biochem*, 122(6): 1122-1128.

Zhou, H., Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci*, 11 (11):2714 –2726.

Zviling, M., Kochva, U., Arkin, I. T. (2007). How important are transmembrane helices of bitopic membrane proteins?. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1768(3): 387-392.

CHAPTER 4

STRUCTURAL ANALYSIS OF CYTOCHROME P450 REDUCTASE (CPR) FROM THE THERMOPHILIC FUNGUS *THIELAVIA TERRESTRIS*

4.1. Introduction

Thielavia terrestris, is a sedentary and thermophilic organism. Like many other fungal species, *T. terrestris* is also dependent on innate metabolic processes to thrive in the environment (Lah *et al*., 2008). Recent studies on recombinant enzymes involving biomass-hydyolyzing activity suggests that *T. terrestris* is one of the most efficient thermophilic fungal organisms to be used for biomass decomposition from moderate to high temperatures (Berka *et al*., 2011). In such complex biochemical activities, cytochrome P450 proteins have a critical function to perform, mainly in primary and secondary metabolic processes (Bernhardt, 2006). Most microsomal P450 systems in eukaryotic organisms consist of a P450 and one or more redox partners (Yadav and Loper, 2000; Denisov *et al*., 2005). NADPH-dependent cytochrome P450 oxidoreductase or cytochrome P450 reductase (CPR) (EC 1.6.2.4; CPR), is a diflavin reductase protein that serves as an electron donor to numerous P450s (Yadav and Loper, 2000). The number of CPRs present in ascomycetes ranges from one to four (Lah *et al*., 2008).

In Chapter 1 of this thesis, different types of CPR classes were explored, indicating that CPRs are involved in electron transfer during diverse catalytic reactions (Denisov *et al*., 2005). Electrons are transferred in a specific trajectory from NADPH (nicotinamide adenine dinucleotide phosphate) to FAD (flavin adenine dinucleotide) to FMN (flavin mononucleotide) and finally to the heme group in the P450 enzyme (Yadav and Loper, 2000).

Co-regulation of both the P450 and CPR has been suggested to play an essential role in achieving optimized P450 activity (Sutter *et al*., 1990; van den Brink *et al*., 1996; Ohkuma *et al*., 1995). According to Lah *et al*., (2008) most studies conducted on CPR: P450 systems, give more focus to P450s and not on the redox partner due to P450s, having more diversity compared to their reductase counterparts that are perceived as extremely conserved and fully characterized in terms of functionality (Lah *et al*., 2008). However, the question lies in the fact, is the structure of CPR similar in all organisms? If not, how do these structures differ in terms of the environment or conditions the organism is exposed to? To understand the function of a protein, it is important to study the structure of the protein, hence only then can the true function of the protein be known. Currently, there are no crystal structures and 3D homology models that exist that give a better indication of the structure of CPR in thermophilic fungi (Table 4.1.).

Table 4.1. Information on crystal structures of cytochrome P450 reductases from the Protein Data Bank (www.pdb.org). Few crystal structures exist that provide us with insight into the mechanisms of CPR activity. From the table it can be observed that CPR from eukaryotic organisms have been characterized.

CPR	Species	Ligand(s)	Literature
NADPH Cytochrome P450 (Oxidoreductase)	*Rattus norvegicus*	FMN, FAD and NADPH	Wang *et al*., 1997; Hamdane *et al*., 2009; Hubbard *et al*., 2014; Xia *et al*., 2011
Polyketone reductase (CPR-C2)	*Candida parapsilosis*	NADPH	Qin *et al*., 2014

Polyketone reductase (CPR-C1)	*Candida parapsilosis*	NADPH	Qin *et al*., 2013
Oxidoreductase	*Homo sapiens*	FMN	Zhao *et al*., 1999
NADPH Cytochrome P450 (reductase)	*Saccharomyces cerevisiae*	FMN, FAD and NADPH	Lamb *et al*., 2006
NADPH Cytochrome P450	*Homo sapiens*	FAD, NADPH	Xia *et al*., 2011(a), (b)

Thermophilic fungi possess an array of thermostable enzymes that can be used in future for the production of human valuable goods at elevated temperatures in the form of cell factories, and as such serve as potential and indispensable biotechnological reservoirs (Berka *et al*., 2011). In this study, the first report on structural analysis of CPR from the thermophilic fungus and ascomycete *T. terrestris* is described, and is therefore a critical step in the molecular characterization of thermophilic P450 systems in *T. terrestris*. Moreover, in expanding the knowledge pertaining to structural analysis on thermophilic fungi, this will also enable us to understand the function of this protein for future studies, especially P450 monooxygenation mechanisms and metabolic manipulations in *T. terrestris* and other thermophilic fungi.

4.2. Materials and Methods

4.2.1. Selection of template

In order to build *T. terrestris* CPR (protein ID: 2115296) 3D model, a template (protein whose crystal structure has been elucidated) needed to be identified. Basic Local Alignment Search Tool (BLAST) search was performed to select the crystal structure of the closest

homolog available in the Protein Data Bank (PDB) (www.pdb.org). The protein that showed highest percent identity was selected as a template for construction of CPR *Tter* model.

4.2.2. Homology modelling and 3D model validation

The 3D model of CPR *Tter* was generated using the homology modelling program Modeller 9v11 (Sali and Blundell, 1993). The Ramachandran's map using PROCHECK (Programs to Check the Stereo Chemical Quality of Protein Structures) (Laskowski *et al*., 1993), and ERRAT graphs (Colovos and Yeates, 1993) were used to analyse the final structure of CPR *Tter*. ERRAT assesses the distribution of different types of atoms with respect to one another in the protein model. The WHAT IF program was used to analyse and also identify residue packing and atomic contact for the detection of bad packing of side chain atoms or unusual residue contacts (Vriend, 1990). The Z-score of Ramachandran's plot was predicted using the WHAT_CHECK software (Hooft *et al*., 1996).

The quality and accuracy of 3D model was then validated and carefully examined using available bioinformatics software: DFire (Zhou and Zhou, 2002), QMEAN (Benkert *et al*., 2011), and Verify3D (Lüthy *et al*., 1992). The protein models generated was visualized using YASARA program (http://www.yasara.org/) (Krieger, 2004).

4.2.3. Active site identification and ligand docking

The 3D model was used for the identification of active sites and for the docking of ligand residues (FMN, FAD and NADPH). The active site was predicted using an alpha shape algorithm to determine potential active sites in 3D protein structures in MOE site finder (Edelsbrunner *et al.*, 1995). Docking studies against the CPR was performed using dock module implemented in MOE2012 (Eldridge *et al.*, 1997).

4.3. Results and Discussion

To date CPR belonging to thermophilic eukaryotes has not been crystallized or modelled. Structural analysis of CPR from thermophilic eukaryotes is necessary in order to understand the molecular basis of thermostablity of this protein. Furthermore, if eukaryotic P450s could be utilized at an industrial scale they need a thermostable electronic partner which is CPR. As discussed in Chapter 1, a study revealed the presence of a large number of thermostable P450s in *T. terrestris* (Syed *et al.*, 2014b). Considering the thermophilic nature of this fungi and the presence of a large number of thermostable P450s, it is reasonable to predict that CPR from this fungus could be thermostable. Characterization of CPR from *T. terrestris* (CPR *Tter*) will pave the way for further utilization of this protein towards the generation of human valuable products. In this study, bioinformatics programs were used to understand its structural properties. This study marks the beginning of our understanding of CPR from thermophilic organisms.

4.3.1. CPR *Tter* modelling and 3D model validation

The 3D model of the CPR *Tter* (Figure 4.1) was constructed based on the template cytochrome P450 reductase (2BPO) from *S. cerevisiae* (*Scer*) (Figure 4.1). The *S. cerevisiae* CPR is the top hit during BLAST analysis at PDB. This indicates that this yeast CPR can serve as the best template for the modelling of CPR *Tter* due to its high percent identity. The 3D model of CPR *Tter* was constructed using Modeller programme as described in materials and methods. Validation of the CPR *Tter* model was performed as described in materials and methods. As shown in Table 4.2 all the parameters employed in assessing the quality of the model were favourable suggesting that the model of CPR *Tter* was of good quality.

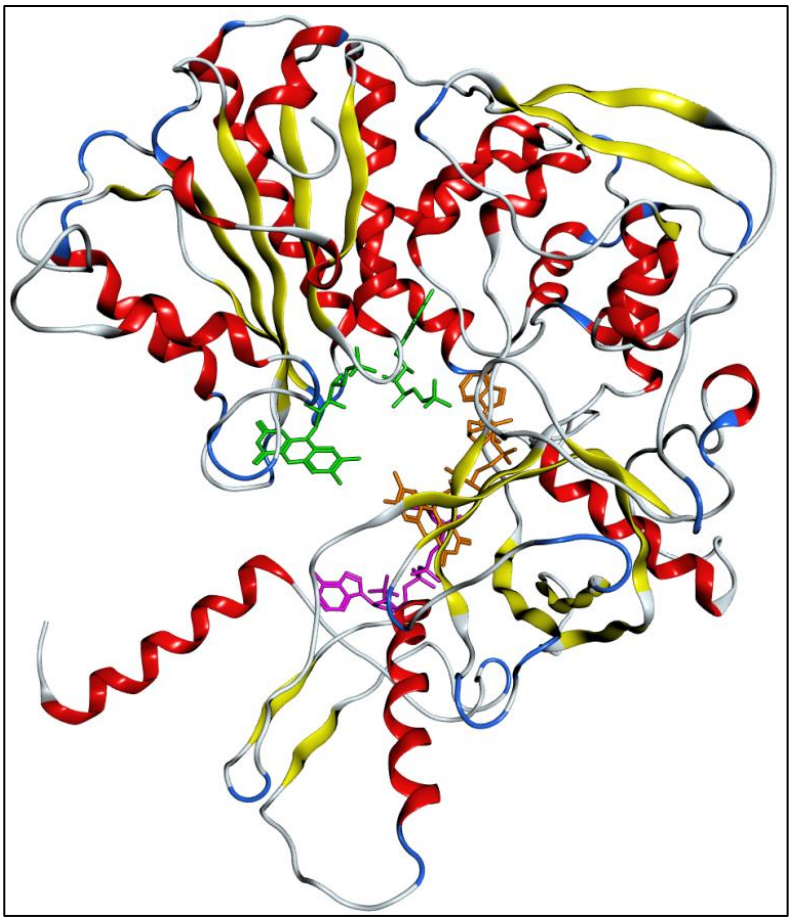

Figure 4.1. Three dimensional structure of cytochrome P450 reductase from *T. terrestris*. Three dimensional structure of the reductase was generated with Modeller as described in materials and methods. Alpha helices are represented in red and beta sheets are in yellow colour. FAD is represented in magenta, FMN is represented in green and NAP is represented in orange colour.

Table 4.2. Validation of CPR *Tter* 3D-model.

Cytochrome P450 Reductase (CPR) name[a]	Sequence Identity (%)[b]	Length[c]	Coverage[c] (%)	dDFIRE[d]	DFIRE2d[d]	QMEAN6 score[e]	Average Verify3D score ±SD[f]

CPR *Tter*	43	551	91%	-1144.90	-904.063	0.609	0.36 ± 0.24

Table notes:

[a]CPR *Tter* was based on the template CPR (PDB ID: 2BPO) from *Saccharomyces cerevisiae* and was generated using Modeller (Yermalitskaya *et al*., 2006).

[b] Sequence identity between CPR *Tter* and the template CPR *Scer* (PDB ID: 2BPO).

[c] Number of P450s amino acids modelled and their percentage compared to the full-length P450s.

[d]dDFire and DFIRE2 pseudo-energy (lower values signify a better model) (Zhou and Zhou, 2002).

[e] QMEAN6 composite score ranging from 0 to 1 (higher values signify a better model) (Benkert *et al*., 2011).

[f] Verify3D scores ranges from -1 (bad score) to +1 (good score) (Lüthy *et al*., 1992). This program analyses the compatibility of an atomic model (3D) with its own amino acid sequence (1D).

4.3.2. Comparative structural analysis of CPR *Tter* and CPR *Scer*

Comparative analysis of 3D model of CPR *Tter* with CPR *Scer* (2BPO) revealed conservation of structural features between both CPRs. (Figure 4.2). As shown in Figure 4.2 CPR *Tter* showed all features present in CPR *Scer* except few differences in the structure. Some of the regions present in CPR *Scer* (yellow coloured regions in Figure 4.2) are absent in CRP *Tter*. Compared to CPR *Scer*, a small extra loop was present in CRP *Tter* (Figure 4.2). Overall, CPR *Tter* showed all features that are characteristic of CPR proteins.

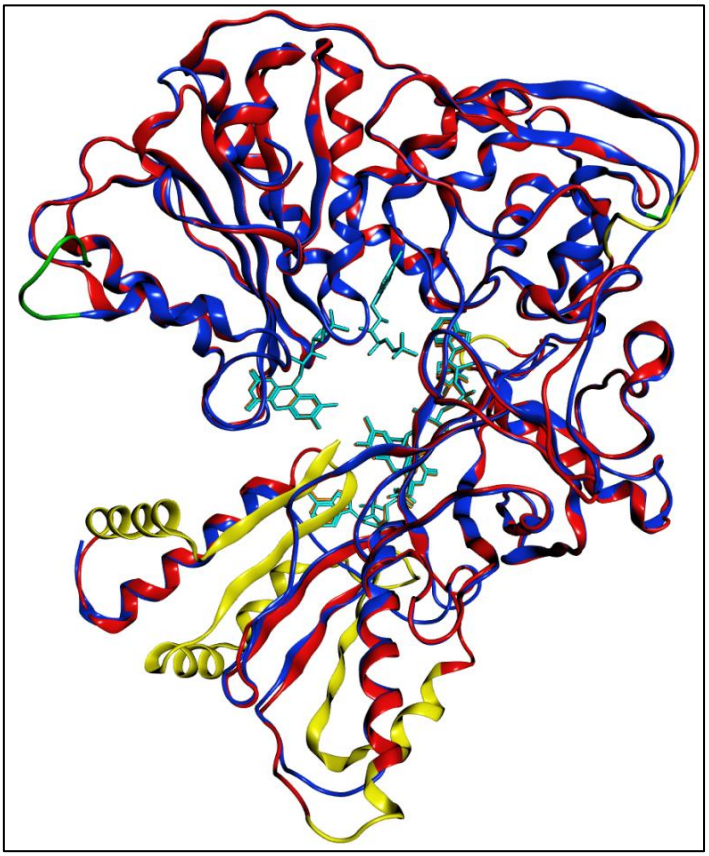

Figure 4.2. Superimposition of CPR *Tter* with the template (CPR *Scer*) generated using MOE software. CPR *Tter* is represented in blue and CPR *Scer* is represented in red. Extra regions of the CPR *Scer* and the CPR *Tter* are represented in yellow and green colours.

4.3.3. Identification of amino acids binding to cofactors in CPR Tter

The main role of CPR is to transfer electrons to the P450. In order to perform this electron transfer CPR should accept electrons from its electron donors and then transfer electrons to the P450. CPR performs this action with the help of cofactors. The characterized CPRs such as CPR *Scer* showed presence of two FMN and one FAD as prosthetic groups and it can bind one NADPH molecule (electron donor) (Yermalitskaya *et al.*, 2006). In order to unravel the nature of cofactors and amino acids responsible to bind these cofactors ligand docking study

was carried out as described in materials and methods. Ligands (FMN, FAD and NADPH) were docked in the CPR *Tter* model (Figure 4.3).

As shown in figure 4.3, CPR *Tter* showed that it was capable of binding to two molecules of FMN, one molecule of FAD and one molecule of NADPH. The binding of a number of cofactors by CPR *Tter* is consistent with reported CPRs where authors suggested CPRs binds two molecules of FMN simultaneously (Lamb *et al.*, 2006). The amino acids involved in binding of these cofactors (Figure 4.2) are conserved in CPR *Tter* when compared to the amino acids from different CPRs (Lamb *et al.*, 2006). Overall, CPR *Tter* showed characteristic structure and cofactors of other CPRs (Lamb *et al.*, 2006) indicating it is capable of transferring electrons to P450s. Future studies involves cloning, expression and functional analysis of CPR *Tter*, where its capability to transfer electrons to P450s will be assessed.

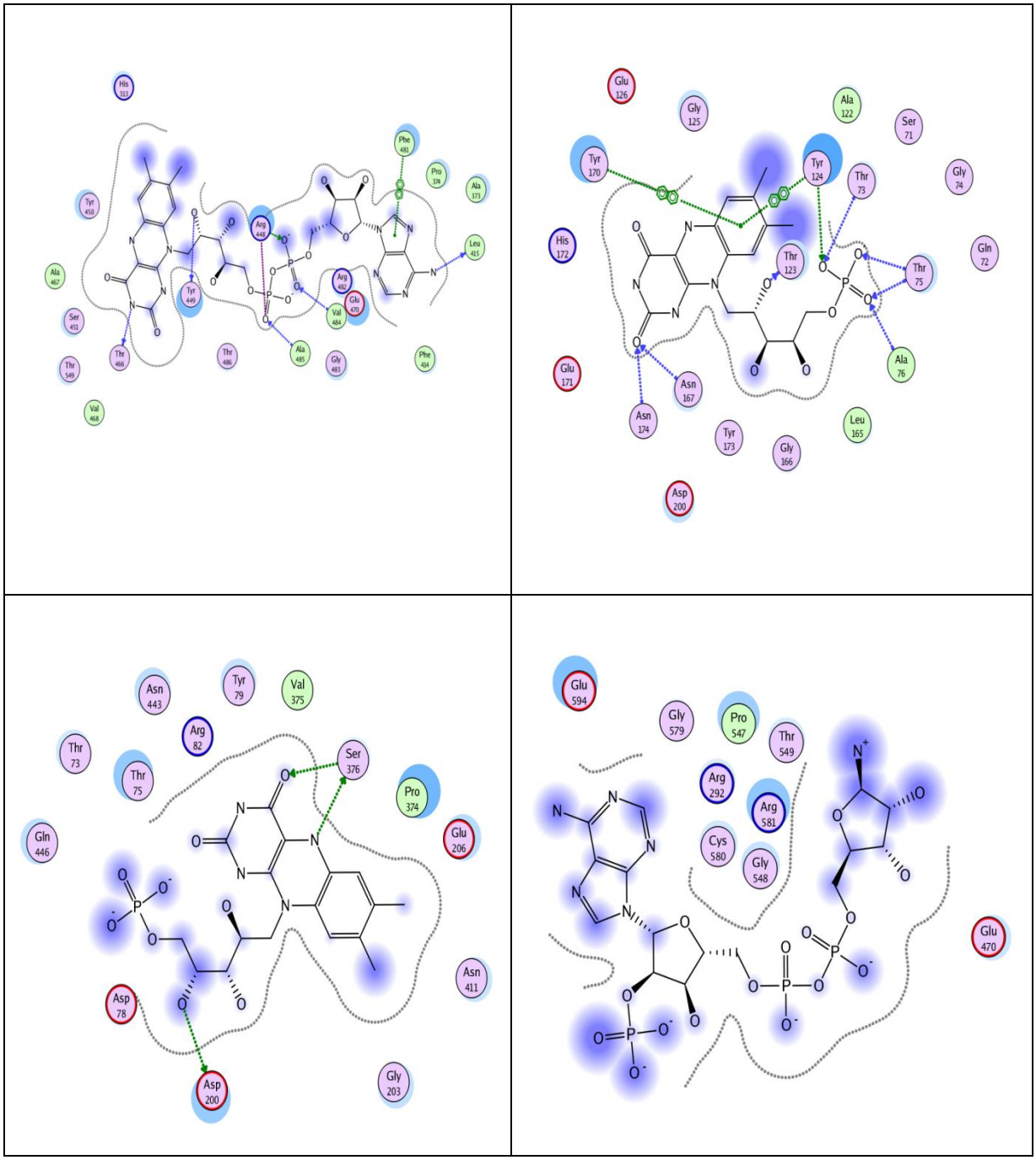

Figure 4.3. Ligand protein interactions of FAD (top left panel), FMN (top right panel), FMN (bottom left) and NAP(D) (bottom right panel) with the CPR *Tter* predicted using MOE software. Amino acids with potential to interact with ligands were shown in the figure.

4.4. Conclusion

In this study, structural analysis of cytochrome P450 reductase from *T. terrestris* (CPR *Tter*) was performed using homology modelling and ligand docking. The 3D model of CPR *Tter* was constructed based on the template (CPR from *S. cerevisiae*). 3D model validation and examination was performed using bioinformatics tools to assess the protein model using specific parameters, which revealed that the model was of a good quality. Comparative analysis and ligand docking studies revealed conservation of characteristic features of CPRs in CPR *Tter*. In future studies CPR *Tter* will be functionally characterized and its thermostablity will be assessed.

REFERENCES

Benkert, P; Biasini, M; Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3): 343–350.

Berka, R. M., Grigoriev, I.V., Otillar, R., Salamov, A., Grimwood, J., Reid, I., Ishmael, N., John, T., Darmond, C., Moisan, M., Henrissat, B., Coutinho, P. M., Lombard, V., Natvig, D. O., Lindquist, E., Schmutz, J., Lucas, S., Harris, P., Powlowski, J., Bellemare, A., Taylor, D., Butler, G., de Vries, R. P., Allijn, I. E., van den Brink, J., Ushinsky, S., Storms, R., Powell, A. J., Paulsen, I. T., Elbourne, L. D H., Baker, S. E., Magnuson, J., LaBoissiere, S., Clutterbuck, A. J., Martinez, D., Wogulis, M., López de Léon, A., Rey, M. W., Tsang, A. (2011). Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nature biotechnology*, 29(10), 922-927.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1): 235-242.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *J Biotechnol.*, 124(1): 128-145.

Chemical Computing Group Inc. Molecular Operating Environment (MOE). 2008. Available at: http://www.chemcomp.com/. [Accessed: 21st April 2015].

Colovos, C., Yeates, T. O. (1993). Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Science*, 2(9), 1511-1519.

Denisov, I. G., Makris T. M., Sligar , S. G., Schlichting, I. (2005). Structure and Chemistry of Cytochrome P450. *Chem. Rev.*, 105: 2253-2277.

Edelsbrunner, H., Facello, M., Fu, P., Liang, J. (1995). Measuring proteins and voids in proteins. In *System Sciences, Proceedings of the Twenty-Eighth Hawaii International Conference*, 5: 256-264.

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design*, 11(5): 425-445.

Hamdane, D., Xia, C., Im, S. C., Zhang, H., Kim, J. J. P., Waskell, L. (2009). Structure and function of an NADPH-cytochrome P450 oxidoreductase in an open conformation capable of reducing cytochrome P450. *Journal of Biological Chemistry*, 284(17), 11374-11384.

Hooft, R. W., Vriend, G., Sander, C., Abola, E. E. (1996). Errors in protein structures. *Nature*, 381(6580), 272-272.

Hubbard, P. A., Shen, A. L., Paschke, R., Kasper, C. B., Kim, J. J. P. (2001). NADPH-cytochrome P450 oxidoreductase structural basis for hydride and electron transfer. *Journal of Biological Chemistry*, 276(31), 29163-29170.

Krieger, E. 2004. Yet Another Scientific Artificial Reality Application (YASARA). Available at: http://www.yasara.org/. [Accessed: 27th April 2015].

Lah, L., Kraševec, N., Trontelj, P., Komel, R. (2008). High diversity and complex evolution of fungal cytochrome P450 reductase: cytochrome P450 systems. *Fungal Genetics and Biology*, 45(4), 446-458.

Lamb, D.C., Kim, Y., Yermalitskaya, L.V., Yermalitsky, V.N., Lepesheva, G.I., Kelly, S.L., Waterman, M.R., Podust, L.M. (2006). A second FMN binding site in yeast NADPH-

cytochrome P450 reductase suggests a mechanism of electron transfer by diflavin reductases. *Structure*, 14(1), 51-61.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26(2), 283-291.

Lüthy, R., Bowie, J.U., Eisenberg, D.R. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356 (6364): 83– 85.

Ohkuma, M., Masuda, Y., Mee Park, S., Ohtomo, R., Ohta, A., Takagi, M. (1995). Evidence that the expression of the gene for NADPH-cytochrome P-450 reductase is n-alkane-inducible in *Candida maltosa*. *Bioscience, biotechnology, and biochemistry*, 59(7), 1328-1330.

Qin, H.-M., Yamamura, A., Miyakawa, T., Kataoka, M., Maruoka, S., Ohtsuka, J., Nagata, K., Shimizu, S., Tanokura, M. (2013). Crystal structure of conjugated polyketone reductase (CPR-C1) from *Candida parapsilosis* IFO 0708 complexed with NADPH. *Proteins: Structure, Function, and Bioinformatics*, 81(11), 2059-2063.

Qin, H.-M., Yamamura, A., Miyakawa, T., Kataoka, M., Nagai, T., Kitamura, N., Urano, N., Maruoka, S., Ohtsuka, J., Nagata, K., Shimizu, S., Tanokura, M. (2014). Structure of conjugated polyketone reductase from Candida parapsilosis IFO 0708 reveals conformational changes for substrate recognition upon NADPH binding. *Applied microbiology and biotechnology*, 98(1), 243-249.

Sali, A., Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.*, 234(3): 779-815.

Sutter, T. R., Sanglard, D., Loper, J. C., Sangard, D. (1990). Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from Candida tropicalis. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins. Journal of Biological Chemistry, 265(27), 16428-16436.

Syed, K., Porollo, A., Lam, Y.W., Grimmett, P.E., Yadav, J.S. (2013) CYP63A2, a catalytically versatile fungal P450 monooxygenase capable of oxidizing higher-molecular-weight polycyclic aromatic hydrocarbons, alkylphenols, and alkanes. *Appl Environ Microbiol.*, 79(8): 2692–702.

van der Brink, J. M., Van Den Hondel, C. A. M. J. J., Van Gorcom, R. F. M. (1996). Optimization of the benzoate-inducible benzoate *p*-hydroxylase cytochrome P450 enzyme system in *Aspergillus niger*. *Applied microbiology and biotechnology*, 46(4), 360-364.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics*, 8(1), 52-56.

Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S. S., Kim, J. J. P. (1997). Three-dimensional structure of NADPH–cytochrome P450 reductase: prototype for FMN-and FAD-containing enzymes. *Proceedings of the National Academy of Sciences*, 94(16), 8411-8416.

Xia, C., Hamdane, D., Shen, A.L., Choi, V., Kasper, C.B., Pearl, N.M., Zhang, H., Im, S.C., Waskell, L., Kim, J.J. (2011a). Conformational changes of NADPH-cytochrome P450 oxidoreductase are essential for catalysis and cofactor binding. *Journal of Biological Chemistry*, 286(18), 16246-16260.

Xia, C., Panda, S. P., Marohnic, C. C., Martásek, P., Masters, B. S., Kim, J. J. P. (2011b). Structural basis for human NADPH-cytochrome P450 oxidoreductase deficiency. *Proceedings of the National Academy of Sciences*, 108(33), 13486-13491.

Yadav, J. S., Loper, J. C. (2000). Cytochrome P450 oxidoreductase gene and its differentially terminated cDNAs from the white rot fungus *Phanerochaete chrysosporium*. *Current genetic*s, 37(1), 65-73.

Yermalitskaya, L.V., Kim, Y., Waterman, M.R., Podust, L.M. (2006). Crystal Structure of the Yeast Cpr Triple Mutant: D74G, Y75F, K78A. PDB ID: 2BPO. Available at: Accessed http://www.rcsb.org/pdb/explore/explore.do?structureId=2bpo. [Accessed: 19th April 2015].

Zhao, Q., Modi, S., Smith, G., Paine, M., McDonagh, P.D., Wolf, C.R., Tew, D., Lian, L.Y., Roberts, G.C., Driessen, H.P. (1999). Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 [Angstrom capital A, ring] resolution. *Protein science*, 8(02), 298-306.

Zhou, H., Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci*., 11(11):2714 –2726.

CHAPTER 5

CONCLUSION

Cytochrome P450 monooxygenases, indeed have remarkable catalytic activity and varied substrate recognition, whereby the capabilities of these enzymes as biocatalysts have been encouraged for the synthesis and production of biotechnological products for human utilization. However, P450s have limitations, which include thermal and chemical, as well as *in vitro* instability. Currently, there is on-going research that aims at exploring and identifying more robust and alternative P450s that would highly adapt and be able to survive extreme industrial conditions and that would also be stable *in vivo*. Apart from biotechnological importance, P450s have been utilized as drug target against pathogens. However, study revealed that pathogens are developing resistance and novel drugs targets need to be identified such that cross-reactions or adverse effects in humans can also be minimized.

This study focused on CYP53 family in fungi. The role of this P450 family as a common alternative anti-fungal drug target has been evaluated. Furthermore, structural aspects of CYP53A and its redox partner, cytochrome P450 reductase (CPR), from the thermophilic fungus *Thielavia terrestris* have been elucidated. This study marks the beginning of our understating on thermostable P450s from eukaryotes.

CYP53A and CPR from *T. terrestris* (*Tter*) have been structurally characterized in this study. Prior to 3D structural analysis, phylogenetic analysis of CYP53 members was performed across the fungal phyla to explore the role of a P450 family (CYP53) in serving as a common drug target against pathogenic ascomycetes and in basidiomycetes, in terms of the wood-degradation process. CYP53A *Tter* is a member of Ascomycota, and like the CYP53

family, is involved in the detoxification of the toxic molecule benzoate. Phylogenetic findings suggested and revealed that this P450 family could serve as a potential common anti-fungal drug target against ascomycete pathogens, because the primary protein structure and gene-structure organization of CYP53 is highly conserved. Furthermore, it was also identified that CYP53 P450s can play a supplementary function in basidiomycetes in the generation of the wood-degrading oxidant, veratryl alcohol, and degradation of other wood-derived compounds. It was suggested that this additional role of basidiomycetes seems to have taken place due to extensive duplication of CYP53 members in their genomes (paralogous evolution). During gene duplication, widespread changes in the primary protein structure ensued to enhance/acquire novel functions for the enrichment of the protein, such as wood degradation.

3D protein modelling and appropriate 3D protein structure validation was conducted for both CYP53A and CPR. This study was the first report on structural analysis of CPR and CYP53A from *T. terrestris*. Validation results for 3D models of CYP53A and CPR indicated that the protein models were of high quality. Structure analysis indicated that CYP53A *Tter* is a biotopic membrane protein. During active site cavity mapping, it was found that 125 amino acid residues lining the active site cavity of CYP53 members, where 78% of the amino acids present are highly conserved. Owing to the fact that the lining of the active site residue of CYP53 members (CYP53A *Tter* used as model P450) is of a highly conserved nature, this suggests that any inhibitor designed for this particular family, could act as common drug target against animal and plant pathogenic ascomycetes.

Currently, immune-compromised individuals with HIV (Human Immunodeficiency Virus) or on treatment with immune-suppressive drugs are at a high risk of invasive and opportunistic fungal infections. Unfortunately, most azole inhibitors developed have been derived from fungal CYP51, which targets fungal and mammalian CYP51 enzymes. Due to

the presence of azole inhibitors, *in vivo*, xenobiotic compounds are also being metabolised leading to toxicity and adverse side effects in such patients such as anaphylaxis. Moreover, due to the prolonged use of azole inhibitors from CYP51, patients develop drug resistance. Therefore, other fungal drug inhibitors need to be identified to circumvent these major medical dilemmas. Therefore, CYP53 could be a promising novel therapeutic drug target due to its potential high substrate specificity and it also does not share any homologs in mammals.

Structural analysis of CPR from *T. terrestris* revealed conservation of basic features of CPR proteins in this redox partner as well. Molecular docking analysis of CPR model revealed amino acids responsible for binding of co-factors such as FMN, FAD and NADPH. Overall, this is first report on structural analysis of CPR from a thermostable eukaryote. The structure of CYP53A and CPR from *T. terrestris* forms the basis for future study focusing on unravelling the factors responsible for thermostability of an eukaryote P450 and their characterization towards use of these enzymes for production of human valuable chemicals at an industrial scale.

OPEN ACCESS Freely available online

PLOS ONE

Cytochrome P450 Monooxygenase CYP53 Family in Fungi: Comparative Structural and Evolutionary Analysis and Its Role as a Common Alternative Anti-Fungal Drug Target

Poojah Jawallapersand[1], Samson Sitheni Mashele[1], Lidija Kovačič[2,3], Jure Stojan[4], Radovan Komel[5], Suresh Babu Pakala[6], Nada Kraševec[5]*, Khajamohiddin Syed[1]*

1 Department of Health Sciences, Faculty of Health and Environmental Sciences, Central University of Technology, Bloemfontein, Free State, South Africa, **2** Conway Institute, School of Medicine and Medical Sciences, University College Dublin, Belfield, Dublin 4, Ireland, **3** Department of Molecular Biomedical Sciences, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, **4** Faculty of Medicine, Institute of Biochemistry, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia, **5** National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia, **6** Department of Biochemistry, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India

Abstract

Cytochrome P450 monooxygenases (CYPs/P450s) are heme-thiolate proteins whose role as a drug target against pathogenic microbes has been explored because of their stereo- and regio-specific oxidation activity. We aimed to assess the CYP53 family's role as a common alternative drug target against animal (including human) and plant pathogenic fungi and its role in fungal-mediated wood degradation. Genome-wide analysis of fungal species revealed the presence of CYP53 members in ascomycetes and basidiomycetes. Basidiomycetes had a higher number of CYP53 members in their genomes than ascomycetes. Only two CYP53 subfamilies were found in ascomycetes and six subfamilies in basidiomycetes, suggesting that during the divergence of phyla ascomycetes lost CYP53 P450s. According to phylogenetic and gene-structure analysis, enrichment of CYP53 P450s in basidiomycetes occurred due to the extensive duplication of CYP53 P450s in their genomes. Numerous amino acids (103) were found to be conserved in the ascomycetes CYP53 P450s, against only seven in basidiomycetes CYP53 P450s. 3D-modelling and active-site cavity mapping data revealed that the ascomycetes CYP53 P450s have a highly conserved protein structure whereby 78% amino acids in the active-site cavity were found to be conserved. Because of this rigid nature of ascomycetes CYP53 P450s' active site cavity, any inhibitor directed against this P450 family can serve as a common anti-fungal drug target, particularly toward pathogenic ascomycetes. The dynamic nature of basidiomycetes CYP53 P450s at a gene and protein level indicates that these P450s are destined to acquire novel functions. Functional analysis of CYP53 P450s strongly supported our hypothesis that the ascomycetes CYP53 P450s ability is limited for detoxification of toxic molecules, whereas basidiomycetes CYP53 P450s play an additional role, i.e. involvement in degradation of wood and its derived components. This study is the first report on genome-wide comparative structural (gene and protein structure-level) and evolutionary analysis of a fungal P450 family.

Citation: Jawallapersand P, Mashele SS, Kovačič L, Stojan J, Komel R, et al. (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): e107209. doi:10.1371/journal.pone.0107209

Editor: Kap-Hoon Han, Woosuk University, Korea, Republic of

Received May 10, 2014; **Accepted** August 8, 2014; **Published** September 15, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files

Funding: KS, PJ, and SSM thank the Central University of Technology (CUT), Bloemfontein, Free State for a grant from the University Research and Innovation fund (to KS and SSM) and Masters in Technology research fellowship (to PJ). SSM also thanks the National Research Foundation (NRF), South Africa for the research grant. NK and RK thank the Slovenian Research Agency Program P1-0104. SBP thanks the University Grants Commission (UGC), New Delhi, India for selecting for the UGC-Faculty Recharge Programme. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: khajamohiddinsyed@gmail.com (KS); nada.krasevec@ki.si (NK)

Introduction

Among microorganisms, fungi, the largest biological kingdom comprising diverse lower eukaryotic microorganisms, have acquired a special place owing to their ability to be pathogens for not only humans but also other animals and plants (Table 1). These lower eukaryotes develop or are constantly developing new strategies to adapt to diverse ecological niches. In order to develop novel drugs by identifying potential novel drug targets and harnessing their potentials for the production of human valuables, a large number of fungal genomes have been sequenced and many fungal genome sequencing projects are currently in progress. Efforts of the Broad Institute of MIT and Harvard (http://www.broadinstitute.org/), Wellcome Trust Sanger Institute (https://www.sanger.ac.uk/), and Joint Genome Institute (JGI) United States Department of Energy (US-DOE) (http://genome.jgi.doe.gov/programs/fungi/index.jsf) resulted in genome sequencing of a large number of fungal species.

Genome sequencing analysis of fungal species revealed the presence of a large number of cytochrome P450 monooxygenases

Table 1. Genome-wide comparative analysis of CYP53 family in fungi.

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Ascomycota								
Magnaporthe grisea	Plant pathogen	1						1
Neurospora crassa	Model organism	1						1
Neurospora discreta	Distantly related to *Neurospora crassa*	1						1
Fusarium graminearum	Plant pathogen	3						3
Fusarium solani f. batatas (Nectria haematococca)	Plant pathogen and animal pathogen (opportunistic human pathogen)	2						2
Fusarium verticillioides	Plant pathogen and animal pathogen (opportunistic human pathogen)	2						2
Fusarium oxysporum	Plant pathogen and animal pathogen (opportunistic human pathogen)	2			1			3
Neosartorya fischeri	Animal pathogen (including human)	1						1
Aspergillus nidulans	Model organism for study of eukaryotic cell biology	1						1
Aspergillus fumigatus	Animal pathogen (opportunistic human pathogen)	1						1
Aspergillus terreus	Human, animal and plant pathogen	1						1
Aspergillus oryzae	Economically important, used for fermentation	2						2
Aspergillus flavus	Plant and animal pathogen (human pathogen)	1						1
Aspergillus niger	Plant and animal pathogen (human pathogen)	1						1
Aspergillus clavatus	Animal pathogen (human pathogen)	1						1
Coccidioides immitis	Animal pathogen (human pathogen)	1						1
Histoplasma capsulatum	Animal pathogen (human pathogen)	0						0
Uncinocarpus reesii	Non-pathogen	1						1
Mycosphaerella fijiensis	Plant pathogen	1						1
Zymoseptoria tritici (formerly named as *Mycosphaerella graminicola*)	Plant pathogen	1						1
Thielavia terrestris	Non-pathogen	1						1
Myceliophthora thermophila	Non-pathogen	1						1
Cochliobolus lunatus	Plant and animal pathogen (human pathogen)	1						1
Total count		**28**			**1**			**29**
Basidiomycota								
Phanerochaete chrysosporium	Model white rot fungus – study of wood degradation			1				1
Postia placenta	Model brown rot fungus – study of wood degradation			1	7			8
Ustilago maydis	Plant pathogen			1				1
Cryptococcus neoformans	Animal pathogen (human)							0
Cryptococcus gattii	Animal pathogen (human)							0
Laccaria bicolor	Symbiotic fungus (ectomycorrhizas)							0
Malassezia globosa	Animal pathogen (human)							0
Puccinia graminis	Plant pathogen		1					1
Sporobolomyces roseus	Non-pathogen		1					1
Phanerochaete carnosa	Model white rot fungus - study of soft wood degradation			6			1	7
Bjerkandera adusta	Wood-degrading white rot fungus			1		7		8
Ceriporiopsis subvermispora	Wood-degrading white rot fungus			4				4
Ganoderma sp.	Wood-degrading white rot fungus			1				1
Ganoderma lucidum	Medicinal mushroom (wood-degrading white rot fungus)			1				1
Phlebia brevispora	Wood-degrading white rot fungus			1				1

Table 1. Cont.

Species	Lifestyle	CYP53 subfamily						Total count
		A	B	C	D	H	NS	
Agaricus bisporus	Litter-degrading fungus			2				2
Serpula lacrymans	Model fungus known as dry rot fungus – study of dry wood degradation			1				1
Stereum hirsutum	Wood-degrading white rot fungus			1				1
Trametes versicolor	Wood-degrading white rot fungus			2				2
Wolfiporia cocos	Wood-degrading brown-rot fungus			9				9
Auricularia delicata	Wood-degrading white rot fungus	1		1				2
Coniophora puteana	Wood-degrading brown rot fungus	1		2				3
Dacryopinax sp.	Wood-degrading brown rot fungus			1				1
Dichomitus squalens	Wood-degrading white rot fungus			1				1
Fomitiporia mediterranea	Wood-degrading white rot fungus			9			1	10
Fomitopsis pinicola	Wood-degrading brown rot fungus			4				4
Gloeophyllum trabeum	Wood-degrading brown rot fungus			1				1
Punctularia strigosozonata	Wood-degrading white rot fungus	1		1				2
Total count		**3**	**2**	**52**	**7**	**7**	**2**	**73**
Total CYP53 members in fungi		**31**	**2**	**52**	**8**	**7**	**2**	**102**

Twenty-three species from ascomycota and 28 species from basidiomycota were used in this study. Identification of CYP53 members in fungal species was carried out as described in the "Materials and methods" section. If no CYP53 member was found in the species, the space was left blank. The abbreviation NS indicates a new subfamily. Fungal species capable of causing diseases in humans were indicated with the word "human" in the table.
doi:10.1371/journal.pone.0107209.t001

(P450s) in their genomes, with some exceptions. P450s are heme-thiolate proteins ubiquitously present across the biological kingdoms [1]. In fungi P450s are known to be involved in both primary and secondary metabolic processes [2,3] and in the degradation of xenobiotic compounds [4]. P450s have been explored as anti-fungal drug targets owing to their key role in fungal physiology through involvement in stereo- and regio-specific oxidation of substrates [5]. Among fungal P450s CYP51, also known as sterol 14α-demethylase, the highly conserved P450 across the biological kingdoms [6], is the primary target of conventional antifungal azole drugs [7]. CYP51 performs demethylation of lanosterol, a key step in biosynthesis of cell membrane ergosterol [6]. Studies have indicated that fungal organisms are developing resistance to azole drugs [8,9]. Furthermore, the currently available anti-fungal drugs have limitations because of similar metabolic pathways between fungi and other organisms (mainly mammals) and hence researchers are in search of alternative novel fungal drug targets [10].

Research on fungal P450s revealed that the P450 family CYP53 can serve as a novel alternative anti-fungal drug target [11]. CYP53 family members are well known as benzoate para-hydroxylases that are involved in the detoxification of a benzoate molecule [12]. Benzoate is a naturally occurring anti-fungal plant material [13] and also a naturally occurring intermediate in the degradation of aromatic compounds in fungi [14–16]. Benzoate exhibits its toxicity by disruption of the membrane, inhibiting essential cellular processes, changing pH balance and inducing stress response in fungi [13,17]. CYP53 P450-mediated para-hydroxylation of benzoate is the only known pathway in fungi that ultimately channels this toxic compound into the β-ketoadipate pathway [18]. Furthermore, the CYP53 gene was found to be essential for fungal species' survival [19]. The CYP53 gene-knock out fungal strain growth was found to be inhibited by the accumulation of toxic intermediate benzoate [19]. This clearly suggests that this P450 is critical in the survival of fungal species, by playing a key role in the detoxification of benzoate.

Considering the fungal resistance to the currently available drugs, especially CYP51 enzyme-based azoles [8], and a preliminary study suggesting that CYP53 P450 family members can serve as novel alternative fungal drug targets [11], in the present study we aimed to understand the role of CYP53 members in fungal physiology per se, performing comparative evolutionary and structural analysis of CYP53 members to check their distribution and structural conservation in fungi. In this way we can determine whether this P450 family can serve as a common drug target against a broad range of fungal pathogens. Furthermore, we also explored its role in adaptation of basidiomycetes to diverse ecological niches such as colonization on wood.

Materials and Methods

Genome data mining and annotation of CYP53 members

Fifty-one fungal species were selected for the analysis of CYP53 member P450s. As shown in Table 1, 23 species from ascomycota and 28 species from basidiomycota were included in this analysis. CYP53 members of the basidiomycete species, such as *Phanerochaete chrysosporium*, *Phanerochaete carnosa*, *Bjerkandera adusta*, *Ganoderma* sp., *Phlebia brevispora*, and *Ceriporiopsis subvermispora*, and ascomycete species, such as *Thielavia terrestris* and *Myceliophthora thermophila,* were retrieved from an author's contributed and original work that has been published and is publicly available [4,20-25]. CYP53 members in the remaining 20 ascomycetes were obtained from the Cytochrome P450 Webpage [26]. Two basidiomycete species, namely *Agaricus bisporus* and *Serpula lacrymans* CYP53 members, were obtained from the Fungal Cytochrome P450 Database (FCPD) [27]. CYP53 members belonging to *Postia placenta* were taken from published literature [28].

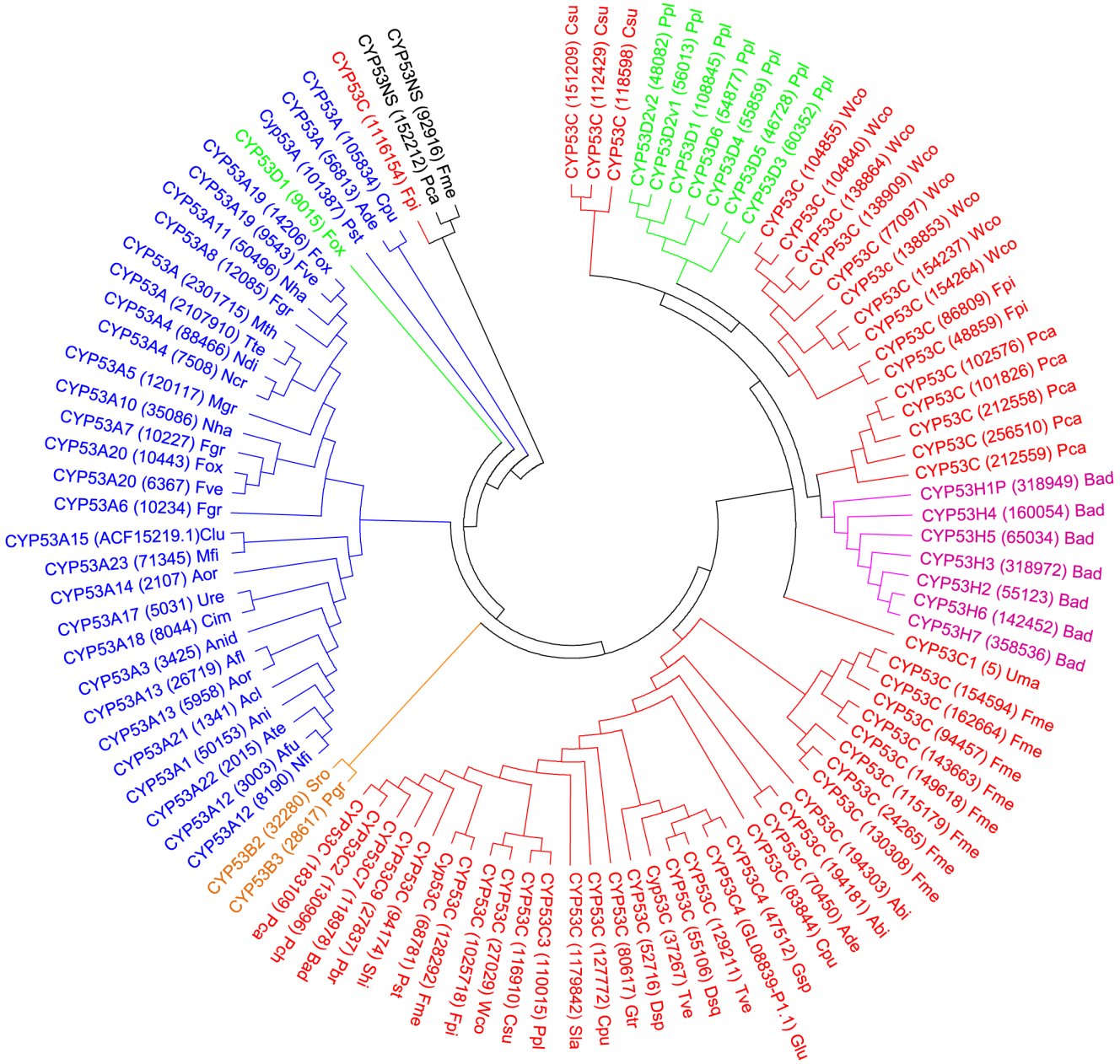

Figure 1. Phylogenetic analysis of CYP53 family in fungi. The tree was constructed with 101 CYP53 P450s belonging to six different CYP53 subfamilies. Phylogeny was inferred using the minimum evolution method [33] and the tree was constructed with MEGA (5.05) software [32]. For details on construction of the tree and the parameter employed for tree construction see the section ''phylogenetic analysis'' under ''Materials and methods''. For ease of visual identity, the tree branch color, protein name, protein ID (parenthesis) and species name were presented in unique color as per sub-family. Fungal species' names were indicated with three letters, where the first letter is taken from the genus name and the other two letters from the species name. Abbreviations: Abi, *Agaricus bisporus*; Acl, *Aspergillus clavatus*; Ade, *Auricularia delicata*; Afl, Aspergillus flavus; Afu, *Aspergillus fumigatus*; Ani, *Aspergillus nidulans*; Aor, *Aspergillus oryzae*; Ate, *Aspergillus terreus*; Bad, *Bjerkandera adusta*; Cim, *Coccidioides immitis*; Clu, *Cochliobolus lunatus*; Cpu, *Coniophora puteana*; Csu, *Ceriporiopsis subvermispora*; Dsq, *Dichomitus squalens*; Fgr, *Fusarium graminearum*; Fme, *Fomitiporia mediterranea*; Fox, *Fusarium oxysporum*; Fpi, *Fomitopsis pinicola*; Fve, *Fusarium verticillioides*; Glu, *Ganoderma lucidum*; Gsp, Ganoderma sp.; Gtr, *Gloeophyllum trabeum*; Mfi, *Mycosphaerella fijiensis*; Mgr, *Magnaporthe grisea*; Mth, *Myceliophthora thermophila*; Ncr, *Neurospora crassa*; Ndi, *Neurospora discreta*; Nfi, *Neosartorya fischeri*; Nha, *Nectria haematococca*; Pbr, *Phlebia brevispora*; Pca, *Phanerochaete carnosa*; Pch, *Phanerochaete chrysosporium*; Pgr, *Puccinia graminis*; Ppl, *Postia placenta*; Pst, *Punctularia strigosozonata*; Shi, *Stereum hirsutum*; Sla, *Serpula lacrymans*; Sro, *Sporobolomyces roseus*; Tte, *Thielavia terrestris*; Tve, *Trametes versicolor*; Uma, *Ustilago maydis*; Ure, *Uncinocarpus reesii*; Wco, *Wolfiporia cocos*.
doi:10.1371/journal.pone.0107209.g001

To identify CYP53 members in the basidiomycete species, such as Wolfiporia cocos, Auricularia delicata, Coniophora puteana, Dacryopinax sp., Dichomitus squalens, Fomitiporia mediterranea, Fomitopsis pinicola, Gloeophyllum trabeum, Punctularia strigosozonata, Stereum hirsutum, and Trametes versicolor, genome data mining was performed as described by one of the authors in his recent publications [24,29], with slight modifications. Blast analysis was performed at the respective species' genome data base

Table 2. Analysis of homology between CYP53 members in fungi.

CYP name	Species name	Homology (%)	CYP name	Species name
Ascomycota				
CYP53A4 (7508)	*Neurospora crassa*	98	CYP53A4 (88466)	*Neurospora discreta*
CYP53A (2107910)	*Thielavia terrestris*	91	CYP53A (2301715)	*Myceliophthora thermophila*
CYP53A19 (9543)	*Fusarium verticillioides*	98	CYP53A19 (14206)	*Fusarium oxysporum*
CYP53A8 (12085)	*Fusarium graminearum*	95	CYP53A19 (9543)	*Fusarium verticillioides*
CYP53A8 (12085)	*Fusarium graminearum*	95	CYP53A19 (14206)	*Fusarium oxysporum*
CYP53A20 (6367)	*Fusarium verticillioides*	99	CYP53A20 (10443)	*Fusarium oxysporum*
CYP53A7 (10227)	*Fusarium graminearum*	97	CYP53A20 (6367)	*Fusarium verticillioides*
CYP53A7 (10227)	*Fusarium graminearum*	97	CYP53A20 (10443)	*Fusarium oxysporum*
CYP53A12 (8190)	*Neosartorya fischeri*	98	CYP53A12 (3003)	*Aspergillus fumigatus*
CYP53A12 (3003)	*Aspergillus fumigatus*	94	CYP53A21 (1341)	*Aspergillus clavatus*
CYP53A13 (5958)	*Aspergillus oryzae*	99	CYP53A13 (26719)	*Aspergillus flavus*
Basidiomycota				
CYP53C4 (47512)	*Ganoderma* sp.	95	CYP53C4 (GL08839-P1.1)	*Ganoderma lucidum*
CYP53D3 (60352)	*Postia placenta*	95	CYP53D5 (46728)	*Postia placenta*
CYP53C (104855)	*Wolfiporia cocos*	67	CYP53C (154237)	*Wolfiporia cocos*

The percentage (%) homology between CYP53 members was obtained from the Cytochrome P450 Webpage [26] based on their highest hit to reference proteins and also estimated using ClustalW2 [38]. P450s showing more than 90% homology were selected and presented in the table. A detailed report on the percentage homology between identified proteins and hit proteins at Cytochrome P450 Webpage [26] was presented in Table S1. As shown in the table, a higher number of CYP53 members from ascomycota showed more than 90% homology, suggesting high conservation of the primary structure in ascomycete species CYP53 members compared to basidiomycete species CYP53 members. For each P450 protein IDs were shown in parenthesis.

doi:10.1371/journal.pone.0107209.t002

that is publicly available [30], using P. chrysosporium CYP53C2 (protein ID: 130996). Considering the presence of CYP53 members in low copies (one or two numbers) in ascomycetes and basidiomycetes, the top 20 hits' proteins were selected for further analysis. The hit proteins were subjected to the NCBI Batch Web CD-Search Tool [31] to separate proteins belonging to the P450 superfamily. This software groups the proteins into different superfamilies based on the conserved domain characteristics of the protein family. The proteins that are grouped under the P450 superfamily were selected for further assignment to the P450 family and subfamily. Assigning the family and subfamily names to the P450 proteins was performed in the same way as described by one of the authors in his recent study [24,29]. Briefly, individual proteins were blasted against all named fungal P450s at the Cytochrome P450 Webpage [26]. A family and subfamily were assigned to the P450 proteins based on standard International P450 Nomenclature criteria, i.e. >40% homology for a family and >55% homology for a subfamily. Among the selected proteins those grouped under the CYP53 family were used in the analysis. The Cytochrome P450 Webpage [26] was visited to check for the presence of CYP53 members, if any, in the basidiomycetes Ustilago maydis, Cryptococcus neoformans, Cryptococcus gattii, Laccaria bicolor, Malassezia globosa, Puccinia graminis and Sporobolomyces roseus. A CYP53 member for Cochliobolus lunatus was obtained from one of an author's contributed work, which is publicly available [11].

Phylogenetic analysis

Phylogenetic analysis of CYP53 members was carried out using the Molecular Evolutionary Genetics Analysis (MEGA) software [32] in the same way as described in one of the author's recent publications [24,25]. Phylogenetic analysis was inferred using the minimum evolution method [33]. The minimum evolution method is widely used in P450 research, based on pairwise distance algorithms for the reconstruction of phylogenies [24,25,37]. In this study we used the minimum evolution method for phylogenetic analysis of CYP53 member P450s. The evolutionary distances were computed using the Poisson correction method [34] and are in the units of the amino acid substitution per site. The minimum evolution tree was searched using the close-neighbor-interchange algorithm [35]. The neighbor-joining algorithm [36] was used to generate the initial tree.

Intron-exon analysis

Gene structure organization of CYP53 family members was carried out as described by an author in his recent publication [25]. Briefly, each CYP53 member gene was accessed at its genome data base at the JGI, US-DOE (http://genome.jgi.doe.gov/programs/fungi/index.jsf; accessed on 5 Feb, 2014) or Broad Institute of MIT and Harvard (http://www.broadinstitute.org/; accessed on 5 Feb, 2014). For each P450 the size of the exons and the location of introns were recorded. A schematic diagram showing horizontal lanes representing the exons and vertical lanes representing the introns' location were drawn. The length of the horizontal lane corresponds to the gene length. CYP53 members that showed high conservation in terms of the size of exons and the location of introns were shown in the figure.

Analysis of homology

To identify the percentage homology between CYP53 members, we performed ClustalW2 multiple sequence analysis [38]. CYP53 members in FASTA format were included in the analysis and the result summary showing the percentage identity matrix was downloaded. After the file had been downloaded, the results were converted into table format and checked for the percentage homology between CYP53 members.

A

```
Conservation:                9                                        95  5      75 7 9 7  9 57  9
CYP53A__2107910__Tte       1 MA------------VINLILSPWA------PAALLV-AAVVYYVYPYLVTYRHLRHI-PAPFPAQFTNWW   50
Conservation:                7    9   9   7   9 5 9  79569 7595 5  79  7999979577   5597 755  57797
CYP53A__2107910__Tte      51 LLLVCRRGDRYATVDKVHKKLGPVVRIQPNHVSILDDEAIQAIYGHGNGFLK--SDFYDAFVSIQRGLFN  118
Conservation:                59 97995999 7 779 96 59   595 5  5   6  57                 5  757 9577
CYP53A__2107910__Tte     119 TRDRAEHTRKRKIVSHTFSVKSVAQFEPYIHSNLELFVRQLDDLIARSTSP-DG----AAHLDCLHWFNY  183
Conservation:                7 7977799597 99979  5 5 7           55 55 575 975557759  9 779 7 5 99
CYP53A__2107910__Tte     184 LAFDVIGDLAFGAPFGMLSSGADMAEVRASPDSPPIYAPAIEILNRRGEVSATLGILPQLKPYAKYLPDP  253
Conservation:                99  9  7   95755 7 7  97          5 979 77  759  9 57   9755977955597
CYP53A__2107910__Tte     254 FFSKGLSAVEKLAGIAIARVKARL--ENPP--PASRKDLLQRLIDGRDEKGQPLGREELTAEALTQLIAG  319
Conservation:                5795977 595775   59 7  77  557   5    59    76 7 99  55 9577 995 5 999
CYP53A__2107910__Tte     320 SDTTSNSSCALLFHAARTPGVLARLRAELDAAVPADLVVPTFDLVRDLPYLSAVVNETLRFHSTSGIGLP  389
Conservation:                9 67     77   5  6 79 77979 9  977 7799 975 9 9799 5    5  99 797997 9
CYP53A__2107910__Tte     390 REVPRD-GQGVHIAGHYFPPGTVLSVPTYSIHHSKEIWGPDADEFRPERW-D--NLTARQKNAFIPFSYG  455
Conservation:                977959996997 76     9 57 7     5   55  97 5997  7 67   9
CYP53A__2107910__Tte     456 PRACVGRNVAEMEMKLIVATWARRYDVSLQQD-HMDTREGFLRKPLGLKIALKRRQ------  510
```

B

```
Conservation:                9
CYP53C2__130996__Pch       1 MA--------------------VIEALTQL--------------DLKSW------LLL--IPALAIVAHI   28
Conservation:                                  567  8    6  9                6   8   7           565
CYP53C2__130996__Pch      29 LIWLL--DPHGIRSY--PGPLLAK--FSDAWL--GYVAAQGHRSEVVHDLHKQYG---------TFVRIA   81
Conservation:                    7  576      5   665       6     8  5          7   56  65 676  7
CYP53C2__130996__Pch      82 ------PNHLSIADP--DALQVVYGH-GTGTLK--SNFYDAFVSI---QRGLFNTRSRSEHARKRKIVSH  137
Conservation:                 76 55    6        5   5                                    6   6   5
CYP53C2__130996__Pch     138 IFSQKSVLEFEPHVRLYVKQLIQQWDRLYEAGAKGL---------------------VWLDCLPW-YNYL  185
Conservation:                 7855658576 57565    5
CYP53C2__130996__Pch     186 AFDIIGDLAFGAPFGMLLAARDAAPVAVD-HEQAMASYGKEK------------SEVQYIPAVQVINDRG  242
Conservation:                    7  6                                    5   5
CYP53C2__130996__Pch     243 TYSASLGVLPPWMRPI----VKLF---PWFRRGQ-KAVKQLAGIAVAAV---------AQRLTTP-----  290
Conservation:                     8 8  6          6         66  5     5 66 5656 5  5  5
CYP53C2__130996__Pch     291 ---TDRVDLLGKLQEGRDDD----GNLM---GKEELTAE-ALTQLIAGSDTTSNSSCAITYYLAKYPDAQ  349
Conservation:                 68  667                          7 77  55 9 68  5  5 7868        5
CYP53C2__130996__Pch     350 RKLQQELDEALGSDD-----------EPVSTFDQVKRLPYLQAVIDEALRIHSTSGIGLPRLVPKG---G  405
Conservation:                    6  5  8  57 7 5  65    78 8   7 9598                          6 58
CYP53C2__130996__Pch     406 MTVCGRFFPEGTVLSVPTYTIHRDEEVWGKDPEVFRPERWF----EQDK-----------NAVQKTYNPF  460
Conservation:                8 96769 86 69  66    65   57
CYP53C2__130996__Pch     461 SFGPRSCIGRNLANMELLIIVSSILRRYDFVLEDPDKP------------------FDTME-GFLRKPVE  511
Conservation:                  5    7
CYP53C2__130996__Pch     512 CVVGIRRRTL-----------  521
```

Figure 2. Analysis of amino acid conservations in CYP53 family members of ascomycota (A) and basidiomycota (B). Analysis of amino acid conservations was carried out using PROMALS3D [39]. CYP53A from *Thielavia terrestris* and CYP53C2 from *Phanerochaete chrysosporium* were presented as a representative of ascomycota (A) and basidiomycota (B) CYP53 members. The residues conserved in CYP53 members of ascomycota (A) and basidiomycota (B) are shown with the conservation index [40] on top of the amino acid residue. Complete alignment of CYP53 members of ascomycota and basidiomycota and a conservation index for the amino acids was presented in Fig. S1.
doi:10.1371/journal.pone.0107209.g002

Analysis of amino acid conservation

The number of amino acids conserved in CYP53 members across the fungi and between ascomycota and basidiomycota was determined using PROfile Multiple Alignment with predicted Local Structures and 3D constraints (PROMALS3D) [39]. PROMALS3D aligns multiple protein sequences and/or structures, with enhanced information from database searches, secondary structure prediction, 3D structures or user-defined constraints and it will also give a conservation index [40]. The conservation index follows numbers above 4, where 9 is the invariantly conserved amino acid across the input sequences.

Homology modeling

3D-modelling of CYP53 member P450s namely; CYP53A (protein ID: 2107910) from *T. terrestris* (Tter) and CYP53C2 (protein ID: 130996) from *P. chrysosporium* (Pchr) was carried out as described by one of the authors in his publication [41], with slight modifications. The Basic Local Alignment Search Tool (BLAST) was used for selecting the closest homologues (template) available in the Protein Data Bank (www.pdb.org; accessed on 5 Feb, 2014). Among the hits at the PDB databank, recently crystallized full-length P450 protein CYP51 from *Saccharomyces cerevisiae* [42] was superior. Hence this P450 was used as a template. The coordinates of the crystal structures of CYP51 (PDB ID: 4LXJ) [42] were used as templates to build the models of CYP53 P450s. 3D-models of the Pchr and Tter were generated using the homology modeling program Modeller 9v11 [43]. The modelling was performed with default parameters using the "allHmodel" protocol to include hydrogen atoms and the "HETATM" protocol to include prosthetic group HEM (heme). The 3D-model's accuracy was validated using DFire [44], QMEAN [45], and Verify3D [46]. Heme-binding residues were identified using 3DLigandSite [47]. Structure alignment between the template and CYP53A of *T. terrestris* and the CYP53C2 of *P. chrysosporium* was performed using PROMALS3D [39]. P450 characteristic secondary structure annotations and substrate

```
Conservation:                                  9
CYP53C__1025718__Fpi_       1  ----------MS----------------------------SVVDQLTG--------LPVAAWA--GLVV   21
Conservation:                                                 99    8 55  95      5  6  7   8 5 8
CYP53C__1025718__Fpi_      22  A--AVVLVHLVPLIT--DPYQVRSY--PGPFL--AKISDAWLGWVAAQGHRSEVVHELHQKHG-------   76
Conservation:                     5567576   6686  6 86  6855  5 5 85  85 5          7  556  755776 57
CYP53C__1025718__Fpi_      77  --KFVQIAPN---HVSVSDPDALQVIYAHGNGTL-KSTFYDAFVSI---QRGLFNTRSRPEHARKRKIVS  137
Conservation:                   66 7 7  685        5  566 5   5  5  5           9    6 56 6 666  55
CYP53C__1025718__Fpi_     138  HIFSQKSVLEFEPYTRMHIKKLMNQWDRLYDLAMKGGSGEEGE-G-----WQGRDGRLWLDILPW-YNYL  200
Conservation:                    8866678768 59796    8 6  5                                  55
CYP53C__1025718__Fpi_     201  AFDIIGDLAFGAPFGMLDACADAAPVAISHEKAMSSYGETDT------------PEITYFPAVQILNDRG  258
Conservation:                          6  8    5      6                 5   6  85             5 8 8  8
CYP53C__1025718__Fpi_     259  EYSASLGVLPPHWRPIVKL-LPWYRKGNKAVQR-LAGIAIAQVAKRLAMHTD--------RSDLLGKLQE  318
Conservation:                    5 5 8      566566    6 7867857 5 5 5  6 55       596 878  6
CYP53C__1025718__Fpi_     319  GKD-DEGNPMGREELTAEALTQLIAGSDTTSNSSC-AITYHLAANPMVQQKLQRELDEALGND------D  380
Conservation:                          7 88 565 9 6955   5 7979 66   58    8  6  95 5868 8  975
CYP53C__1025718__Fpi_     381  DPVSTFEQVKRLPYLEAVINEGLRLHSTSGIGLPRLVP--EGGLTVCGRFFPEGTVLSVPSYTIHRDQDV  448
Conservation:                    78 8   669697                          6 688 97879 97568 567   5655 558 5
CYP53C__1025718__Fpi_     449  WGSDADAFRPERWF-EQDE-----------KAIQKTFNPFSFGPRSCVGRNLASMELLIILSSILRRYHF  506
Conservation:                     9
CYP53C__1025718__Fpi_     507  VLEHPEQG-------------------------------------------------------------  514
Conservation:                       5  56  5 58    7
CYP53C__1025718__Fpi_     515  ---LDTKEGFLRKPVECRVG-IKRRTV---------  537
```

Figure 3. Analysis of amino acid conservations in CYP53C subfamily of basidiomycota. Analysis of amino acid conservations was carried out using PROMALS3D [39]. CYP53C from *Fomitopsis pinicola* is presented as a representative of CYP53C members. The residues conserved in CYP53C members are shown with the conservation index [40] on top of the amino acid residue. Complete alignment of CYP53C members of basidiomycota and a conservation index for the amino acids was presented in Fig. S1.
doi:10.1371/journal.pone.0107209.g003

recognition sites (SRS) in modelled P450s were identified according to their alignment with the template P450s and standard SRS localization regions, as described in the literature [48,49]. Protein models were visualized using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/; accessed on 5 Feb, 2014).

Active site cavity residues mapping

To identify the amino acid residues lining the active site cavity CASTp was used [50]. The CYP53A of *T. terrestris* (abbreviated as Tter) 3D-model generated in this study was used to predict the protein active site cavity using CASTp. The cavity showing the higher volume and covering the SRS regions was selected and the program automatically listed the residues lining the active site. The active site cavity structure and the residues lining the cavity were presented as a figure using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC.

Results and Discussion

CYP53 distribution in fungi

The CYP53 family is one of the P450 families apart from CYP51 and CYP61 that are conserved between the phyla ascomycota and basidiomycota [2,27]. In this study we screened 51 fungal species belonging to ascomycota (23 species) and basidiomycota (28 species) for analysis of CYP53 family members in their genomes. Genome data mining of ascomycetes (23 species) and basidiomycetes (28 species) revealed the presence of one to nine copies of CYP53 members in their genomes (Table 1). The CYP53 family member count ranged from one to three in ascomycetes and one to ten in basidiomycetes. The basidiomycete species *F. mediterranea* showed the maximum number of CYP53 members (10 CYP53 P450s) in its genome. No CYP53 member was identified in the ascomycete *Histoplasma capsulatum* and in the basidiomycetes *C. neoformans*, *C. gattii* and *M. globosa* or the symbiotic *L. bicolor* (Table 1). Overall, ascomycete species showed a lower number of CYP53 member P450s in their genomes compared to basidiomycete species (Table 1), suggesting a possible duplication of CYP53 members after the phylum divergence. Moreover, our analysis revealed the complete absence of CYP53 member P450s in phyla zygomycota and chytridiomycota. Furthermore, in ascomycota only species belonging to subphyla pezizomycotina showed CYP53 members in their genomes and CYP53 member P450s were not found in species of the subphyla saccharomycotina and taphinomycotina, which is in accordance with the smaller size of the P450ome in relation to the growth form of the fungus [51]. Overall, contrary to the established assumption that this family is conserved in fungi, our study showed that CYP53 is not conserved across the fungal species. In future, further genome sequencing analysis of species belonging to chytridiomycota and zygomycota and the subphylum taphrinomycotina could be performed that may provide more information on the presence of this protein family in their genome. However, considering the life style and small size genomes of saccharomycotina species, the absence of CYP53 family members is expected.

Analysis of the CYP53 family suggested the dominance of specific CYP53 subfamilies in ascomycota and basidiomycota (Table 1). Ascomycete species showed only the CYP53A subfamily in their genomes, with the exception of *Fusarium oxysporum,* which showed a single copy CYP53 member belonging to the CYP53D subfamily (Table 1). In contrast to ascomycete species, basidiomycete species showed divergence in CYP53 subfamilies. Five subfamilies were observed in basidiomycetes, i.e. CYP53A, CYP53B, CYP53C, CYP53D, and CYP53H (Table 1). Our analysis of CYP53 members in basidiomycetes revealed the presence of two new CYP53 subfamilies in *P. carnosa* and *F. mediterranea*. Among the CYP53 subfamilies observed for basidiomycota, the CYP53C subfamily was dominant, with 52 members, followed by CYP53D (eight members) and CYP53H (seven members). A single copy of CYP53A members was found in *A. delicata*, *P. strigosozonata*, and *C. puteana* (Table 1). Considering the presence of CYP53A and CYP53D subfamilies in both phyla, one can assume that after the divergence of phyla, ascomycete species might have lost CYP53 subfamilies such as CYP53B, C and H. On the other hand, basidiomycete species enhanced CYP53 numbers in their genome, possibly by genome duplication of CYP53 members in view of the possible require-

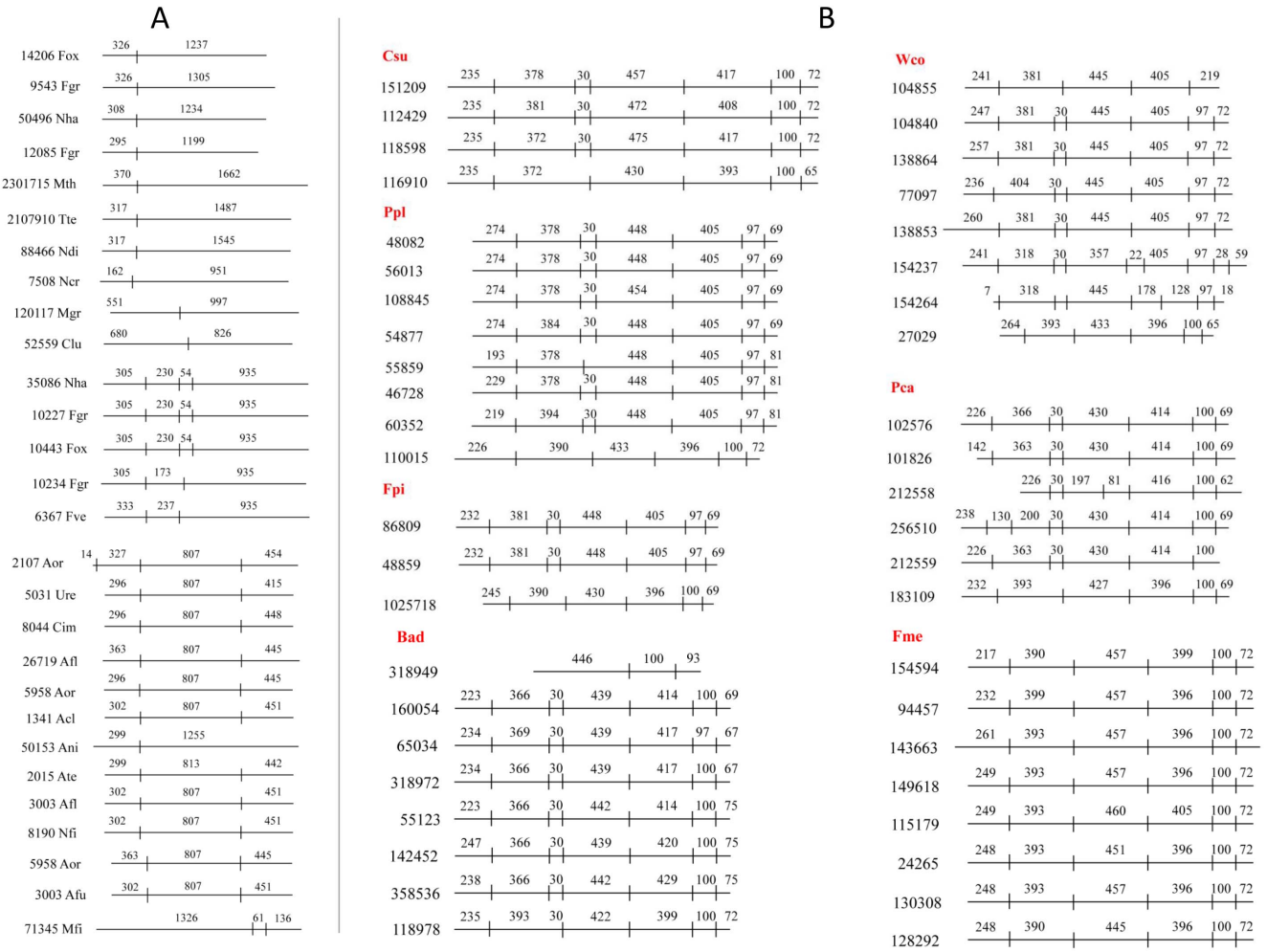

Figure 4. Gene-structure analysis of CYP53 family in (A) ascomycete species and (B) basidiomycete species. Intron-exon analysis was carried out as described in the "Material and methods" section. Horizontal lines indicate gene size and vertical lines indicate introns. For each CYP53 gene the size of the exons (base pairs) and protein ID from the JGI US-DOE [30] is shown in the figure. Abbreviations: Acl, *Aspergillus clavatus*; Afl, Aspergillus flavus; Afu, *Aspergillus fumigatus*; Ani, *Aspergillus nidulans*; Aor, *Aspergillus oryzae*; Ate, *Aspergillus terreus*; Bad, *Bjerkandera adusta*; Cim, *Coccidioides immitis*; Clu, *Cochliobolus lunatus*; Csu, *Ceriporiopsis subvermispora*; Fgr, *Fusarium graminearum*; Fme, *Fomitiporia mediterranea*; Fox, *Fusarium oxysporum*; Fpi, *Fomitopsis pinicola*; Fve, *Fusarium verticillioides*; Mfi, *Mycosphaerella fijiensis*; Mgr, *Magnaporthe grisea*; Mth, *Myceliophthora thermophila*; Ncr, *Neurospora crassa*; Ndi, *Neurospora discreta*; Nfi, *Neosartorya fischeri*; Nha, *Nectria haematococca*; Pca, *Phanerochaete carnosa*; Ppl, *Postia placenta*; Tte, *Thielavia terrestris*; Ure, *Uncinocarpus reesii*; Wco, *Wolfiporia cocos*.
doi:10.1371/journal.pone.0107209.g004

ment of these P450 family members to adapt to diverse ecological niches.

Phylogenetic analysis of CYP53 P450 family

In order to understand the evolution of the CYP53 family and its distribution in fungi, we performed evolutionary analysis of the CYP53 family using the minimum evolution method [33]. Minimum evolution analysis of CYP53 members showed subfamily-specific and species-specific alignment/grouping of CYP53 members (Fig. 1), suggesting that after divergence of phyla (ascomycota and basidiomycota) CYP53 members have been subjected to phylum-specific amino acid changes in their structure. The most striking feature was that CYP53 members belonging to a particular basidiomycete species were grouped together (Fig. 1). This clearly indicates that paralogous evolution of CYP53 members, possibly *via* genome duplication, occurred in basidiomycete species. In a recently published study [25], we observed the same phenomenon of genome duplication of member P450s in basidiomycete species. Furthermore, we also showed that these P450 duplications were necessitated by the fungal species to adapt to diverse ecological niches [25]. Interestingly, CYP53D1 of *F. oxysporum* (ascomycete) did not align with its counterpart present in *P. placenta* (basidiomycete) (Fig. 1), suggesting that extensive changes specific to phyla might have occurred in their primary structure.

High conservation of primary structure of CYP53 members in ascomycota

From the above study it is highly positive that after divergence of ascomycota and basidiomycota, CYP53 members have been subjected to phyla-specific changes or conservation in their primary structure. In order to understand these phyla-specific changes or conservations in CYP53 members, we followed two methods. Firstly we analyzed the percentage homology and secondly we deduced amino acids conserved in CYP53 members in both ascomycetes and basidiomycetes.

Table 3. CYP53A Tter and CYP53C2 Pchr 3D-models validation.

CYP name[a]	Sequence Identity (%)[b]	Length[c]	Coverage[c] (%)	dDFIRE[d]	DFIRE2d[d]	QMEAN6 score[e]	Average Verify3D score ± SD[f]
CYP53A Tter	16	510	100%	−1101.04	−849.785	0.57	0.23±0.22
CYP53C2 Pchr	18	519	99.6%	−1028.1	−789.42	0.57	0.32±0.24

Both CYP53 P450 models were based on the template CYP51 (PDB ID: 4LXJ) from *Saccharomyces cerevisiae* [42] and were generated using Modeller [43]. Abbreviations: Tter *Thielavia terrestris*; Pchr, *Phanerochaete chrysosporium*.

[a] Models were based on the template CYP51 (lanosterol 14α-demethylase) from *Saccharomyces cerevisiae* (PDB ID: 4LXJ) [42].

[b] Sequence identity between CYP53 P450s (CYP53A Tter and CYP53C2 Pchr) and the template CYP51 (PDB ID:4LXJ).

[c] Number of P450s amino acids modeled and their percentage compared to the full-length P450s.

[d] dDFire and DFIRE2 pseudo-energy (lower values signify a better model) [44].

[e] QMEAN6 composite score ranging from 0 to 1 (higher values signify a better model) [45].

[f] verify3D scores ranges from -1 (bad score) to +1 (good score). This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D) [46].

doi:10.1371/journal.pone.0107209.t003

ClustalW2 analysis of CYP53 members revealed a high percentage homology among CYP53 members (Table 2) in ascomycota; some of the members showed >90% homology compared to CYP53 members in basidiomycota. The observed high percentage homology in CYP53 members of ascomycota (Table 2) might be due to the dominance of a single CYP53A subfamily. It is noteworthy that although the CYP53C subfamily is dominant in basidiomycota (Table 1), most of its members seem to be subjected to major amino acid changes, as the percentage homology between CYP53C members is not high with exception of a few P450s, as observed for CYP53A members for ascomycota (Table 2).

To link the high percentage homology observed for CYP53 members of ascomycetes towards conservation of amino acid in their primary structure, we performed amino acid conservation studies using PROMALS3D (Figures 2 and 3; Fig. S1). PROMALS3D analysis of CYP53 members across fungi suggested conservation of eight amino acids (Fig. S1A). Conservation of only eight amino acids in CYP53 members across fungi is understandable, considering the high diversity of CYP53 members across fungal species (five subfamilies and two new subfamilies). The most striking difference was observed in the number of amino acids conserved in the CYP53 members of ascomycota and basidiomycota (Figures 2 and 3 and Fig. S1). A hundred and three amino acids were found conserved in CYP53 members of ascomycota compared to CYP53 members of basidiomycota, which showed only seven amino acids conserved in their primary structure (Fig. 2). This strongly suggests that the observed high percentage homology between CYP53 members of ascomycota is due to the high conservation of amino acids in their primary structure.

One can argue that the high conservation of amino acids (103 amino acids) in CYP53 members of ascomycota (Fig. 2A) is due to the presence of a single CYP53A subfamily whereas five subfamilies and two new subfamilies exist in basidiomycota. To rule out this argument, we present two types of evidence. Firstly, we collected CYP53A members from ascomycete species belonging to 11 different genera (Table 1), suggesting the high diversity of host species, which should thus reflect in CYP53A primary structure as well. However, this was not true, as ascomycete CYP53 members showed high conservation in the primary structure (Fig. 2). Secondly, we estimated the number of amino acids conserved in the CYP53C subfamily alone (Fig. 3), the subfamily that is dominant in basidiomycota. Interestingly, our analysis revealed conservation of only 20 amino acids in CYP53C subfamily members in basidiomycota (Fig. 3), further strengthening our hypothesis that basidiomycota CYP53 members have been subjected to extensive primary structure changes. Further studies were carried out to map the location of conserved amino acids to extrapolate the effect of the conservation in CYP53 substrate specificity or catalytic activity, if any.

Gene conservation and genome duplications of CYP53 members

The above study indicated high conservation of CYP53 members' primary structure (at a protein level) in ascomycetes compared to basidiomycetes. To gain insight into this aspect, we further analyzed the gene structure of CYP53 members (Fig. 4). Analysis of the size of exons and the location of introns indicated high conservation of the gene-structure in CYP53 members belonging to both fungal phyla ascomycota (Fig. 4A) and basidiomycota (Fig. 4B). Gene structure analysis suggested that some ascomycete species, such as *F. oxysporum*, *Fusarium solani f. batatas (Nectria haematococca)*, *Fusarium verticillioides*, and *Fusarium graminearum,* contain two types of ortholog P450s in

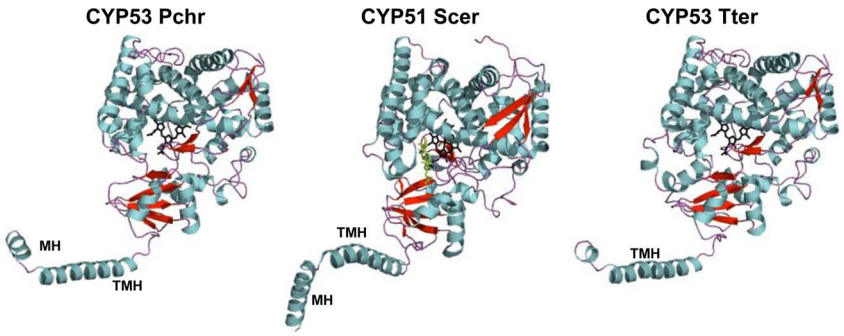

Figure 5. Structural analysis of CYP53 P450s. 3D-models for CYP53C2 from *Phanerochaete chrysosporium* (CYP53 Pchr) [4] and CYP53A from *Thielavia terrestris* (CYP53 Tter) [24] were constructed based on the template CYP51 from *Saccharomyces cerevisiae* (CYP51 Scer) (PDB ID: 4LXJ) [42]. The CYP51 Scer crystal structure was downloaded from the Protein Data Bank (www.pdb.org) with ID: 4LXJ. The overall structures of P450s were presented using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/). The parameters used for validation of 3D-models of CYP53 Tte and CYP53 Pchr were presented in Table 3. The heme prosthetic group is shown in black color and the bound substrate for CYP53 Scer in green color. Alpha-helices and beta-strands are shown with blue and red. The membrane helix (MH) and the trans-membrane helix (TMH) are indicated in the models. P450 characteristics secondary structure notations are presented in Fig. 6.
doi:10.1371/journal.pone.0107209.g005

```
                                MH                  TMH
Pchr          1  ---------AVIEALTQLDLKSWLLLIPALAIVAHILIWLLDPHGIRSYPGPLLAKFSDAWL--GYVAAQ  59
4LXJ          1  SIVGEALEYVNIGLSHFLALPLAQRISLIIIIPFIYNIVWQLLYSLRKDRPPLVFYWIPWVG--SAVVYG  68
Tter          1  -----------MAVINLILSPWAPAALLVAAVVYYVYPYLVTYRHLR--HIPAPFPAQFTNWWLLLVCRR  57
                        αA        β1_1       β1_2     αB        β1_5  αB'              αC
Pchr         60  GHRSEVVHDLHKQYGTFVRIA---PNHLSIADPDALQVVYGHGTGTLKSNFYDAF-V--SIQRGLFNTRS  123
4LXJ         69  MKPYEFFEECQKKYGDIFSFVLLGRVMTVYLGPKGHEFVFNAKLADVSAEAAYAHLTTPVFGKGVIYDCP  138
Tter         58  GDRYATVDKVHKKLGPVVRIQ---PNHVSILDDEAIQAIYGHGN-GFLKSDFYDAFV--SIQRGLFNTRD  121
                                                                          SRS-1
                       αC                  αD                        β3_1        αE
Pchr        124  RSEHARKRKIVSHIFSQKSVLEFEPHVRLYVKQLIQQWDRLYEAGAK----GLVWLDCLPWYNYLAFDII  189
4LXJ        139  NSRLMEQKKFVKGALTKEAFKSYVPLIAEEVYKYFRDSK-NFRLNER----TTGTIDVMVTQPEMTIFTA  203
Tter        122  RAEHTRKRKIVSHTFSVKSVAQFEPYIHSNLELFVRQLD--DLIARSTSPDGAAHLDCLHWFNYLAFDVI  189
                      αF'                 αF        αF''                 αG
Pchr        190  GDLAFGAPFGMLLA--------------ARDAAPVAVDHEQAMAS---YGK-EKS------EVFPWFRRG  235
4LXJ        204  SRSLLGKEMRAKLD--------------TDFAYLYSDLDKGFTPI---NFV-FPN------LPLEHYRKR  249
Tter        190  GDLAFGAPFGMLSSGADMAEVRASPDSPPIYAPAIEILNR-RGEVSATLGILPQLKPYAKYLPDPFFSKG  258
                                          SRS-2                                SRS-3
                          αG                  αH                  αI
Pchr        236  QKAVKQ-LAGIAVAAVAQRLT-TPTDRVDL-LGKLQEGRDDDGNLMGKEELTAEALTQLIAGSDTTSNSS  302
4LXJ        250  DHAQKA-ISGTYMSLIKERRKNNDIQDRDLIDSLMKNSTYKDGVKMTDQEIANLLIGVLMGGQHTSAATS  318
Tter        259  LSAVEKLAGIAIARVKARLEN-PPPASRKDLLQRLIDGRDEKGQPLGREELTAEALTQLIAGSDTTSNGS  327
                                                                    SRS-4
                          αJ                  αJ'       αK                  β1_4
Pchr        303  CAITYYLAKYPDAQRKLQQELDEALGSDDEPVSTFDQVKRLPYLQAVIDEALRIHSTSGIG-LPRLVP--  369
4LXJ        319  AWILLHLAERPDVQQELYEEQMRV-LDGGKKELTYDLLQEMPLLNQTIKETLRMHHPLHSL-FRKVMK--  384
Tter        328  CALLFHAARTPGVLARLRAELDAA-VPADLVVPTFDLVRDLPYLSAVVNETLRFHSTSGIGLPREVPRDG  396
                                                                    SRS-5
                    β2_1   β2_2   β1_3 αK'
Pchr        370  KGGMTVCGRFFPEGTVLSVPTYTIHRDEEVWGKDPEVFRPERWFEQDKNAV-----------------QK  422
4LXJ        385  DMHVPNTSYVIPAGYHVLVSPGYTHLRDEYF-PNAHQFNIHRWNNDSASSYSVGEEVDYGFGAISKGVSS  453
Tter        397  QGVHI-AGHYFPPGTVLSVPTYSIHHSKEIWGPDADEFRPERWDNLTARQ-----------------KN  447
                                      αL          β3_3          β4_1 β4_2      β3_2
Pchr        423  TYNPFSFGPRSCIGRNLANMELLIIVSSILRRYDFVLE--DPDKPFDTMEGFLRKPVECVVGIRRRT---  487
4LXJ        454  PYLPFGGGRHRCIGEHFAYCQLGVLMSIFIRTLKWHYPEGKTVPPPDFTSMVTLPTGPAKIIWEKRNPEQ  523
Tter        448  AFIPFSYGPRACVGRNVAEMEMKLIVATWARRYDVSLQ--QDHMDTR-EGFLRKP-LGLKIALKRRQ---  510
                                                                  SRS-6
Pchr             -------
4LXJ        524  KIGGRHH  530
Tter             -------
```

Figure 6. Structural alignment of CYP53C2 (Pchr) and CYP53A (Tter) models with CYP51 (4LXJ) using PROMALS3D [39]**.** 3D-models for CYP53C2 from *Phanerochaete chrysosporium* (Pchr) [4] and CYP53A from *Thielavia terrestris* (Tter) [24] were constructed using the template CYP51 (4LXJ) from *Saccharomyces cerevisiae* [42]. P450 characteristic notations for α-helices (shown in red font) and β-strands (shown in blue font) and SRS were mapped as per the template (4LXJ) [42] and published literature [48,49]. Residues highlighted in green and turquoise appear in contact with the heme and substrate.
doi:10.1371/journal.pone.0107209.g006

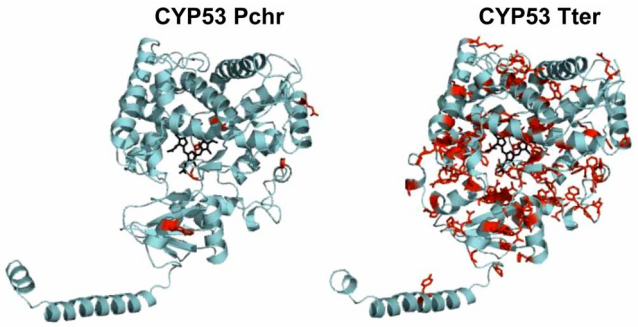

Figure 7. Analysis of amino acid conservations in CYP53 member P450s. 3D-models of CYP53C2 from *Phanerochaete chrysosporium* (Pchr) and CYP53A from *Thielavia terrestris* (Tter) as a representative of basidiomycota and ascomycota CYP53 member P450s are shown in the figure. The conserved residues among CYP53 members of two different phyla identified in this study (Fig. 2) are highlighted with red and shown in stick form. The heme-prosthetic group is presented in black. The models are presented using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/).
doi:10.1371/journal.pone.0107209.g007

their genome. The first type contains a single intron and the second one contains three introns (Fig. 4A). Paralog P450s were found in *F. graminearum* (protein IDs: 10234 and 10227) and *Aspergillus oryzae* (protein IDs: 2107 and 5958), suggesting the genome duplication of these P450s. Overall, ascomycete species CYP53 members showed simple gene structure with single and triple introns (Fig. 4A).

It is evident from Fig. 4B, especially considering the exon sizes and location of introns, that basidiomycete species enriched CYP53 members in their genome by genome duplications (paralogous evolution). The high conservation in the size of exons and location of introns of CYP53 members of basidiomycetes strongly suggests that CYP53 members are genome-duplicated. In comparison to ascomycete species, CYP53 members of basidiomycete species showed more introns in their structure (Fig. 4B). An interesting discovery we made was that basidiomycete species selectively enriched a single type of CYP53 member in their genome (Fig. 4B). In support of this argument, we present a few examples: (i) in *P. placenta* we observed two orthologs, of which one duplicated seven times while no duplication was observed for the second ortholog (protein ID: 110015); (ii) in *W. cocos* three orthologs were found: one ortholog duplicated seven times whereas no duplications were observed for the remaining two orthologs (protein IDs: 138909 and 27029); (iii) in *F. pinicola*, *P. carnosa* and *F. mediterranea* two orthologs were found in their genomes; in these species one ortholog was duplicated whereas no duplication was observed for the second ortholog (protein ID: 1025718 (*F. pinicola*); protein ID: 183190 (*P. carnosa*); protein ID: 162664 (*F. mediterranea*); (iv) *B. adusta* showed three ortholgos; two orthologs (protein IDs: 318949 and 118978) have remained the same since the divergence of this species.

From the above results it is clear that the higher number of CYP53 members in basidiomycetes is due to the genome duplication of selective CYP53 members. Despite the conservation of gene structure and the paralogous evolution of CYP53 members in basidiomycete species, the low percentage of homology among them suggests that during the genome duplication events, extensive changes in the primary structure occurred. Most of the changes might be destined to acquire novel functions to serve

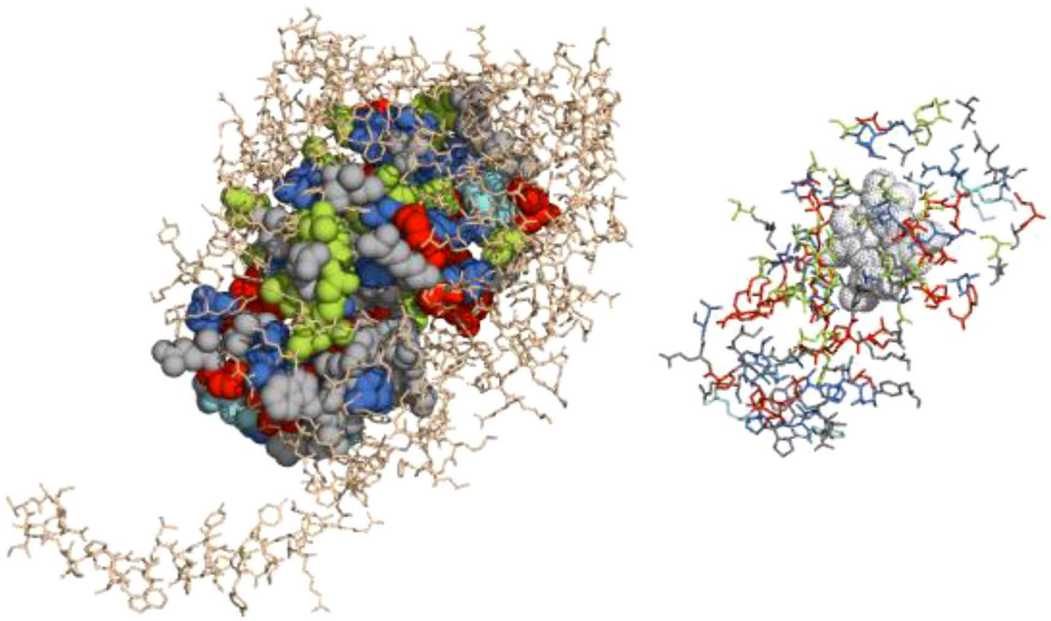

Figure 8. CYP53A Tte active site cavity mapping and analysis of the nature of the active site amino acids. The active site cavity and amino acid residues surrounding the cavity were identified using CASTp [50]. Conservation of amino acid residues in ascomycete species CYP53 P450s were identified using PROMALS3D [39] (Fig. 2). The CYP53A Tte model was used as a representative of ascomycete CYP53 members. The figure on the left shows the mapped active site cavity (space filled) and the protein backbone is presented in lines style, in the right-hand figure the conserved residues lining the cavity mapping. Different colors correspond their conservation index [40], 9 (conserved residues) – red, conservation index 7 – blue, conservation index 6 – cyan, conservation index 5 – green and residues with no conservation index – grey. The heme group is shown in grey dots. For details on the conserved nature of amino acid residues lining the active site cavity, see Table 4. The activity center is represented using PyMOL Molecular Graphics System, Version 1.7.1.1. Schrödinger, LLC (http://www.pymol.org/).
doi:10.1371/journal.pone.0107209.g008

Table 4. Analysis of amino acid conservation in CYP53A Tter active site cavity.

Conservation index	Amino acids	Number of Amino acids	Percent (%) contribution in active site cavity
9	R76, P79, G99, L101, K102, Y106, F117, R120, R122, H125, R129, F136, L192, P241, E311, T314, A318, T322, L370, S382, G387, L388, P389, R390, G409, S413, F449, P451, F452, C459, G461, R462, A465, E466, K498	35	28
7	L51, H81, N98, L116, N118, V132,V188, I189, L227, A235, T236, L243, L288, L313, G319, D321, N325, S326, V364, L378, P393, G398, V399, P408, V411, L412, S453, R457, A458, M467, E468, G494	32	26
6	Q78, R365, V392, F406, V464, M469	6	5
5	F100, D104, F105, S111, T119, F144, I148, A193, A221, A223, I224, I226, N228, E232, A310, Q315, I317, S320, T323, V374, T383, G385, V460, L496	24	19
None	N48, W49, L52, R57, Y61, V74, S103, I112, L152, A185, I219, E225, I239, L240, Q242, S371, N375, S384, I386, E391, R394, Q397, P407, T410,K470, L471, R492, F495	28	22
Total		**125**	**100**

Active site cavity residues were identified using CASTp [50] (Fig. 8). Conservation of amino acid residues in ascomycete species CYP53 P450s was identified using PROMALS3D [39] (Fig. 2). Conservation index 9 means amino acid is conserved. CYP53A Tter amino acid residues were used as a representative of ascomycete CYP53 members. CYP53A Tter P450 amino acids and their numbering are presented as representative of ascomycetes CYP53 P450s. Abbreviation Tter indicates *Thielavia terrestris*.

doi:10.1371/journal.pone.0107209.t004

fungal species (basidiomycete) to adapt to diverse ecological niches. In this direction, we further investigated whether amino acid changes play any role in CYP53 substrate specificity and/or catalytic activity.

Structure and amino acid conservation analysis of CYP53 members

Primary structure analysis and gene-structure organization studies suggested that ascomycete CYP53 members are highly conserved and basidiomycete CYP53 members are subjected to evolutionary pressure to change their primary structure composition. To identify the role of amino acids conserved as observed for ascomycete CYP53 members (Fig. 2) and variants as observed for basidiomycete CYP53 members (Figures 2 and 3) in CYP53 substrate specificity and/or catalytic activity, we performed comparative CYP53 homology modeling studies to map these conserved/variant residue locations in the protein structure. In the present study we selected two CYP53 members, CYP53A from *T. terrestris* that was recently identified and characterized in an author's laboratory [24] as representative of ascomycete CYP53 P450s and CYP53C2 from the model white rot *P. chrysosporium* (abbreviated as Pchr) [4] as representative of basidiomycete CYP53 P450s. The 3D-models of CYP53 P450s were constructed and validated as described in materials and methods. As shown in Table 3 all the parameters employed in assessing the quality of the models were favorable suggesting that models of CYP53 P450s were of good quality.

As shown in Fig. 5, CYP53A Tter and CYP53C2 Pchr showed all P450 motifs in the same way as CYP51 of *S. cerevisiae* [42]. Interestingly the membrane helix (MH) and transmembrane helix (TMH) found in CYP51 were also observed in CYP53C2 Pchr (Fig. 5). However, CYP53A Tter showed only the membrane helix (Fig. 5). This suggests that CYP53A Tter and CYP53C2 Pchr are biotopic membrane proteins with one transmembrane helix. A detailed secondary structure analysis including heme-binding residues, substrate binding residues and substrate recognition sites (SRS1-SRS6) is shown in Fig.6.

After successful construction and analysis of 3D-models of CYP53C2 Pchr and CYP53A Tter we proceeded to map the conserved amino acids observed for CYP53 members of ascomycota and basidiomycota (Fig. 2) to investigate the role of these residues in substrate specificity and/or catalytic activity. As shown in Fig. 7, CYP53 members of ascomycetes possess conserved amino acid residues throughout the protein structure, whereas CYP53 members of basidiomycetes show conservation at P450 signature motifs such as EXXR and CXG. In order to understand how many of these conserved amino acids are actually part of the active site cavity, we identified the active site cavity of CYP53 Tter using CASTp (Fig. 8 and Table 4) [50]. As shown in Fig. 8 and Table 4, among 125 amino acids lining the active site cavity, 35 (28%) are conserved (conservation index 9) and 62 amino acid residues (50%) are moderately conserved (conservation index 5–7) across the CYP53 members of ascomycetes. Overall, the high conservation of amino acids (78%) in the active site cavity and in the rest of the protein structure (Fig. 2) strongly suggests that the active site cavity and overall structure of CYP53 members of ascomycete species are highly conserved. Considering the structural conservation, any inhibitor developed against one of the CYP53 members could possibly act as common inhibitor against CYP53 members of ascomycete species and hence could act as a common anti-fungal (towards pathogenic ascomycetes) agent. On the other hand, basidiomycete CYP53 members showed much less conserved residue in their structure (Figures 2 and 7), suggesting basidiomycete members have been subjected to evolutionary pressure to acquire novel functions to help the organism adapt to diverse ecological niches.

Functional significance of CYP53 family and its potential role as a common anti-fungal drug target

CYP53 family members play a key role in fungal primary metabolism, i.e. the β-ketoadipate pathway [12,19], and secondary metabolism, i.e. detoxification of phenolic compounds [28,52]. The β-ketoadipate pathway is a convergent pathway for aromatic compound degradation [18] that is widely distributed in soil bacteria and fungi. Fungal-mediated degradation of aromatic compounds such as phenylalanine, toluene, and cinnamic acid leads to the formation of benzoate [14–16]. As part of the β-ketoadipate pathway CYP53 is involved in detoxification of this

toxic compound and key intermediate molecule. CYP53 hydroxylates benzoate to 4-hydroxybenozate [12], the prime reaction in the benzoate metabolism that subsequently leads to protocatechuate as the ring fission substrate [53]. This reaction is critical for fungal organisms in order to detoxify the benzoate; to date this hydroxylation reaction carried out by CYP53 is the only way to detoxify this compound. Further support of CYP53's critical role in fungal primary metabolism can be obtained from a study where CYP53 deletion proved to be lethal for fungal organisms' survival (19). This suggests that the CYP53 family can serve as a novel alternative drug target against fungal pathogens, especially ascomycete pathogens. Results from this work showing high conservation of the primary and tertiary structure of CYP53 members (Figures 2 and 7) across the ascomycetes (consisting of animal and plant pathogen fungal species) indicate that any inhibitor developed against a CYP53 member could serve as a novel common drug against a large number of pathogenic ascomycete fungi. Results from authors laboratory showed inhibitors directed at this P450 effectively inhibited CYP53 activity [11] and also growth inhibition of different fungal species such as *C. lunatus*, *Aspergillus niger* and *Pleurotus ostreatus* [54]. Furthermore, this P450 family offers an advantage over the CYP51 family, the currently exploited target against fungal infections, as CYP53 does not have a homolog in higher eukaryotes. This will offer researchers the opportunity to design selective and potent inhibitors of pathogenic fungi.

Overall, the facts discussed above, such as (i) the critical role of CYP53 in fungal primary metabolism, (ii) high conservation of the primary and secondary structure of CYP53 members in ascomycetes and (iii) CYP53 not having any homolog in higher eukaryotes (advantage over CYP51 family), strongly support our hypothesis that the CYP53 family can be a potential novel alternative anti-fungal drug target and an inhibitor designed against this P450 family can serve as a common drug against pathogenic ascomycetes.

The most interesting aspect of the CYP53 family's role in basidiomycete fungi extends beyond detoxification of benzoate. Our study showed that most of the ascomycetes contain a single CYP53 member in their genomes, whereas basidiomycetes showed multiple CYP53 members (Table 1). Results from this study (Fig. 4B) revealed that the number of CYP53 members increase in basidiomycete species genomes by duplication of CYP53 members after speciation (paralogous evolution). Here we propose the critical role of these CYP53 members in basidiomycetes that forced basidiomycetes to enhance this P450 family member in their genomes.

First, basidiomycetes are well known for their role as biodegraders of wood [55]. Wood is composed of many aromatic compounds, including benzoic acid derivatives and other phenolic compounds, among others eugenol, isoeugenol and guaiacol [56]. Most of these compounds are anti-fungal and toxic to fungi [13]. The multi-factorial phenomenon of toxicity of these compounds, including membrane disruption, inhibition of essential metabolic reactions, changes in pH homeostasis, and accumulation of toxic anions, has been proposed toward fungi [17]. If basidiomycete species want to colonize on wood they need an enzyme that can detoxify the benzoate molecule, as this molecule is an intermediate in detoxification of wood components comprising many aromatic compounds. Since there is an enormous need for successful wood colonization, wood-degrading basidiomycetes amplified their CYP53 members in their genomes.

Secondly, synthesis of aryl-metabolites, including veratryl alcohol by basidiomycete fungi, involves the formation of benzoate and para-hydroxybenzoic acid as intermediate molecules [14]. Veratryl alcohol is a secondary metabolite and plays a key role in lignin-peroxidase-mediated oxidation of wood components [57]. In a recent study, veratryl alcohol was shown to be the dominant extracellular ligninolytic oxidant in decaying wood [58]. The presence of a high number of CYP53 members and the generation of benzoate and para-hydroxybenzoate as an intermediate in the biosynthesis of veratryl alcohol suggest that in basidiomycete species CYP53 members also play a role in the generation of veratryl alcohol and help basidiomycete species directly in the degradation and subsequent colonization of wood.

Thirdly, demethylation of stilbene, a class of molecule found in plants, by CYP53D subfamily members from *P. placenta* (basidiomycete) [28] indicates that CYP53 family members play a critical role in the detoxification or degradation of plant compounds and help fungi in the colonization of wood. It is noteworthy that CYP53D members are present in the highest numbers (seven P450s) in *P. placenta* and are all evolved *via* paralogous evolution (Fig. 4B). This strongly indicates that *P. placenta* duplicated CYP53D members in its genome in order to colonize successfully on wood.

The above-mentioned role of CYP53 in wood-degrading basidiomycete species physiology (primary or secondary metabolism) is based on the available data and further experimentation would provide more insight into this aspect.

Collectively, the above results indicate that in ascomycetes the CYP53 role is limited to the detoxification of toxic molecules, whereas in basidiomycetes CYP53 plays an additional role, i.e. involvement in the generation of veratryl alcohol and degradation of wood-derived compounds.

Conclusion

In this advanced scientific era, understanding of animal (including human) and plant pathogenic fungal organisms in terms of controlling their causative diseases and developing effective drugs is still poorly understood. Currently available drugs and drug targets are becoming ineffective because fungal species develop resistance. Genome sequencing analysis of the fungal species gives researchers the opportunity to look for novel drug targets against these pathogens and to search for novel enzymes for the generation of human valuables. The present study is such an example; we explored fungal genome sequencing results to understand the role of a P450 family (CYP53) in serving as a common drug target against pathogenic ascomycetes and in basidiomycetes, particularly in terms of the wood-degradation process. The CYP53 family plays a key role in the detoxification of the toxic molecule benzoate and this family has proven to be essential for the organism's survival. Our findings suggest that this P450 family can serve as a common anti-fungal (toward pathogenic ascomycetes) drug target in view of its highly conserved protein structure in ascomycetes. The most striking features of ascomycete CYP53 P450s were a large number of amino acids conserved in their active site cavity (78%), strongly indicating that any inhibitor developed for this family can act against a wide range of animal and plant pathogenic ascomycetes. We also identified that CYP53 P450s can play an additional role in basidiomycetes, i.e. in the generation of the wood-degrading oxidant veratryl alcohol and degradation of wood-derived compounds. This additional role of basidiomycetes seems to have enriched this P450 family by extensive duplication of CYP53 members in their genomes (paralogous evolution). During the duplication process extensive changes in the protein primary structure occurred to enhance/acquire novel functions, such as involvement in wood degradation.

Supporting Information

Figure S1 Comparative-structural analysis and subsequent identification and estimation of conserved amino acids of CYP53 family members in fungi. Amino acid conservation was observed at three levels, i.e. (i) kingdom level (Fig. S1A), (ii) phylum level: ascomycota (Fig. S1B) and Basidiomycota (Fig. S1C) and (iii) family level (Fig. S1D). The first line in each block shows conservation indices for positions with a conservation index above 5. Each representative sequence has a magenta name and is colored according to PSIPRED [1] secondary structure predictions (red: alpha-helix, blue: beta-strand). A representative sequence and the immediate sequences below it with black names, if there are any, form a closely related group (determined by the option "Identity threshold"). Sequences within each group are aligned in a fast way. The groups are aligned using profile consistency with predicted secondary structures. The last two lines show a consensus amino acid sequence (Consensus_aa) and consensus-predicted secondary structures (Consensus_ss). Representative sequences have magenta names and they are colored according to predicted secondary structures (red: alpha-helix, blue: beta-strand). If the sequences are in aligned order, the sequences with black names directly below a representative sequence are in the same pre-aligned group and are aligned in a fast way. The first and last residue numbers of each sequence in each alignment block are shown before and after the sequences respectively. Consensus-predicted secondary structure symbols: alpha-helix: h; beta-strand: e. Consensus amino acid symbols are: conserved amino acids in bold and uppercase letters; aliphatic (I, V, L): *l*; aromatic (Y, H, W, F): @; hydrophobic (W, F, Y, M, L, I, V, A, C, T, H): *h*; alcohol (S, T): o; polar residues (D, E, H, K, N, Q, R, S, T): p; tiny (A, G, C, S): t; small (A, G, C, S, V, N, D, T, P): s; bulky residues (E, F, I, K, L, M, Q, R, W, Y): b; positively charged (K, R, H): +; negatively charged (D, E): −; charged (D, E, K, R, H): c.
(PDF)

Table S1 Analysis of homology between CYP53 members in fungi. Annotation of CYP53 P450 family members in fungi. Hit proteins were blasted at the Cytochrome P450 Webpage [26]. CYP53 members were assigned to subfamilies based on their percentage homology to reference proteins. The reference proteins were also included in the table.
(XLSX)

Acknowledgments

The authors want to thank Ms Barbara Bradley, Pretoria, South Africa for English language editing.

Author Contributions

Conceived and designed the experiments: KS NK. Performed the experiments: PJ SSM LK JS RK SBP NK KS. Analyzed the data: PJ SSM LK JS RK SBP NK KS. Contributed reagents/materials/analysis tools: PJ SSM LK JS RK SBP NK KS. Contributed to the writing of the manuscript: PJ SSM LK JS RK SBP NK KS.

References

1. Nelson DR (2013) A world of cytochrome P450s. Phil. Trans. R. Soc. B. Biol. Sci. 368: 1612 20120430.
2. Črešnar B, Petrič S (2011) Cytochrome P450 enzymes in the fungal kingdom. Biochim Biophys Acta 1814: 29–35.
3. Hlavica P (2013) Evaluation of structural features in fungal cytochromes P450 predicted to rule catalytic diversification. Biochim Biophys Acta 1834: 205–220.
4. Syed K, Yadav JS (2012) P450 monooxygenases (P450ome) of the model white rot fungus *Phanerochaete chrysosporium*. Crit Rev Microbiol 38: 339–363.
5. Yoshida Y (1988) Cytochrome P450 of fungi: primary target for azole antifungal agents. Cur Top Med Mycol 2:388–418.
6. Lepesheva GI, Waterman MR (2004) CYP51—the omnipotent P450. Mol Cell Endocrinol 215: 165–170.
7. Kelly SL, Kelly DE (2013) Microbial cytochrome P450: biodiversity and biotechnology, where do cytochrome P450 come from, what do they do and what can they do for us? Phil Trans R Soc B Biol Sci 368: 1612 20120430.
8. Sanglard D (2002) Resistance of human fungal pathogens to antifungal drugs. Curr Opin Microbiol 5: 379–385.
9. Hof H (2001) Critical annotations to the use of azole antifungals for plant protection. Antimicrob Agents Chemother 45: 2987–2990.
10. Sangamwar AT, Deshpande UD, Pekamwar SS (2008) Antifungals: need to search for a new molecular target. Indian J Pharm Sci 70: 423–430.
11. Podobnik B, Stojan J, Lah L, Krasevec N, Seliskar M, et al. (2008) CYP53A15 of *Cochliobolus lunatus*, a target for natural antifungal compounds. J Med Chem 51: 3480–3486.
12. Faber BW, van Gorcom RFM, Duine JA (2001) Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. Arch Biochem Biophys 394: 245–254.
13. Amborabe B-E, Fleurat-Lessard P, Chollet J-F, Roblin G (2002) Anti-fungal effects of salicylic acid and other benzoic acid derivatives towards *Eutypa lata*: structure-activity relationship. Plant Physiol Biochem 40: 1051–1060.
14. Lapadatescu C, Ginies C, Le Quere JL, Bonnarme P (2000) Novel scheme for biosynthesis of aryl metabolites from L-phenylalanine in the fungus *Bjerkandera adusta*. Appl Environ Microbiol 66: 1517–1522.
15. Durham DR, Mcnamee CG, Stewart DB (1984) Dissimilation of aromatic-compounds in *Thodotorula-graminis*. Biochemical-characterization of pleiotropically negative mutants. J Bacteriol 160: 771–777.
16. Jensen KA, Evans KMC, Kirk TK, Hammel KE (1994) Biosynthetic-pathway for veratryl alcohol in the ligninolytic fungus *Phanerochaete chrysosporium*. Appl Environ Microbiol 60: 709–714.
17. Brul S, Coote P (1999) Preservative agents in foods. Mode of action and microbial resistance mechanisms. Int J Food Microbiol 50: 1–17.
18. Harwood CS, Parales RE (1996) The beta-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol 50: 553–590.
19. Fraser JA, Davis MA, Hynes MJ (2002) The gene *gmdA*, encoding an amidase and *bzuA*, encoding a cytochrome P450, are required for benzamide utilization in *Aspergillus nidulans*. Fungal Genet Biol 35: 135–146.
20. Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, et al. (2012) Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. BMC Genomics 13: 444.
21. Fernandez-Fueyo E, Ruiz-Dueñas FJ, Ferreira P, Floudas D, Hibbett DS, et al. (2012) Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. Proc Natl Acad Sci USA 109: 5458–5463.
22. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science 336: 1715–1719.
23. Syed K, Nelson DR, Riley R, Yadav JS (2013) Genome-wide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the Polyporale species *Bjerkandera adusta*, *Ganoderma* sp. and *Phlebia brevispora*. Mycologia 105: 1445–1455.
24. Syed K, Shale K, Nazir KHMNZ, Krasevec N, Mashele SS, et al. (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 Genom 36: 321–333.
25. Syed K, Shale K, Pagadala NS, Tuszynski J (2014) Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 Monooxygenase families enriched in model basidiomycete fungi. PLoS ONE 9(1): e86683.
26. Nelson DR (2009) The cytochrome P450 homepage. Hum Genomics 4: 59–65.
27. Moktali V, Park J, Fedorova-Abrams ND, Park B, Choi J, et al. (2012) Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. BMC Genomics 13: 525.
28. Ide M, Ichinose H, Wariishi H (2012) Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete *Postia placenta*. Arch Microbiol 194: 243–53.
29. Syed K, Mashele SS (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): e95616.
30. Grigoriev IV, Cullen D, Goodwin SB, Hibbet D, Jeffries TW, et al. (2011) Fueling the future with fungal genomics. Mycology 2: 192–209.
31. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39: D225–9.

32. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
33. Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum evolution trees. Mol Biol Evol 9: 945–967.
34. Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. Evolving Genes and Proteins. Academic Press, New York. 97–166.
35. Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York.
36. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.
37. Chen S, Xu J, Liu C, Zhu Y, Nelson DR, et al. (2012) Genome sequence of the model medicinal mushroom Ganoderma lucidum. Nature Commun 3: 913.
38. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clutal X version 2.0. Bioinformatics 23: 2947–2948.
39. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: A tool for multiple sequence and structure alignment. Nucleic Acids Res 36: 2295–2300.
40. Pei J, Grishin NV (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics 17: 700–712.
41. Syed K, Porollo A, Lam YW, Grimmett PE, Yadav JS (2013) CYP63A2, a catalytically versatile fungal P450 monooxygenase capable of oxidizing higher-molecular-weight polycyclic aromatic hydrocarbons, alkylphenols, and alkanes. Appl Environ Microbiol 79: 2692–702.
42. Monk BC, Tomasiak TM, Keniya MV, Huschmann FU, Tyndall JD, et al. (2014) Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. Proc Natl Acad Sci USA 111: 3865–3870.
43. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815.
44. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 11:2714–2726.
45. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27: 343–350.
46. Lüthy R, Bowie JU, Eisenberg DR (1992) Assessment of protein models with three-dimensional profiles. Nature 356: 83–85.
47. Wass MN, Kelly LA, Sternberg MJ (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res 38:W469–473.
48. Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem 267: 83–90.
49. Podust LM, Stojan J, Poulos TL, Waterman MR (2001) Substrate recognition sites in 14α-sterol demethylase from comparative analysis of amino acid sequences and X-ray structure of Mycobacterium tuberculosis CYP51. J. Inorganic Biochemistry 87: 227–235.
50. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34: W116–W118.
51. Lah L, Kraševec N, Trontelj P, Komel R (2008) High diversity and complex evolution of fungal cytochrome P450 reductase: Cytochrome P450 systems. Fungal Genet Biol 45: 446–458.
52. Fujii T, Nakamura K, Shibuya K, Tanase S, Gotoh O, et al. (1997) Structural characterization of the gene and corresponding cDNA for the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. Mol Gen Genet 256: 115–120.
53. Wright JD (1993) Fungal degradation of benzoic-acid and related compounds. World J Microbiol Biotechnol 9: 9–16.
54. Korošec B, Sova M, Turk S, Kraševec N, Novak M, et al. (2014) Antifungal activity of cinnamic acid derivatives involves inhibition of benzoate 4-hydroxylase (CYP53). J Appl Microbiol 116: 955–966.
55. Martinez AT, Speranza M, Ruiz-Duenas FJ, Ferreira P, Camarero S, et al. (2005) Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. Int Microbiol 8: 195–204.
56. Haupert LJ, Owen BC, Marcum CL, Jarrell TM, Pulliam CJ (2012) Characterization of model compounds of processed lignin and the lignome by using atmospheric pressure ionization tandem mass spectrometry. Fuel 95: 634–641.
57. Ten Have R, Rietjens IM, Hartmans S, Swarts HJ, Field JA (1998) Calculated ionization potentials determine the oxidation of vanillin precursors by lignin peroxidase. FEBS Lett 430: 390–392.
58. Hunt CG, Houtman CJ, Jones DC, Kitin P, Korripally P, et al. (2013) Spatial mapping of extracellular oxidant production by a white rot basidiomycete on wood reveals details of ligninolytic mechanism. Environ Microbiol 15: 956–966.

Central University of Technology, Free State

FACULTY OF HEALTH AND ENVIRONMENTAL SCIENCES

Memorandum

To: **Prof SS Mashele**
Dean: Faculty of Health and Environmental Sciences

cc:

From: **Dr K Syed – Supervisor**

Date: 14 August 2015

Priority: High

Deadline:

Our Ref: **Your Ref:**

Subject: **Ms P Jawallapersand : M Tech Biomedical Technology**

Dear Prof Mashele

I hereby declare that the dissertation for **Ms P Jawallapersand** is ready for submission. Considering the advantage of having soft copy over hard copy to save money and paper I hereby requested to make four copies. I hereby confirm that external examiners and internal examiners are happy with soft copies.

Furthermore, I request examination department not to display **Ms P Jawallapersand** thesis on CUT website considering the chapter 4 has not been published. Once the chapter 4 published I will inform CUT to include her thesis on the CUT website.

Yours truly,

S. Khajamoheddin

Dr K Syed

14 August 2015