Phylogenetic and structural and functional analysis of cytochrome P450 monooxygenase CYP5619A1 from *Saprolegnia diclina*

By

Hans-Denis Bamal

Dissertation submitted in fulfilment of the requirements of the degree

Doctor of Health Science in Biomedical Technology

Department of Health Sciences

Central University of Technology, Free State

13 November 2017

Supervisor: Prof Khajamohiddin Syed

Co-supervisor: Prof Samson Sitheni Mashele

DECLARATION

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

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

HANS DENIS BAMAL **DATE**

M.B. BRADLEY

P.O. Box 37326
Faerie Glen
Pretoria 0043

072 369 5149

DECLARATION ON EDITING

Customer:	**Date**
Prof K Syed Department of Health Sciences Faculty of Health and Environmental Sciences Central University of Technology	2016/11/06
Contact: Prof Syed	

Thesis submitted for editing

Titles of the thesis as indicated:

Phylogenetic and structural and functional analysis of cytochrome P450 monooxygenase CYP5619A1 from *Saprolegnia diclina*

The above entitled thesis was submitted to me for language editing, which was done on 23 October 2017.

M.B. BRADLEY (MA) - Language editor

[redacted]

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

- Professor Jack Tuszyinski of the Cross Cancer Institute, for the IT support from his laboratory at the University of Alberta, Edmonton, AB, Canada.

- Professor Samson Sitheni Mashele for his mentorship, guidance and support.

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

- Dr. Emmanuel I Edoun from the University of the Witwatersrand, for making this possible from the very beginning of my student life.

- My colleagues at the Faculty of Health and Environmental Sciences, especially the Department of Health Sciences at the Central University of the Free State for their continuous support, assistance and encouragement.

- All members of the Tuszynski lab, Faculty of Medicine and Dental Sciences, University of Alberta, Edmonton, AB, Canada.

- To my parents, sisters and brothers 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 my dearest friends and family. 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.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

%	Percentage / Percent
>	Over
≥	Exact value or more than
°C	Degree Celsius
µm	Micrometer
µM	MicroMolar
3D	Three-dimensional
Å	Angstrom
A. astaci	*Aphanomyces astaci*
ACD	Arachidonic acid
AI	Adequate intake
Amp	Ampicillin
Arg	Arginine
A-T-G-C	Adenine-Thymine-Guanine-Cytosine
B.C.	Before Christ
BAC	Bacterial Artificial Chromosome
BLASTp	Protein-protein Basic Local Alignment Search Tool
bp	Base pairs
$CaCl_2$	Calcium chloride

CDD	Conserved Domain Database
CHO	Chinese hamster ovary
CO	Carbon monoxide
CPR	Cytochrome P450 Reductase
C-terminal	Carboxy-terminus
CYP	Cytochrome P450
Cys	Cysteine
DCR	Icosanoic acid
DHA	docosahexaenoic acid
E. coli	*Escherichia coli*
EFA	Essential fatty acid
EIC	Linoleic acid
EPA	Eicosapentaenoic acid
ERRAT	Server detecting errors in protein models
EST	Expressed Sequence Tag
et al.	*Et alia* (and others)
FAD	Flavin adenine dinucleotide
FAO	Food and Agriculture Organization
Fe^{2+}	Iron (II) cation
FeS	Iron sulfur
FMN	Flavin mononucleotide

g	Grams
g/d	Grams per day
g/mol	Grams/mole
GLA	Gamma-linolenic acid
Glu	Glutamate
Gly	Glycine
HB	Hydrogen bond
HCl	Hydrogen chloride
HEK	Human embryonic kidney
HEM	Heme group
HEPES	HydroxyEthyl PiperazineEthaneSulfonic acid
Hid	Histidine neutral delta-protonated
Hie	Histidine neutral epsilon-protonated
i.e.	*id est* (that is)
ID	Identity
IDT	Integrated DNA Technology
Inc.	Incorporation
KB	Kilo base
kcal/mol	Kilocalories per mole
KCl	Potassium chloride
KOH	Potassium hydroxide

l	Litre
LB	Luria-Bertani
Leu	Leucine
LNL	Alpha-linolenic acid
M	Molar
MCS	Multiple cloning site
MEGA	Molecular Evolutionary Genetics Analysis
Met	Methionine
mg/d	Milligrams per day
$MgCl_2$	Magnesium chloride
MGR	Malachite green
$MgSO_4$	Magnesium sulfate
min	Minutes
ml	Millilitres
mM	MilliMolar
$MnCl_2$	Manganese (II) chloride
MOE	Molecular Operating Environment
MYR	Myristic acid
MYZ	Myristoleic acid
NaCl	Sodium chloride
NADH	Reduced nicotinamide adenine dinucleotide

NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NIH	National Institute of Health
nm	Nanometre
NMR	Nuclear magnetic resonance
No.	Number
NSF	New subfamily
N-terminal	Amino terminal end
OD600	Optical Density measured at a wavelength of 600 nanometres
OLE	Oleic acid
O-linked	Oxygen-linked
P. ramorum	*Phytophthora ramorum*
P. sojae	*Phytophthora sojae*
P450	Cytochrome P450
PAM	Palmitoleic acid
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pH	Potential of hydrogen
PLM	Palmitic acid
Pro	Proline

PromalS3D	PROfile Multiple Alignment with predicted Local Structures and 3D constraints
PUFA	Polyunsaturated fatty acid
RE	Restriction enzyme
RMSD	Root Mean Square Deviation
rpm	revolutions per minute
S. diclina	*Saprolegnia diclina*
S. ferax	*Saprolegnia ferax*
S. parasitica	*Saprolegnia parasitica*
SOB	Super Optimised Broth
SOC	Super Optimal broth with Catabolite repression
STE	Stearic acid
TB (buffer)	Tris base, Boric acid
Thr	Threonine
T_m	Melting temperature
™	Trademark
Trp	Tryptophan
Tyr	Tyrosine
US$	United States Dollars
USA	United States of America
v/v	Volume per volume
WHO	World Health Organization

www	World wide web
xg	Relative centrifugal force measured in multiples of the standard acceleration due to gravity at the Earth's surface
Z-score	Indicates overall model quality
α	Alpha
β	Beta

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Genome sequencing of lower eukaryotes such as fungi revealed high diversity of cytochrome P450 monooxygenases (P450s/CYPs) in their genomes compared to other biological kingdoms. For example, not only the presence of a large number of P450s was detected in many of their genomes, but also high diversity in terms of the number of P450 families. P450s are heme-thiolate proteins distributed across the biological kingdoms with immense catalytic diversity, which has prompted the use of these enzymes as potential catalysts for the production of fine chemicals, pharmaceutical compounds, antibiotics, fragrances and detoxification of carcinogenic and/or mutagenic compounds. Progress has been made in understanding P450s from lower eukaryotic organisms, which has led to the unravelling of their potential as anti-fungal drug targets.

The lower eukaryotes belonging to the kingdom *Stramenopila*, especially phylum *Oomycota* species P450s, have been underexplored. Oomycetes are "hard-wired parasites" that remain a serious problem in agriculture and aquaculture and are counted among the most widespread and deadliest disease-causing agents of plants and crops worldwide. Their destructive behaviour lies in their ability to breach the host surface and break it down, promptly resulting in extensive destruction that hinders agricultural growth. The impact of oomycete species on the economy triggered various investigations on pathogenesis and control methods for these pathogens.

In the quest to find a remedy, genome sequencing of oomycetes was carried out. Recently, the Unit for Drug Discovery Research's laboratory (Department of Health Sciences, Faculty of Health and Environmental Sciences at the Central University of Technology) performed comprehensive comparative P450 genomics in 13 oomycete pathogens and discovered six novel P450s that can be used as drug targets against these

pathogens, particularly fish pathogens. The novel P450s belong to the CYP5619 family and were found in the fish pathogen *Saprolegnia diclina*. In order to use these P450s as novel drug targets, it is of the utmost importance to perform biochemical and biophysical characterisation of the family members. Hence, I am herewith proposing to perform *in silico* structural and functional analysis of CYP5619A1 from *S. diclina,* including cloning and generation of recombinant *E. coli* cells containing the *CYP5619A1* gene in a novel expression vector and comprehensive phylogenetic analysis. This study is the first of its kind on analysis of the novel P450 family CYP5619 in microbes.

Phylogenetic analysis of CYP5619 family members across biological kingdoms revealed the presence of this novel P450 family in other oomycetes and in a phytoplankton. However, the number of CYP5169 members in organisms varied. Nine CYP5619 members were found in *Achlya hypogyna* and six were found in *S. parasitica* (both oomycetes). The oomycetes, *Thraustotheca clavata* and *Aphanomyces invadans,* were found to have three and two CYP5619 members, respectively. *Emiliania huxleyi,* a phytoplankton, was found to have two CYP5619 family members, but the smallest count was attributed to an oomycete, *A. astaci* (one CYP5619 member). This suggests that the CYP5619 family is present in other organisms apart for oomycetes.

After performing phylogenetic analysis, a 3D model of CYP5619A1 from *S. diclina* was built by homology modeling and assessed for its binding affinity with different predicted substrates and with malachite green, a remedy used to treat *S. diclina* infections. The study revealed that eight of the compounds required low energy to bind to the target protein, with binding energies below -6.00 kcal/mol. This suggests that these ligands can act as possible substrates of CYP5619A1. Among all ligands, linoleic acid and malachite green showed a high binding affinity with the CYP5619A1 model. Linoleic acid is a polyunsaturated fatty

acid with 18 carbon atoms and two double bonds in its structure. Malachite green is an organic compound that is widely used in aquaculture to treat *S. diclina* infections. These two compounds appeared to be the compounds with the best affinities to the target protein. In this regard, it is reasonable to believe that linoleic acid-like compounds could be potential substrates for CYP5619A1 and malachite green possibly inhibiting CYP5619A1 in *S. diclina*.

In order to validate *in silico* results, *CYP5619A1* was cloned into the newly designed vector *pINK-d* using *in silico* and *in vitro* techniques. The gene was cloned into the novel expression vector using the software pDRAW and the sequence of the designed primers, vector and gene of interest were sent to GenScript. The vector and vector with *CYP5619A1* obtained from GenScript were subjected to restriction enzyme analysis. The results obtained were satisfactory, as the gene was perfectly cloned into the vector, which was verified by running the plasmid through an agarose gel on the one hand, and the plasmid restricted by the *KpnI* and *XbaI* enzymes on the other hand. The gel analysis confirmed the presence of *CYP5619A1*.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1. Essential fatty acids

The human body synthesises fats it needs from the food we consume. Nevertheless, some important fats have to be taken in directly from natural sources, since they cannot be synthesised by the body, and are therefore termed "essential". These essential fatty acids (EFAs) are also called polyunsaturated fatty acids (PUFAs), as they contain more than one double bond between carbon atoms and therefore, are not saturated with hydrogen. There are two families of PUFAs, namely the alpha-linolenic acid (LNL), and the linoleic acid (EIC) families. They are used to build specialised fats: the omega-3 and the omega-6 fatty acids, respectively (Groff *et al.*, 1995).

Once ingested, the LNL is converted to two long-chain PUFAs: eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), the two types of omega-3 fatty acids most readily used by the body. On the other hand, EIC is converted by the body upon ingestion, to the long-chain PUFAs gamma-linolenic acid (GLA) and arachidonic acid (www.nutri-facts.org).

1.1.1. Role of essential fatty acids in humans

It is necessary to have an adequate intake of PUFAs, as they play a crucial role in the proper development and functioning of organs. Indeed, they account for one of the most important requirements for the development of the brain and the maintenance of its function, as well as for good vision (www.nutri-facts.org). Almost two thirds of the human brain is made up of phospholipids, which makes it the fattiest organ of the human body. For good development and integrity, it requires PUFAs such as DHA and EPA, as well as other micronutrients such as vitamins B, C and E, iron, zinc, copper or taurine. DHA is widely distributed in the

cerebral cortex, membranes of synaptic communication centres, mitochondria and the retina's photoreceptors. It also accounts for 50% of neuronal membrane weight (Haag, 2003; Singh, 2003).

Besides the role played by the PUFAs in brain development and functioning and the vision process, EFAs also help in the regulation of immune and inflammatory responses, as well as in hormonal regulation by producing hormone-like molecules. According to Hippocrates, "*the food that is good for the heart is likely to be good for the brain*". Omega-3 fatty acids DHA and EPA contribute to the regulation of blood pressure and the maintenance of triglycerides' concentrations. Singh (2005) also reported some benefits of omega-3, including a reduction of thromboxane level and an increase of prostacyclin level (which lead to vasodilation and greater oxygen delivery to the organs), an increase of serotonin ("feel good" neurotransmitter) and acetylcholine ("memory boosting" chemical) levels, and the neutralisation of oxygen-free radicals.

Deficiencies in EFAs are found in patients fed intravenously, those with cystic fibrosis and those with chronic poor fat absorption. It is also found in infants who did not receive enough nutrients from their mothers during their pregnancies. Symptoms include vision and nerve problems, probably developed during the growth of the child, poor memory, heart and circulation issues, fatigue, mood swings, dry skin (www.nutri-facts.org), liver and kidney abnormalities, reduced growth rates and impaired immune functions (Linscheer & Vergroesen, 1994; Barnard & Raymond, 1999; Harvard Mental Health Letter, 2003).

1.1.2. Daily requirements for humans

Recently, the National Academies released the Dietary Reference Intakes Report for Energy and Macronutrients (Trumbo *et al.*, 2002). Adequate intakes (AIs) have been set for EIC and LNL. The AI for EIC is 17 g/d and 12 g/d for men and women aged 19 to 50 years old,

respectively. The AI for LNL is 1.6 g/d and 1.1 g/d for men and women aged 19 to >70 years old, respectively.

In 1999, the National Institutes of Health (NIH) sponsored an international workshop on the essentiality and recommended dietary intakes for omega-6 and omega-3 fatty acids. The NIH working group proposed AIs of 2–3% of total calories for EIC, 1% of total calories for LNL, and 0.3% of total calories for EPA and DHA. The working group further recommended intakes of EPA and DHA of ≥650 mg/d and a minimum of 300 mg/d of DHA during pregnancy and lactation (Simopoulos *et al.*, 1999).

Health Canada suggests a minimum of 3% of energy from omega-6 fatty acids and 0.5% from omega-3 fatty acids or 1% for infants who do not receive a preformed source of EPA and DHA (Scientific Review Committee, 1990). The United Kingdom recommends that 1% of energy be from LNL and 0.5% from EPA and DHA combined (Hmso, 1994).

While there are no official recommendations for vegetarians and vegans, it is not possible for this population to achieve the NIH working group's proposed AIs for EPA and DHA. Even with the use of DHA-enriched eggs, some seaweed, and/or DHA supplements, the best vegetarians could do is to meet the recommended intakes for DHA. Some experts suggest that vegetarians (and others receiving no direct sources of EPA and DHA) at least double the recommended intakes of LNL (Davis & Kris-Etherton, 2003). This would suggest an intake of LNL in the range of 1–2%. The ratio of omega-6 to omega-3 fatty acids is often used to assess the balance between EFA in the diet, although there is some controversy as to its practical significance. For vegetarians and others who consume little, if any, EPA and DHA, the omega-6-to-omega-3 ratio is of greater relevance than for individuals who consume significant daily sources of EPA and DHA.

A number of recommendations have been made on the basis of the ratio of omega-6 to omega-3 fatty acids. The World Health Organisation and the Food and Agriculture Organisation (WHO and FAO) suggest a ratio of 5:1–10:1, Sweden recommends a ratio of 5:1 (Nordic Working Group on Diet and Nutrition) (Becker *et al.*, 2004), Canada recommends 4:1–10:1 (Scientific Review Committee), and Japan recently changed its recommendation from 4:1 to 2:1 (Kris-Etherton *et al.*, 2000). On the basis of the proposed AIs, the NIH suggests a ratio of 2:1–3:1 (Simopoulos *et al.*, 1999). One study found that a ratio of 4:1 allows for adequate conversion to DHA in healthy vegetarians (Indu & Ghafoorunissa, 1992). Another research group suggested that the optimal ratio to maximise the conversion of LNL to DHA is 2.3:1 (Masters, 1996). Given the rate of conversion of LNL to EPA and DHA, it has been suggested that a safe and adequate ratio for the vegetarian and vegan populations would be in the range of 2:1 to 4:1 (Davis & Melina, 2000). This can best be achieved by increasing LNL in the diet and decreasing EIC, if indicated.

1.1.3. Sources of essential fatty acids

Supplements containing omega-3 and omega-6 fatty acids are available, but the EFAs can also be found naturally in plant and animal products. Table 1.1 shows some common forms and different sources of those fatty acids, as well as some potential health-promoting benefits. From the information presented in table 1.1 it is clear that fish and fish products are the common source for EPA and DHA and are readily available for consumption. Thus, it is necessary to protect aquatic animals, especially fish, to ensure a continuous supply of EFAs to humans.

Table 1.1. Common forms, food sources and health benefits related to omega-3 and omega-6 fatty acids (Franzen-Castle & Ritter-Gooder, 2010)**.**

	Omega-3 fatty acids	Omega-6 fatty acids
Most commons forms	Eicosapentaenoic acid (EPA) Docosahexanoic acid (DHA), and Alpha-linolenic acid (LNL)	Linoleic acid (EIC) mostly (85-90% of omega-6 fatty acids)
Common food sources	EPA and DHA – fatty fish (salmon, white tuna, mackerel, rainbow trout, herring, halibut and sardines) LNL – canola or soybean oil, walnuts and ground flaxseed or flaxseed oil	Vegetable oils (corn, sunflower, safflower and soy), salad dressing, nuts, whole wheat bread and chicken
Research suggests potential health-promoting benefits	Reduce inflammation in heart disease, inflammatory bowel disease, and rheumatoid arthritis Help prevent blood from clotting and sticking to artery walls Help lower risk of blocked blood vessels and heart attacks Prevent hardening of the arteries Decrease risk of sudden death and abnormal heart rates Decrease triglyceride levels Lower blood pressure	Neutral or lower levels of inflammatory markers Replacing saturated and transfat with omega-3 fatty acids associated with decreasing risk of heart disease Improve insulin resistance and reduce the incidence of diabetes Lower blood pressure Lower cholesterol levels

1.2. Oomycetes

Long considered a class within the kingdom *Fungi* based on similarities in growth patterns, oomycetes have been subject to multiple studies. Traditional groupings based on morphology have been replaced by more advanced methods such as (i) the observation of structural characteristics through the transmission electron microscope developed in the 1970s, (ii) taxonomic analyses of phenotypic characteristics and (iii) sequence comparisons.

1.2.1 Classification of oomycetes

Oomycetes form a class under the superphylum *Heterokonta*, and the kingdom *Stramenopila*. Dick (2013) reported two subclasses, with six prominent orders. Below are some of the genera identified in the various orders:

Subclass: *Peronosporomycetidae*

Order: Peronosporales, e.g: Many genera of downy mildews, *Albugo*

Order: Pythiales, e.g.: *Pythium*, *Phytophthora*

Subclass: *Saprolegniomycetidae*

Order: *Saprolegniales*, e.g.: Aphanomyces

Order: *Sclerosporales*, e.g.: Downy mildews of the *Poaceae*, such as: *Sclerospora*, *Peronosclerospora*, *Sclerophthora*

Order: *Salilagenidiales*, e.g. Lagena

Order: *Leptomitales*

Saprolegnia is the only genus of oomycete pathogens that does not contain plant pathogens, but it contains pathogens of different water-borne organisms such as crayfish and fish. Although *Saprolegnia* species are considered secondary pathogens, when given the

appropriate circumstances they would act as primary pathogens and cause mycoses. Typically, once an organism is infected the disease is fatal. Scientists believe that extensive mortality of salmon and trout in Europe have been caused by *Saprolegnia* infection. *Saprolegnia* can parasitise fins and flesh, gaining initial infection through wounds. It can also parasitise eggs and is often visible as a white cottony mass on the surface of eggs or fish in home aquaria.

Recently, *Saprolegnia ferax* has been linked to the decline in amphibian populations. Apparently, climate change induced shallower water levels, which exposed eggs to higher levels of ultraviolet radiation. UV-B radiations have been reported to cause damage to the skin and have a strong immunosuppressive effect in fish. In the case of *S. ferax* infections, in addition to the above-mentioned effects, the number of mucus-producing goblet cells was observed to be reduced in UV-B-injured fish skin (Helbling *et al.*, 2003).

1.2.2. Characteristics of oomycetes

Absorption of nutrients, production of mycelium, filamentous growth at the vegetative stage and formation of spores for sexual and asexual reproduction led researchers to consider oomycetes as fungi. However, recent studies revealed new characteristics of those pathogens that classify them as a distinct group, namely the heterokont algae.

These heterokont algae form a distinctive group from the general algae in having motile zoospores with two types of flagella, to which the term “heterokont” refers. One of the flagella on the oospore is a whiplash oriented posteriorly and the other one is oriented anteriorly, with a fibrous, ciliated structure.

Another characteristic of the oomycetes that distinguishes them from fungi is their sexual reproduction. In oomycetes, haploid gametes fertilise haploid oospheres, producing

diploid oospores as zygotes. This is termed oogamous reproduction. One may observe a large, single oospore, or rather a cluster of small oospores, inside the oogonium.

A third pattern that distinguishes oomycetes from true fungi is the composition of their cell walls. Fungi have cell walls made of chitin and cellulose. Nonetheless, this second pattern is rarely observable in the fungi. In the oomycetes, on the other hand, cellulose is an integral part of the cell wall, together with beta glucans, which are potent and proven non-specific immune response activators and modulators (Beta Glucan Research – Saccharomyces cerevisiae, from http://www.betaglucan.org/).

Other particular characteristics of oomycetes are the nuclear state of the vegetative mycelium and the mitochondrial cristae's shape. Indeed, the true fungi's mycelium is septate and contains haploid nuclei; cells may as well be dikaryotic with two or more haploid nuclei. In the *Oomycota*, however, the organism is primarily in a diploid state: the vegetative cells generally consist of coenocytic (without septa) hyphae that contain diploid nuclei. On the other hand, fungi possess flattened mitochondrial cristae, while *Oomycota*'s are tubular, with protoplasmic- and nuclear-associated microtubules. This last characteristic was observed by dint of the development of the transmission electron microscope, allowing researchers to perceive differences in ultrastructure characteristics. More studies using molecular sequence data and mathematical algorithms came to support the established difference between *Oomycota* and fungi, by comparing sequence similarities in particular gene regions and analysing changes among the gene regions' base pairs, leading to the conclusion that *Oomycota* mostly relate to the other members of the heterokont algae, rather than to fungi.

1.2.3. Oomycete species and targets

Oomycetes are organisms considered to be "hard-wired" for parasitism (Lamour & Kamoun, 2009; Beakes, 2012) although some species from the order Peronosporales are mainly

saprophytes (Kamoun, 2003; Lamour & Kamoun, 2009). They cause diseases in both plants and animals (Kamoun, 2003; Phillips *et al.*, 2008; Lamour & Kamoun, 2009). Oomycetes are counted among the most widespread and deadliest disease-causing agents of plants and crops worldwide. Their destructive behaviour in agriculture lies in their ability to breach the host plant's surface and break it down, promptly resulting in extensive destruction that hinders agricultural growth (Soanes *et al.*, 2007). There has been a huge impairment of aquaculture owing to oomycetes and as for plants, serious diseases are caused not only in agriculturally and ornamentally important plants, but also other plants in the environment. A summary of diseases caused by oomycetes is listed in Table 1.2.

Table 1.2. Oomycete species and general information, such as their host and diseases caused by these pathogenic species.

Species name	Host	General information
Phytophthora sojae and *Phytophthora ramorum*	Plants	These species are considered model species for the *Phytophthora* genus owing to well-developed genetic and genomics resources, including genetic maps, bacterial artificial chromosome (BAC) libraries and expressed sequence tag (EST) libraries. *P. sojae* causes soybean root and stem rot and leads to substantial yield losses annually. *P. ramorum* causes sudden oak death, tanoak and ramorum blight on woody ornamental forest under canopy plants. It causes stem cankers on trees and leaf blight or stem dieback on ornamentals and under-storey forest species.
Phytophthora infestans	Plants	It is the causative agent of late blight disease in potato and tomato plants. Responsible for the famous Irish potato famine in the mid-nineteenth century.
Phytophthora parasitica	Plants	A model species of oomycete pathogens. Causes destructive diseases in a wide variety of crops including tomato, eggplant, pepper, tobacco, potato, walnuts, fruits and a wide range of nursery and ornamental plants and forest ecosystems.

Phytophthora capsici	Plants	Attacks the roots, stems, leaves and fruit of pepper, resulting in damping-off, seedling blight, foliar blight and death. Other plants that are affected include tomato, eggplant, cucumber, watermelon, pumpkin, squash and cocoa.
Hyaloperonospora arabidopsidis (formerly *Hyaloperonospora parasitica*)	Plant	The causal agent of the downy mildew of the plant model organism *Arabidopsis thaliana*
Pythium aphanidermatum	Plants	Causes damping off, root and stem rots, and blights of grasses and fruits, papaya, beets, pepper and cotton
Pythium irregular	Plants	Highly pathogenic on wide range of cereal and leguminous plants
Pythium awayamai	Plants	Isolated mainly from monocotyledon plants. Causes snow-rot disease of cereal plants including wheat and barley
Pythium ultimum	Plants	Causes damping-off and root rot to diverse plant hosts, including crops and forests
Pythium vexan	Plants	Causes canker, damping-off and rot disease in many economically important crops, including rubber trees, potato and sugar cane
Saprolegnia parasitica	Animals	One of the most important fish pathogens causing millions of dollars of losses to the aquatic culture business worldwide. It attacks a wide variety of fish,

		amphibians and crustaceans. Members of this genus *Saprolegnia* causes "Saprolegniosis", a disease that is characterised by visible white or grey patches of filamentous mycelium on the body and fins of freshwater fish.
Saprolegnia diclina	Animals	Pathogen of amphibians, fish and insects. Plays a role in decline of natural populations of amphibians. Outbreaks leading to severe reductions and even extinction of amphibians have been attributed to this species. It is also a large problem in fish hatcheries where it infects eggs of salmon and trout.

As shown in the table above, oomycetes also infect animals, mainly aquatic and to date, they have remained a serious problem in agriculture and aquaculture (Kamoun, 2003; Lamour & Kamoun, 2009; Phillips *et al.*, 2008). Oomycete diseases are not commonly easy to control. Moreover, some oomycete species, particularly *Phytophthora* species, have the ability to build up resistance against chemicals by producing new genetically tougher strains. Plants are also very sensitive to oomycete attacks owing to their weak disease resistance.

1.2.4. Economic importance of oomycetes

Oomycetes are eukaryotic organisms that superficially resemble filamentous fungi, but are phylogenetically related to diatoms and brown algae in the kingdom *Stramenopila* (Gunderson *et al.*, 1987; Lamour & Kamoun, 2009; Thines & Kamoun, 2010; Jiang & Tyler, 2012; Thines, 2014). The impact of oomycetes on humankind is well documented as both a persistent threat to subsistence and commercial farming and as destructive pathogens of native plants (Erwin, 1996; Agrios, 2005; Lamour & Kamoun, 2009). As a result, news related to plant diseases caused by oomycetes tends to capture the interest of the general public and is frequently featured in the media.

S. ferax and *S. diclina* are causing declines in amphibian populations (Kiesecker *et al.*, 2001; Fernandez Benéitez *et al.*, 2008). *S. parasitica* is known to be a devastating pathogen affecting many freshwater species of fish, and *S. diclina* is a potent pathogen of fish eggs. Both are currently causing significant economic damage in the global fish farming industry (van West 2006; Phillips *et al.*, 2008). Previously, *S. parasitica* and *S. diclina* infections were controlled with the biocide malachite green. However, there was a worldwide ban on the use of this chemical because of its potential toxicological and carcinogenic effects on both fish and consumers of fish (reviewed in Alderman, 1994; Marking *et al.*, 1994). As a result of this ban, there has been a significant increase in *Saprolegnia* infections (Robertson *et al.*, 2009).

On fish, saprolegniosis is characterised by white or gray patches of cotton wool-like filamentous mycelia (Hatai & Hoshiai, 1992). In general, infection initially appears on epidermal tissues of the head, tail and fins of the fish (Tiffney, 1939; Hatai & Hoshiai, 1992; Fregeneda Grandes *et al.*, 2001; Hussein & Hatai, 2002), and it subsequently spreads to the rest of the body. Lesion areas may be soft, necrotic and ulcerated, and the surrounding areas may show oedema and necrosis (Gieseker *et al.*, 2006). Saprolegniosis has been documented in a range of fish species, including Atlantic salmon (*Salmo salar* L.), brown trout, (*Salmo trutta* L.), coho salmon (*Oncorhynchus kisutch*), perch (*Perca fluviatilis*), masu salmon (*Oncorhynchus masou*), rainbow trout (*Oncorhynchus mykiss*), Japanese char (*Salvelinus leucomenis*), sockeye salmon (*Oncorhynchus nerka*), channel catfish (*Ictalurus punctatus*) (Tiffney, 1939; Hatai & Hoshiai, 1992; Bly *et al.*, 1993; Hussein & Hatai, 2002; Stueland *et al.*, 2005) and freshwater crayfish (Diéguez-Uribeondo *et al.*, 1994), as well as in amphibians (Kieseker *et al.*, 2001; Fernandez-Benéitez *et al.*, 2008).

Because of overfishing in the seas in recent years, fish production has become dependent on fish farming for an adequate supply and as a result, aquaculture has become the world's fastest growing food sector. This fact is reflected in the astonishing increase in fish production through fish farming. In 2004, a staggering 45.5 million tons of fish worth US$ 63.4 billion were farmed worldwide. This represents a 12.6% and 15.3% increase, respectively, over reported figures in 2002 (FAO Fishery Information: www.fao.org). Aquaculture has both social and economic importance in many countries, which include China, Canada, Chile, Japan, Norway, the United States and the United Kingdom. The greatest cause of economic loss in aquaculture is diseased fish, with bacterial diseases being the most common, closely followed by fungal and oomycete infections (Meyer, 1991). *S. parasitica*, a ubiquitous freshwater pathogen (Diéguez-Uribeondo *et al.*, 2007), can infect not only a wide range of wild and farmed fish, but may also cause infection in hobby fish tanks.

In Japanese freshwater ponds, *Saprolegnia* species are among the main causes of infection of salmon fish (Hussein & Hatai, 2002). *S. parasitica* has been reported to cause mass mortality of coho salmon in Japanese salmon farms (Hatai & Hoshiai, 1992). *S. parasitica* is also the cause of the disease termed ‘‘winter kill,’’ which affects catfish in the United States and causes significant financial losses in the farming of this fish (Bly *et al.*, 1992). In Chile, one of the main fish-producing countries, *S. parasitica* has become a major threat to farms of Atlantic salmon, coho salmon, and rainbow trout (Zaror *et al.*, 2004). In addition to aquaculture, *S. parasitica* has been reported to have a major impact on wild fish populations globally. *Saprolegnia* species have been isolated from head burn lesions of wild adult Chinook and steelhead salmon. The disease is also having a significant impact on the populations of these species in the Northwest United States, where up to 22% of returning salmon die as a result of headburn lesions that have become infected with *S. parasitica* (Neitzel *et al.*, 2004).

1.3. Cytochrome P450 monooxygenases

1.3.1. Generalities on cytochrome P450 monooxygenases

Cytochrome P450 monooxygenases (also called CYPs or P450s) are heme-thiolate proteins that form a very diverse protein superfamily and play a role in metabolising drugs and numerous other xenobiotics (Nelson, 2013). They are found in almost every living species, with extremely few exceptions, such as the gram-negative bacteria *Escherichia coli.* Nonetheless, on a greater scale, they have been documented to belong to every phylogenetic domain of life, ranging from microscopic prokaryotes to complex eukaryotes (Nelson, 2013). The highest number of P450 genes belongs to the kingdom *Plantae*. Nevertheless, the kingdom *Fungi* possesses the highest number of P450 families, which denotes higher diversity in that kingdom. As a matter of fact, 399 P450 families are located throughout the 2 784 annotated fungal P450s in comparison to plant genomes having only 129 P450 families

located throughout the 4267 annotated plant P450s (Nelson, 2009; Nelson, 2011). To date, more than 21 000 P450s have been sequenced, annotated and described (Nelson, 2009). Through evolution, they have enabled organisms to adapt to the different changes that occurred, geographically and ecologically. Examples are documented and include adaptation to elevated hydrostatic pressures and temperatures of solfataric hot springs (Park *et al.*, 2002; David *et al.*, 2013; Nelson, 2013) and adaption in the utilisation of photosynthetically fixed carbon such as wood (Syed *et al.*, 2014a, 2014b).

The term 'cytochrome P450' was coined in 1962 as a temporary name for a coloured substance in the cell (Omura & Sato, 1962). This pigment (abbreviated 'P' in the naming), when reduced and bound with carbon monoxide, produced an unusual absorption peak at a wavelength of 450 nm (Figure 1.1), thus P (for pigment) and 450 (for the wavelength). The cysteine-thiolate group is the prime reason for this phenomenal spectral display observed in P450s. 'Cytochrome' is a misnomer, given that P450s are enzymes rather than true cytochromes or single entities. On the other hand, there is an observed formation of the fifth ligand of the heme-iron, hence the name heme-thiolate proteins or haemoproteins is given to P450 enzymes (Omura & Sato, 1962). Despite this, the name 'cytochrome P450' has stuck and is so widely accepted that any change would be impractical. At first, P450s were believed to represent single enzymes. Today, it seems likely that humans and other mammals have approximately 50 distinct P450 enzymes. The total number may be higher in plants, possibly as high as several hundred. In the last 15 years of the 20th century, research was largely concerned with defining P450 multiplicity in humans and a diverse range of other organisms.

Several diverse P450 nomenclature systems have emerged based on their molecular weights or preferences for substrates. The resultant plethora of names and accompanying confusion prompted prominent workers in the field to devise a standard nomenclature for the P450 gene family based on amino acid sequence comparisons and the evolutionary

relationships of the corresponding genes (Nebert *et al.*, 1987). First proposed in 1987, the nomenclature was devised on the premise that it would be updated as frequently as the identification of new P450 enzymes necessitated. Several updates of the P450 gene superfamily have been published (Nebert *et al.*, 1989; Nebert *et al.*, 1991; Nelson *et al.*, 1993; Nelson *et al.*, 1996). In addition, an official website has been established, based at the University of Memphis, to provide up-to-date information on P450 multiplicity in all species (Nelson, 2006).

Nomenclature recommendations CYP450 (Nebert *et al.*, 1987; Nebert *et al.*, 1991; Nelson, 2009) for naming a gene include:

- The root symbol CYP for 'cytochrome P450';
- An Arabic number for the P450 family (at least 40% sequence homology);
- A letter for the subfamily (more than 55% sequence homology); and
- An Arabic numeral for the individual gene.

When describing a P450 gene, all letters and numerals are written in italics. The same nomenclature is recommended for the enzyme, but it is written in non-italicised form. Thus, the *CYP2D6* gene encodes the CYP2D6 enzyme.

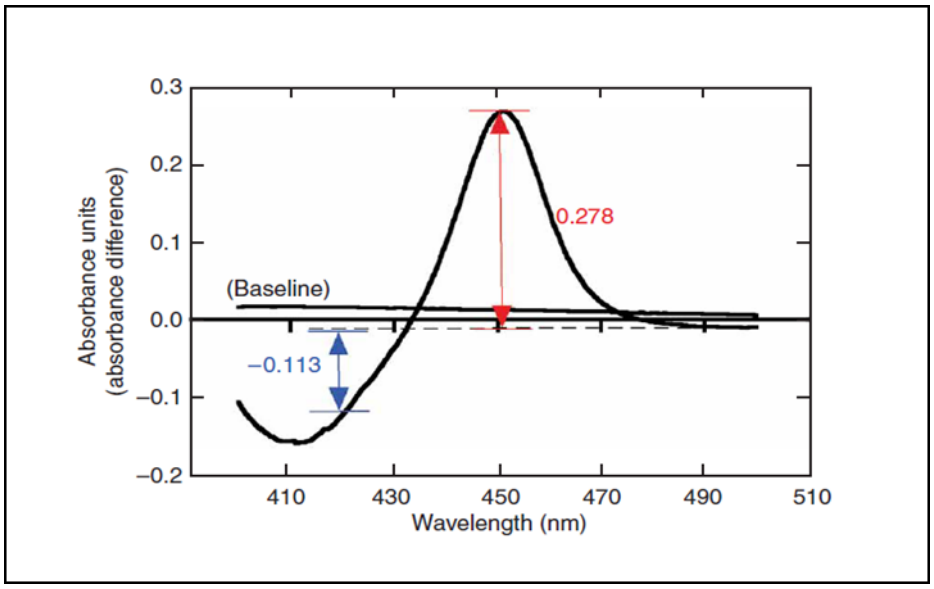


Figure 1.1: Typical cytochrome P450 reduced-CO difference spectrum (taken from Guengerich *et al.*, 2009). The P450 spectral assay is based on the principle that the ferrous form (Fe^{2+}) of the hemoprotein reacts with the carbon monoxide (CO) to form a CO-bound complex that distinctively generates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

1.3.2. P450s' role in drug development or as drug targets

The *P450* gene family, which encodes the P450 enzyme system, is one of the most widely studied topics in drug development. The drug-metabolising enzymes, which belong to the P450 family, are polymorphic (meaning that they have more than one variant of the gene).

P450s have been shown to play highly important roles in the metabolism of drugs and xenobiotics, as well as in the biosynthesis of a variety of endogenous compounds, many of them displaying hormonal function. The role of P450s as therapeutic targets is still inadequately recognised, although several P450 inhibitors became efficient drugs that even reached blockbuster status.

Whereas the majority of clinically used drugs are inactivated by P450s, several prodrugs are bio-converted to their active species by these enzymes. Therefore, this mechanism has been exploited to a greater extent, e.g. by taking advantage of the different P450 enzymes to achieve targeted drug delivery, to improve efficacy or to decrease the unwanted adverse effects of existing and novel drug molecules (Huttunen *et al.* 2008).

P450s are key enzymes in cancer formation and cancer treatment. They mediate the metabolic activation of numerous pre-carcinogens and participate in the inactivation and activation of anticancer drugs (Rodriguez-Antona & Ingelman-Sundberg, 2006). For that reason, a major objective of cancer research is the development of therapeutic agents specifically targeted at tumour cells.

P450s expressed at higher levels in the tumour cells than in the surrounding normal tissue offer therapeutic options by the activation of prodrugs specifically in the cancer cells and avoiding undesirable systemic effects (Riddick *et al.*, 2005). In this respect, there are therapeutic options and opportunities arising from both the enhanced endogenous expression

of P450 in tumours and P450-mediated gene therapy. Concerning endogenous overexpression of individual forms of P450 enzymes in tumour cells, CYP1B1 has been the best studied example, because although several CYP1As, CYP2Cs and CYP3As exhibit enhanced expression in some tumour cells, these enzymes also display considerable expression in normal tissue, mainly in the liver. On the other hand, CYP1B1 mRNA and protein expression have been found in a wide range of malignant tumours and in metastatic disease (McFadyen *et al.*, 2001), but the CYP1B1 protein is generally not detected in normal tissue at important levels (Gibson *et al.*, 2003).

Another P450 used as a drug target is the CYP2W1. In cancer tissues and transfected cells, CYP2W1 gives multiple immunoreactivity bands, suggesting that the protein might be subjected to posttranslational modifications. *In vitro* and *in vivo* studies revealed that CYP2W1 undergoes glycosylation at Asn177, which is enabled by the unique inverted topology of the protein in the endoplasmic reticulum membrane. Regardless of the reversed emplacement, CYP2W1 retains its catalytic function in cell systems and is capable of converting inactive substrates to potent cytotoxic species, demonstrating the potential for prodrug activation. Efficient prodrugs have been developed, providing a basis for a novel therapeutic approach in colon cancer chemotherapy (Travica *et al.*, 2013).

Moreover, P450s are the target of special interest in the development of drugs for skin diseases because most – if not all – drugs available in the armamentarium of dermatologists are substrates, inducers, or inhibitors of this enzyme family. The functional significance of drug metabolism in skin and the implication of P450s in skin pathology and therapy is an area for future investigation. Detailed insight into the mechanism of action of various cutaneous P450s, being capable of modulating the drug bioavailability, will be helpful in the development of better strategies for novel therapy against constantly increasing skin disorders (Ahmad & Mukhtar, 2004).

1.4. CYP5619: A novel P450 family

Researchers across the world are trying to understand oomycetes to control the diseases they cause and develop novel drugs against these pathogens. In this regard, researchers from the Unit for Drug Discovery Research, Department of Health Sciences, Faculty of Health and Environmental Sciences, Central University of Technology (CUT), analysed P450 proteins in 13 pathogenic oomycete genomes (Sello *et al*., 2015) (Table 1.3). Genome-wide identification and annotation of P450s in the 13 oomycetes belonging to two different classes and three different orders revealed the presence of a moderate number of P450s in their genomes. The analysis also resulted in the discovery of a novel P450 monooxygenase protein family, namely the CYP5619. Members of this novel P450 family are fusion proteins with an N-terminal P450 domain fused to a heme-dioxygenase/peroxidase domain (Sello *et al.* 2015) (Figure 1.2). This protein family was discovered in *S. diclina*. The authors suggested that members of this P450 family can serve as novel drug targets against oomycete pathogens.

Table 1.3. Comparative analysis of P450s in oomycetes (taken from Sello *et al*., 2015).

Species name	No. of P450s	No. of P450 families	No. of P450 subfamilies
Phytophthora sojae	30	4	18
Phytophthora parasitica	31	4	18
Phytophthora ramorum	24	4	17
Phytophthora infestans	20	3	14
Phytophthora capsici	28	3	17
Hyaloperonospora arabidopsidis	7	2	7
Pythium irregulare	41	3	17
Pythium aphanidermatum	31	4	18
Pythium ultimum	19	3	12
Pythium iwayamai	42	3	19
Pythium vexan	20	4	15
Saprolegnia parasitica	24	6	16
Saprolegnia diclina	39	9	26

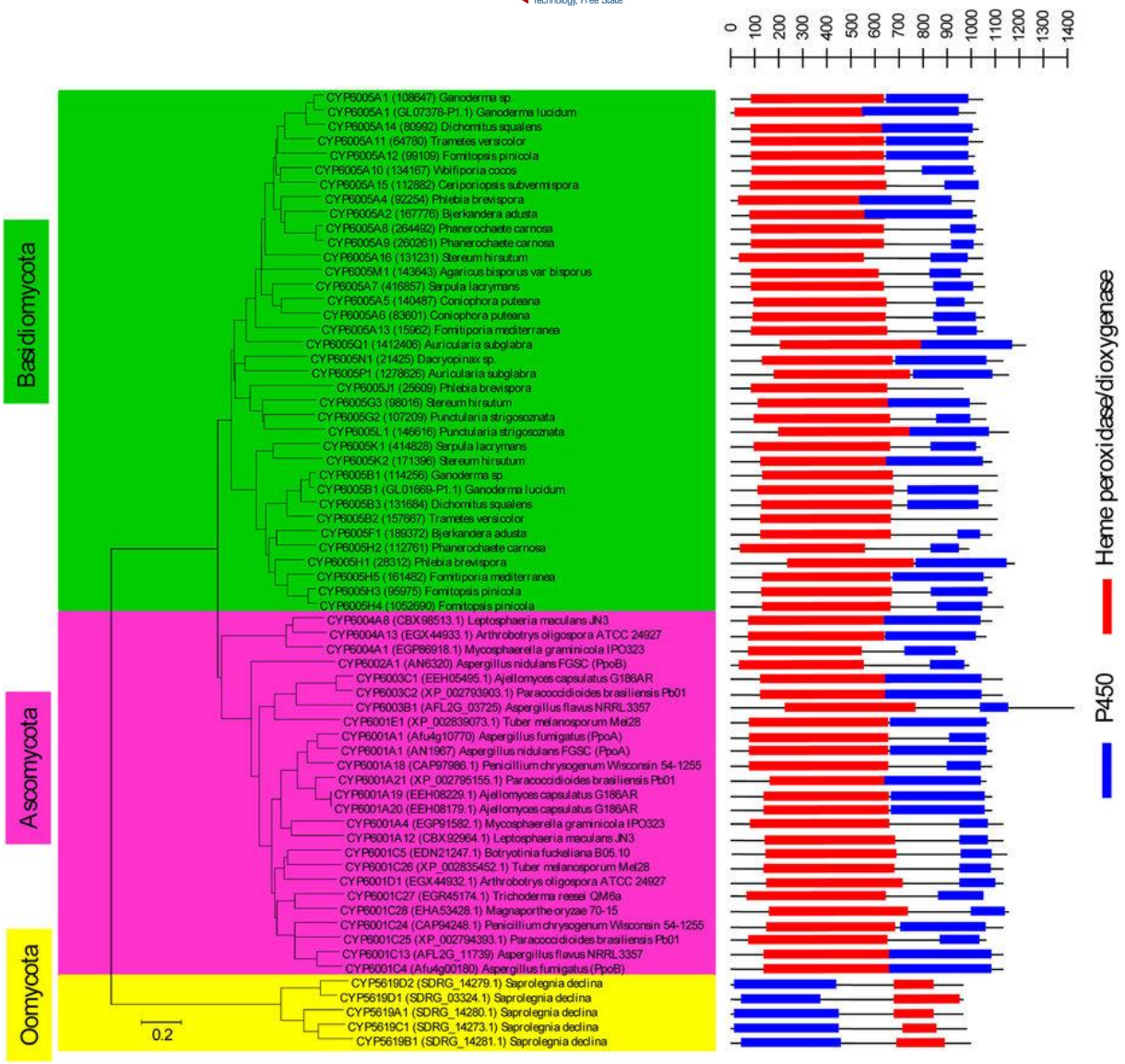


Figure 1.2. Phylogenetic analysis of oomycetes novel P450 family CYP5619 with other fused P450 family members from fungi (taken from Sello *et al*., 2015).

1.5. Rationale and aims of the study

Genome sequencing of lower eukaryotes such as fungi revealed high diversity in their genomes compared to other biological kingdoms (Sello *et al*., 2015). For example, not only was the presence of a large number of P450s detected in many of their genomes, but also high diversity in terms of the number of P450 families (Sello *et al*., 2015). P450s are heme-thiolate proteins distributed across the biological kingdoms (Nelson, 2013) with immense catalytic activity and diverse substrate acceptance, which have prompted the use of these enzymes as potential catalysts for the production of fine chemicals (Guengerich, 2002), pharmaceutical compounds (Ingelman-Sundberg, 2004; Guengerich, 2006), antibiotics, fragrances and detoxification of carcinogenic and/or mutagenic compounds (Urlacher & Eiben, 2006). Progress has been made in understanding P450s from lower eukaryotic organisms, which led to the unravelling of their potential as anti-fungal drug targets (Yoshida, 1988; Jawallapersand *et al*., 2014).

The lower eukaryotes belonging to the kingdom *Stramenopila*, especially phylum *Oomycota* species P450s, have been underexplored. Oomycetes are “hard-wired parasites” that remain a serious problem in agriculture and aquaculture (Phillips *et al*., 2008) and are counted among the most widespread and deadliest disease-causing agents of plants and crops worldwide. The impact of oomycete species on the economy triggered various investigations on pathogenesis and control methods for these pathogens.

In the quest to find a remedy, genome sequencing of oomycetes was carried out (Haas *et al*., 2009). Recently, comprehensive comparative P450 genomics in 13 oomycete pathogens was performed by laboratory of the Unit for Drug Discovery Research at CUT, leading to the discovery of a novel P450 that can be used as a drug target against these pathogens, particularly fish pathogens (Sello *et al*., 2014). The novel P450s belong to the CYP5619 family. In order to use this family’s members as novel drug targets, it is of the utmost

importance to perform biochemical and biophysical characterisation of the members. Hence, the present research is proposing to perform *in silico* structural and functional analysis of CYP5619A1 from *S. diclina* as well as cloning and generation of recombinant *E. coli* cells containing the CYP5619A1 gene in a novel expression vector.

References

- Agrios, G.N., 2005. Plant diseases caused by fungi. *Plant pathology*, *4*.
- Ahmad, N. and Mukhtar, H., 2004. Cytochrome p450: A target for drug development for skin diseases. *Journal of Investigative Dermatology*, *123*(3), pp. 417-425.
- Alderman, D.J., 1994. Control of oomycete pathogens in aquaculture. *Salmon Saprolegniasis. Bonneville Power Administration, Div. Fish and Wildlife, Portland, OR*, pp. 111-130.
- Barnard, N.D. and Raymond, J., 1999. Foods that fight pain: Revolutionary new strategies for maximum pain relief. Harmony.
- Beakes, G.W., Glockling, S.L. and Sekimoto, S., 2012. The evolutionary phylogeny of the oomycete "fungi". *Protoplasma*, *249*(1), pp. 3-19.
- Becker, W., Lyhne, N., Pedersen, A.N., Aro, A., Fogelholm, M., Phorsdottir, I., Alexander, J., Anderssen, S.A., Meltzer, H.M. and Pedersen, J.I., 2004. Nordic Nutrition Recommendations 2004 - Integrating nutrition and physical activity. *Scandinavian Journal of Nutrition*, *48*(4), pp. 178-187.
- Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology*, *124*(1), pp. 128-145.
- Bly, J.E., Lawson, L.A., Dale, D.J., Szalai, A.J., Durburow, R.M. and Clem, L.W., 1992. Winter saprolegniosis in channel catfish. *Diseases of Aquatic Organisms*, *13*(3), pp. 155-164.
- Bly, J.E., Lawson, L.A., Szalai, A.J. and Clem, L.W., 1993. Environmental factors affecting outbreaks of winter saprolegniosis in channel catfish, Ictalurus punctatus (Rafinesque). *Journal of Fish Diseases*, *16*(6), pp. 541-549.
- Brodhun, F., Göbel, C., Hornung, E. and Feussner, I., 2009. Identification of PpoA from Aspergillus nidulans as a fusion protein of a fatty acid heme

dioxygenase/peroxidase and a cytochrome P450. *Journal of Biological Chemistry, 284*(18), pp. 11792-11805.

- David, J.P., Ismail, H.M., Chandor-Proust, A. and 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. *Phil. Trans. R. Soc. B, 368*(1612), p. 20120429.
- Davis, B. and Melina, V., 2000. Becoming vegan: The complete uide to adopting a healthy plant-based diet. Summertown, TN.
- Davis, B.C. and Kris-Etherton, P.M., 2003. Achieving optimal essential fatty acid status in vegetarians: Current knowledge and practical implications. *The American Journal of Clinical Nutrition, 78*(3), pp. 640S-646S.
- Denisov, I.G., Makris, T.M., Sligar, S.G. and Schlichting, I., 2005. Structure and chemistry of cytochrome P450. *Chemical Reviews, 105*(6), pp. 2253-2278.
- Dick, M.W., 2013. Straminipilous Fungi: Systematics of the Peronosporomycetes including accounts of the marine straminipilous protists, the plasmodiophorids and similar organisms. Springer Science & Business Media.
- Diéguez-Uribeondo, J., Cerenius, L. and Söderhäll, K., 1994. Saprolegnia parasitica and its virulence on three different species of freshwater crayfish. *Aquaculture, 120*(3-4), pp. 219-228.
- Diéguez-Uribeondo, J., Fregeneda-Grandes, J.M., Cerenius, L., Pérez-Iniesta, E., Aller-Gancedo, J.M., Tellería, M.T., Söderhäll, K. and Martín, M.P., 2007. Re-evaluation of the enigmatic species complex *Saprolegnia diclina–Saprolegnia parasitica* based on morphological, physiological and molecular data. *Fungal Genetics and Biology, 44*(7), pp. 585-601.

- Erwin, D.C. and Ribeiro, O.K., 1996. Phytophthora diseases worldwide. *American Phytopathological Society (APS Press).*
- Fernández-Benéitez, M.J., Ortiz-Santaliestra, M.E., Lizana, M. and Diéguez-Uribeondo, J., 2008. *Saprolegnia diclina*: Another species responsible for the emergent disease 'Saprolegnia infections' in amphibians. *FEMS Microbiology Letters, 279*(1), pp. 23-29.
- Franzen-Castle, L.D. and Ritter-Gooder, P., 2010. Omega-3 and Omega-6 fatty acids. *NebGuide. University of Nebraska. Institute of Agriculture and Natural Resources. Nebraska (US).*
- Fregeneda Grandes, J.M., Fernandez Diez, M. and Aller Gancedo, J.M., 2001. Experimental pathogenicity in rainbow trout, Oncorhynchus mykiss (Walbaum), of two distinct morphotypes of long-spined Saprolegnia isolates obtained from wild brown trout, Salmo trutta L., and river water. *Journal of Fish Diseases, 24*(6), pp. 351-359.
- Gibson, P., Gill, J.H., Khan, P.A., Seargent, J.M., Martin, S.W., Batman, P.A., Griffith, J., Bradley, C., Double, J.A., Bibby, M.C. and Loadman, P.M., 2003. Cytochrome P450 1B1 (CYP1B1) is overexpressed in human colon adenocarcinomas relative to normal colon: Implications for drug development 1. *Molecular Cancer Therapeutics, 2*(6), pp. 527-534.
- Gieseker, C.M., Serfling, S.G. and Reimschuessel, R., 2006. Formalin treatment to reduce mortality associated with Saprolegnia parasitica in rainbow trout, Oncorhynchus mykiss. *Aquaculture, 253*(1), pp. 120-129.
- Groff, J.L., Gropper, S.S. and Hunt, S.M., 1995. The fat soluble vitamins. *Advanced Nutrition and Human Metabolism.* Minneapolis: West Publishing Company, pp. 284-324.

- Guengerich, F.P., 2002. Cytochrome P450 enzymes in the generation of commercial products. *Nature Reviews Drug Discovery, 1*(5), pp. 359-366.
- Guengerich, F.P., 2006. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *The AAPS Journal, 8*(1), pp. E101-E111.
- Guengerich, F.P., Martin, M.V., Sohl, C.D. and Cheng, Q., 2009. Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nature Protocols, 4*(9), pp. 1245-1251.
- Gunderson, J.H., Elwood, H., Ingold, A., Kindle, K. and Sogin, M.L., 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proceedings of the National Academy of Sciences, 84*(16), pp. 5823-5827.
- Haag, M., 2003. Essential fatty acids and the brain. *The Canadian Journal of Psychiatry, 48*(3), pp. 195-203.
- Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T. and Bozkurt, T.O., 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature, 461*(7262), pp. 393-398.
- Hatai, K. and Hoshiai, G., 1992. Mass mortality in cultured coho salmon (Oncorhynchus kisutch) due to Saprolegnia parasitica coker. *Journal of Wildlife Diseases, 28*(4), pp. 532-536.
- Helbling, E.W., Zagarese, H.E. and Zagarese, H., 2003. UV effects in aquatic organisms and ecosystems. Comprehensive Series in Photochemical and Photobiological Sciences. Roy. Soc. Chem, Cambridge.
- Hmso, U.K., 1994. Nutritional aspects of cardiovascular disease (report on health and social subjects No. 46).

- Hussein, M. and Hatai, K., 2002. Pathogenicity of Saprolegnia species associated with outbreaks of salmonid saprolegniosis in Japan. *Fisheries Science, 68*(5), pp. 1067-1072.
- Huttunen, K.M., Mahonen, N., Raunio, H. and Rautio, J., 2008. Cytochrome P450-activated prodrugs: Targeted drug delivery. *Current Medicinal Chemistry, 15*(23), pp.2346-2365.
- Indu, M., 1992. n-3 fatty acids in Indian diets — Comparison of the effects of precursor (alpha-linolenic acid) Vs product (long chain n-3 poly unsaturated fatty acids). *Nutrition Research, 12*(4-5), pp. 569-582.
- 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), pp. 193-200.
- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and Syed, K., 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: Comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. *PloS one, 9*(9), p. e107209.
- Jiang, R.H. and Tyler, B.M., 2012. Mechanisms and evolution of virulence in oomycetes. *Annual Review of Phytopathology, 50*, pp. 295-318.
- Kamoun, S., 2003. Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell, 2*(2), pp. 191-199.
- Kiesecker, J.M., Blaustein, A.R. and Belden, L.K., 2001. Complex causes of amphibian population declines. *Nature, 410*(6829), pp. 681-684.
- Klingenberg, M., 1958. Pigments of rat liver microsomes. *Archives of Biochemistry and Biophysics, 75*(2), pp. 376-386.

- Kris-Etherton, P.M., Taylor, D.S., Yu-Poth, S., Huth, P., Moriarty, K., Fishell, V., Hargrove, R.L., Zhao, G. and Etherton, T.D., 2000. Polyunsaturated fatty acids in the food chain in the United States. *The American Journal of Clinical Nutrition*, *71*(1), pp. 179S-188S.
- Lamour, K. and Kamoun, S., 2009. *Oomycete genetics and genomics: Diversity, interactions and research tools.* John Wiley & Sons.
- Linscheer, W.G., and Vergroesen, A.J. *Lipids*. in: Shils, M.E., Olson, J.A. and Shike, M. eds., 1994. *Modern nutrition in health and disease* (Vol. 2, pp. 940-945). Philadelphia: Lea & Febiger.
- Marking, L.L., Rach, J.J. and Schreier, T.M., 1994. Search for antifungal agents in fish culture. *Salmon Saprolegniasis. Bonneville Power Administration, Div. Fish and Wildlife, Portland, OR*, pp. 131-148.
- Masters, C., 1996. Omega-3 fatty acids and the peroxisome. *Molecular and Cellular Biochemistry*, *165*(2), pp. 83-93.
- McFadyen, M.C.E., Cruickshank, M.E., Miller, I.D., McLeod, H.L., Melvin, W.T., Haites, N.E., Parkin, D. and Murray, G.I., 2001. Cytochrome P450 CYP1B1 over-expression in primary and metastatic ovarian cancer. *British Journal of Cancer*, *85*(2), p. 242.
- Meyer, F.P., 1991. Aquaculture disease and health management. *Journal of Animal Science*, *69*(10), pp. 4201-4208.
- Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W. and Phillips, I.R., 1987. The P450 gene superfamily: Recommended nomenclature. *DNA*, *6*(1), pp.-11.
- Nebert, D.W., Nelson, D.R., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B. and Levin, W., 1989.

The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *DNA*, *8*(1), pp.-13.

- 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. and Loper, J.C., 1991. The P450 superfamily: Update on new sequences, gene mapping, and recommended nomenclature. *DNA and Cell Biology*, *10*(1), pp. 1-14.
- Neitzel, D.A., Elston, R.A. and Abernethy, C.S., 2004. *Prevention of prespawning mortality: Cause of salmon headburns and cranial lesions*. Pacific Northwest National Laboratory.
- Nelson, D.R., 1999. Cytochrome P450 and the individuality of species. *Archives of Biochemistry and Biophysics*, *369*(1), pp. 1-10.
- Nelson, D.R., 2006. Cytochrome P450 nomenclature, 2004. *Cytochrome P450 Protocols*, pp. 1-10.
- Nelson, D.R., 2009. The cytochrome p450 homepage. *Human Genomics*, *4*(1), p. 59.
- Nelson, D.R., 2011. Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1814*(1), pp. 14-18.
- Nelson, D.R., 2013. A world of cytochrome P450s.
- Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., Gotoh, O. and Okuda, K., 1993. The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and Cell Biology*, *12*(1), pp. 1-51.
- Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W. and Gunsalus, I.C.,

1996. P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics and Genomics*, *6*(1), pp. 1-42.

- Omura, T. and Sato, R., 1962. A new cytochrome in liver microsomes. *Journal of Biological Chemistry*, *237*(4), pp. PC1375-PC1376.
- Park, S.Y., Yamane, K., Adachi, S.I., Shiro, Y., Weiss, K.E., Maves, S.A. and Sligar, S.G., 2002. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: High resolution structure and functional properties. *Journal of Inorganic Biochemistry*, *91*(4), pp. 491-501.
- Phillips, A.J., Anderson, V.L., Robertson, E.J., Secombes, C.J. and van West, P., 2008. New insights into animal pathogenic oomycetes. *Trends in Microbiology*, *16*(1), pp. 13-19.
- Poulos, T.L., Finzel, B.C. and Howard, A.J., 1987. High-resolution crystal structure of cytochrome P450cam. *Journal of Molecular Biology*, *195*(3), pp. 687-700.
- Riddick, D.S., Lee, C., Ramji, S., Chinje, E.C., Cowen, R.L., Williams, K.J., Patterson, A.V., Stratford, I.J., Morrow, C.S., Townsend, A.J. and Jounaidi, Y., 2005. Cancer chemotherapy and drug metabolism. *Drug Metabolism and Disposition*, *33*(8), pp. 1083-1096.
- Robertson, E.J., Anderson, V.L., Phillips, A.J., Secombes, C.J., Diéguez-Uribeondo, J. and van West, P., 2009. Saprolegnia—fish interactions. *Oomycete Genetics and Genomics: Diversity, Interactions, and Research Tools*, pp.407-424.
- Rodriguez-Antona, C. and Ingelman-Sundberg, M., 2006. Cytochrome P450 pharmacogenetics and cancer. *Oncogene*, *25*(11), pp. 1679-1691.
- Rylott, E.L., Jackson, R.G., Sabbadin, F., Seth-Smith, H.M., Edwards, J., Chong, C.S., Strand, S.E., Grogan, G. and Bruce, N.C., 2011a. The explosive-degrading cytochrome P450 XplA: Biochemistry, structural features and prospects for

bioremediation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1814*(1), pp. 230-236.

- Rylott, E.L., Lorenz, A. and Bruce, N.C., 2011b. Biodegradation and biotransformation of explosives. *Current Opinion in Biotechnology, 22*(3), pp. 434-440.
- Scientific Review Committee, 1990. Nutrition recommendations. *Ottawa: Canadian Government Publishing Centre, Supply and Services Canada*.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific Reports, 5*, p. 11572.
- Simopoulos, A.P., Leaf, A. and Salem Jr, N., 1999. Workshop on the essentiality of and recommended dietary intakes for omega-6 and omega-3 fatty acids. *Journal of the American College of Nutrition (18)* , Iss. 5.
- Singh, M., 2003. Nutrition, brain and environment: How to have smarter babies?. *Indian Pediatrics, 40*(3), pp. 213-220.
- Singh, M., 2005. Essential fatty acids, DHA and human brain. *Indian Journal of Pediatrics, 72*(3), pp. 239-242.
- Soanes, D.M., Richards, T.A. and Talbot, N.J., 2007. Insights from sequencing fungal and oomycete genomes: What can we learn about plant disease and the evolution of pathogenicity? *The Plant Cell, 19*(11), pp. 3318-3326.
- Stueland, S., Hatai, K. and Skaar, I., 2005. Morphological and physiological characteristics of Saprolegnia spp. strains pathogenic to Atlantic salmon, Salmo salar L. *Journal of Fish Diseases, 28*(8), pp. 445-453.

- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and 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 and Genomics, 36*(3), pp. 321-333..
- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014a. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One, 9*(1), p. e86683.
- Thines, M. and Kamoun, S., 2010. Oomycete–plant coevolution: Recent advances and future prospects. *Current Opinion in Plant Biology, 13*(4), pp. 427-433.
- Thines, M., 2014. Phylogeny and evolution of plant pathogenic oomycetes — A global overview. *European Journal of Plant Pathology, 138*(3), pp. 431-447.
- Tiffney, W.N., 1939. The host range of Saprolegnia parasitica. *Mycologia, 31*(3), pp. 310-321.
- Travica, S., Pors, K., Loadman, P.M., Shnyder, S.D., Johansson, I., Alandas, M.N., Sheldrake, H.M., Mkrtchian, S., Patterson, L.H. and Ingelman-Sundberg, M., 2013. Colon cancer – Specific cytochrome P450 2W1 converts duocarmycin analogues into potent tumor cytotoxins. *Clinical Cancer Research, 19*(11), pp. 2952-2961.
- Trumbo, P., Schlicker, S., Yates, A.A. and Poos, M., 2002. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. *Journal of the American Dietetic Association, 102*(11), pp. 1621-1630.
- Urlacher, V.B. and Eiben, S., 2006. Cytochrome P450 monooxygenases: Perspectives for synthetic application. *Trends in Biotechnology, 24*(7), pp. 324-330.
- van West, P., 2006. Saprolegnia parasitica, an oomycete pathogen with a fishy appetite: New challenges for an old problem. *Mycologist, 20*(3), pp. 99-104.

- WHO and FAO Joint Consultation: fats and oils in human nutrition. *Nutr Rev* 1995; 53, pp. 202–5.
- Yoshida, Y., 1988. Cytochrome P450 of fungi: Primary target for azole antifungal agents. In *Current topics in medical mycology* (pp. 388-418). New York: Springer.
- Zaror, L., Collado, L., Bohle, H., Landskron, E., Montaña, J. and Avendaño, F., 2004. Saprolegnia parasitica en salmones y truchas del sur de Chile. *Archivos de medicina veterinaria, 36*(1), pp. 71-78.

CHAPTER 2

PHYLOGENETIC AND SECONDARY STRUCTURE ANALYSIS OF NOVEL P450 FAMILY CYP5619 FROM *SAPROLEGNIA DICLINA*

2.1. Introduction

Cytochrome P450 monooxygenases (CYPs or P450s) have potential biotechnological values, including their use as drug targets against pathogens (Kelly & Kelly, 2013; Jawallapersand *et al.*, 2014). Some of their functions include the catabolism of compounds used as carbon sources by the host organism, as well as the synthesis of biologically active compounds such as prostaglandins, steroids and arachidonate metabolites. A high percentage of P450s' substrates are hydrophobic compounds. Products of P450s activity are made more water-soluble by monooxygenation. Because of this reaction of P450 and their substrates, some authors suggested that P450s could be engineered to break down environmentally toxic compounds such as polycyclic aromatic hydrocarbons and fluorocarbons (Urlacher & Eiben, 2006). Another perspective in P450 application is the use of these proteins in the monooxygenation of compounds in a region- and stereo-selective manner to obtain products that have been proven to be of great value in biological processes. For example, P450s can be used to convert arachidonic acid to 14,15- epoxyeicosatrienoic acid, found to play a major role in the regulation of potassium and calcium influx in kidneys (Graham-Lorence *et al.*, 1997; Chen *et al.*, 1999). It is therefore of the utmost importance to have a good understanding of P450s' structures and eventually their catalytic mechanism, to accomplish these projects.

P450s have been reported to have a common structural fold for all determined structures (Graham & Peterson, 1999; Sirim *et al.*, 2010). They have enough diversity in the primary, secondary, and tertiary sequences to accommodate specific substrates and redox

partners and additionally to target the cellular location of the protein. Redox partners are key proteins in the functionality of P450s (Hannemann *et al.*, 2007). They form systems that have been shown to be extremely diverse. Usually fused to their redox partner, some P450s have nevertheless been reported to stand alone, and others to be separated from their protein partner (Guengerich & Munro, 2013) (Figure 2.1).

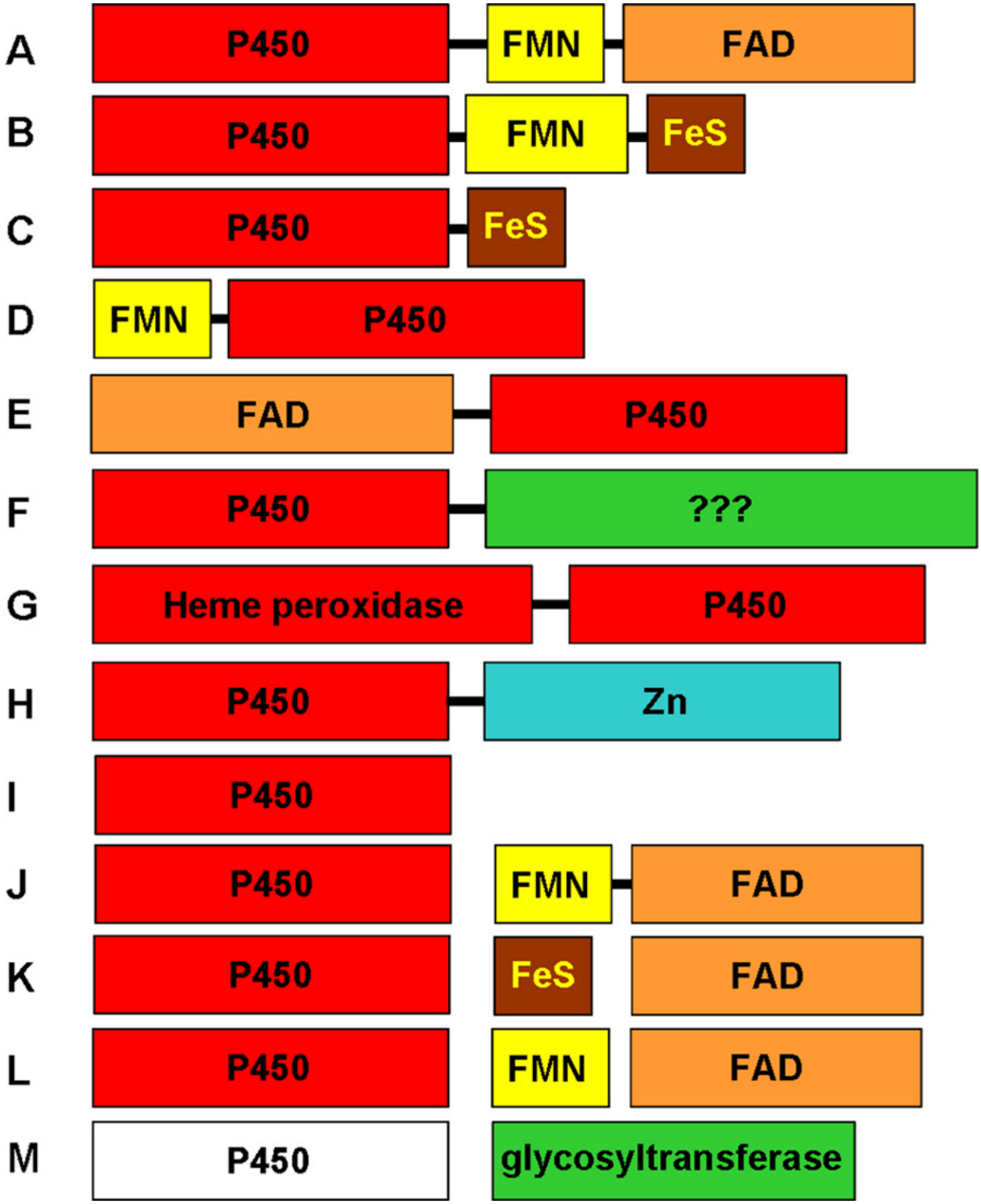


Figure 2.1. Diversity of P450 redox systems and P450 fusion proteins (taken from Guengerich & Munro, 2013). A selection of distinct types of P450 enzymes and (where relevant) their redox partner systems is shown. The sizes of the boxes are indicative of the lengths of the protein modules. Bound prosthetic groups are indicated in the colour-coded

domains. A, P450BM-3 (CYP102A1)-type P450-CPR fusion, also seen for fungal P450foxy (CYP505)-type systems (Munro *et al.*, 2002). B, CYP116B-type P450-phthalate dioxygenase reductase fusion (Warman *et al.*, 2012). C, *M. capsulatus* P450-FDx fusion CYP51FX (Jackson *et al.*, 2002). D, *R. rhodochrous* P450-flavodoxin fusion XplA, involved in reductive degradation of explosives (Rylott *et al.*, 2011). E, *Pseudomonas fluorescens* PfO-1 acyl-CoA dehydrogenase-P450 fusion CYP222A1. This protein is depicted with FAD bound in its N-terminal domain, but there is no report to date of characterisation of this protein. F, Mimivirus CYP5253A1, with a P450 fused to a C-terminal domain of uncertain function but containing several potential sites for post-translational modification. G, PpoA dioxygenase/peroxidase-P450 fusion enzyme from *A. nidulans*, involved in Psi factor production (Brodhun *et al.*, 2009). H, P450-hydrolase fusion CYP631B5, involved in mycophenolic acid production (Hansen *et al.*, 2012). I, "stand-alone" P450 that acts without partner proteins, typified by P450nor (CYP55A)-type nitric-oxide reductase enzymes that interact directly with NAD(P)H, peroxygenase CYP152 P450s that use H_2O_2 to oxidize substrates, P450s that isomerize substrates (e.g. CYP5A1/8A1), and allene oxide synthase (CYP74A) dehydratase P450s. J, typical eukaryotic Class II P450 systems with separate membrane-associated P450 and a CPR partner. K, Class I (mitochondrial) P450 system that interacts with the iron-sulphur protein ADx, which is in turn reduced by ADR. Most bacterial systems use a similar redox apparatus (Munro *et al.*, 2007). L, variation on system K, in which a flavodoxin replaces the iron-sulphur protein. This type of system supports CYP176A1 (P450cin), enables *Citrobacter braakii* to catabolise cineole and can reduce CYP107H1 (P450BioI), involved in *Bacillus subtilis* biotin synthesis (Lawson *et al.*, 2004; Hawkes et al., 2010). M, heme-free EryCII P450-like protein devoid of a cysteine proximal ligand. EryCII is an allosteric activator of the glycosyltransferase EryCIII in the production of erythromycin D in *Saccharopolyspora erythraea* (Moncrieffe *et al.*, 2012).

P450s have been underexplored in the kingdom *Stramenopila*, especially in the phylum *Oomycota*. Species from that phylum are "hard-wired" parasites, fungi-like organisms, well known for causing diseases in plants and animals, resulting in a huge financial loss in aquafarming and agriculture (Phillips *et al.*, 2008). Research has been done to understand these pathogens and in order to find a remedy, genome sequencing of oomycetes has been carried out (Haas *et al.*, 2009). In this direction recently the Unit for Drug Discovery Research (UDDR)'s laboratory at CUT completed comparative P450 genomics in the 13 oomycetes (Sello *et al.*, 2015). Annotation of P450s in 13 oomycetes resulted in the discovery of a novel P450 family, namely the CYP5619 family in *Saprolegnia diclina*, with a P450 domain at the N-terminal, and a peroxidase domain at the C-terminal (Sello *et al.*, 2015) (Figure 2.2).

However, to date, study on analysis of CYP5619 P450s in other biological kingdoms or secondary structure analysis with respect to P450 helices and sheets has not been reported. This chapter is aimed at performing phylogenetic and secondary structure analysis of the CYP5619 family.

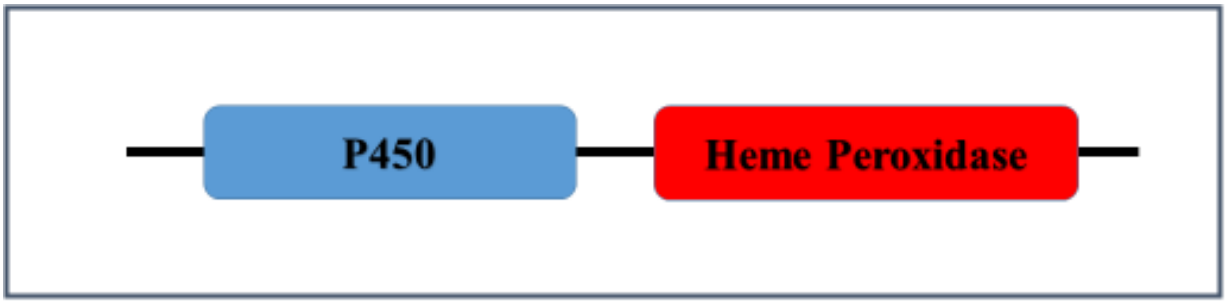


Figure 2.2. P450 fusion proteins pattern observed in *Saprolegnia diclina*. The N- terminal P450 is fused to a C-terminal heme-peroxidase domain.

2.2. Methodology

2.2.1. Sequence retrieval

The protein sequences of six members of the CYP5619 family, namely CYP5619A1 (ID: SDRG_14280.1), CYP5619B1 (ID: SDRG_14281.1), CYP5619B2 (ID: SDRG_14277.1), CYP5619C1 (ID: SDRG_14273.1), CYP5619D1 (ID: SDRG_03324.1) and CYP5619D2 (ID: SDRG_14279.1), were obtained from an article published by researchers at the CUT laboratory (Sello *et al.*, 2015). Protein IDs of CYP5619 family members are shown in parenthesis.

2.2.2. Identification of CYP5619 homologs

The CYP5619 members' amino acid sequences were subjected to a protein Basic Local Alignment Search Tool (BLASTp) analysis against all sequences at the NCBI website. For each P450, a set of 100 homologs was downloaded. The homologs for each P450 were subjected to a batch search on the conserved domain database (CDD) (Marchler-Bauer *et al.*, 2010, 2014) in order to identify the domains present in the sequences. The results were downloaded and sorted. Hits exhibiting the presence of both a P450 and lipoxygenase domains were retained and the others were removed. Sequences with a different conformation compared to CYP5619 family members (P450 domain at the N- terminal and lipoxygenase domain at the C-terminal) were also removed from the analysis.

2.2.3. Annotation and classification

Based on the identity percentage, family and subfamily names were assigned. For assigning the family and subfamily names, the standard rule set by the International P450 Nomenclature Committee was followed, i.e. P450s within a family share more than 40% amino acid homology and members of subfamilies share more than 55% amino acid homology.

Furthermore, P450s that showed less than 40% homology with the target P450s were assigned to a new family.

2.2.4. Phylogenetic analysis

Phylogenetic analysis of P450s was carried out in the same way as described in recent publications from the UDDR's laboratory at the CUT (Syed *et al.*, 2014a, 2014b). Briefly, evolutionary analysis was carried out using the minimum evolution method16. The phylogenetic analysis was carried out using Molecular Evolutionary Genetics Analysis (MEGA 5.05) software (Tamura *et al.*, 2011).

2.2.5. Secondary structure prediction

The PROfile Multiple Alignment with Local Structures and 3D constraints (PROMALS3D) (Pie *et al.*, 2008) was used to align protein sequences and to predict secondary structure alignments. It is important to note that PROMALS3D aligns multiple protein sequences and/or structures, with enhanced information from database searches, secondary structure prediction, 3D structures or user-defined constraints and would also give the conservation index.

2.2.6. Analysis of amino acids conservation

This section was conducted using the results from the target-template alignment provided by the PROMALS3D server.

2.3. Results and discussion

2.3.1. CYP5619 members' homologs

The BLAST search revealed the presence of a moderate number of P450 sequences with the same conformation as that of CYP5619 family members. Based on their percentage identities,

the sequences were assigned to new families and subfamilies. In total, 24 P450s belonging to six species were found to have both P450 and heme-peroxidase/lipoxygenase motifs, with the former being N-terminal and the latter fused at its C-terminal. The P450s were grouped into two families (CYP5619 and new family) families and five subfamilies (A, B, C, D, and new subfamily) (Figure 2.1 and Tables 2.1 and 2.2).

2.3.2. Phylogenetic analysis

The BLAST search coupled with the phylogenetic analysis of CYP5619 family members across biological kingdoms revealed the presence of this novel family in other oomycete organisms and in a phytoplankton (Figure 2.1. and Table 2.2). However, the number of CYP5169 members in organisms varied (Table 2.2). Nine CYP5619 members were found in *Achlya hypogyna* and six were found in *S. parasitica* (both oomycetes). *Thraustotheca clavata* and *Aphanomyces invadans* (both oomycetes) were found to have three and two CYP5619 members, respectively. One organism (*Emiliania huxleyi*) of the phytoplankton was found to have two CYP5619 family members, but the smallest count was attributed to an oomycete: *A. astaci*.

Phylogenetic analysis resulted in alignment of the same subfamily members together, suggesting the nomenclature is correct (Figure 2.3). The new family and subfamilies are grouped together (Figure 2.3).

Table 2.1. Annotation of hit proteins. Standard P450 nomenclature was followed to assign families and subfamilies to different P450s.

P450 name	Protein ID	Specie source	Reference P450		
			% Id	Homolog P450	Specie source
NF1SF1	XP_005786468.1	*Emiliania huxleyi* CCMP 1516	96.42	XP_005778763.1	*Emiliania huxleyi* CCMP 1516
NF1SF1	XP_005778763.1	*Emiliania huxleyi* CCMP 1516	96.42	XP_005786468.1	*Emiliania huxleyi* CCMP 1516
CYP5619NSF1	2OQR84833.1	*Achlya hypogyna*	50.37	CYP5619C1	*Saprolegnia diclina* VS20
CYP5619NSF1	XP_012203940. 1	*Saprolegnia parasitica* CBS 223.65	50.47	CYP5619C1	*Saprolegnia diclina* VS20
	2OQR84833.1	*Achlya hypogyna*	62.59	XP_012203940.1	*Saprolegnia parasitica* CBS 223.65
CYP5619D	XP_012194083.1	*Saprolegnia parasitica* CBS 223.65	92.38	CYP5619D1	*Saprolegnia diclina* VS20
CYP5619D	AIG56338.1	*Achlya hypogyna*	64.77	CYP5619D1	*Saprolegnia diclina* VS20
CYP5619D	AIG56100.1	*Achlya hypogyna*	62.83	CYP5619D1	*Saprolegnia diclina* VS20
CYP5619D	AIG56283.1	*Achlya hypogyna*	62.30	CYP5619D1	*Saprolegnia diclina* VS20
CYP5619C	XP_008878127.1	*Aphanomyces invadans*	56.15	CYP5619C1	*Saprolegnia diclina* VS20
CYP5619C	XP_008879406.1	*Aphanomyces invadans*	57.17	CYP5619C1	*Saprolegnia diclina* VS20
CYP5619C	XP_009834503.1	*Aphanomyces astaci*	77.22	XP_008879406.1	*Aphanomyces invadans*
CYP5619A	OQS03666.1	*Thraustotheca clavata*	58.50	CYP5619A1	*Saprolegnia diclina* VS20
CYP5619A	OQS07119.1	*Thraustotheca clavata*	64.92	OQS03666.1	*Thraustotheca clavata*
CYP5619C	XP_012203939.1	*Saprolegnia parasitica* CBS 223.65	95.65	CYP5619C1	*Saprolegnia diclina* VS20

CYP5619C	OQR84821.1	*Achlya hypogyna*	74.67	CYP5619C1	*Saprolegnia diclina* VS20
CYP5619C	1OQR84833.1	*Achlya hypogyna*	79.57	CYP5619C1	*Saprolegnia diclina* VS20
CYP5619B	1OQR84828.1	*Achlya hypogyna*	69.08	CYP5619B1	*Saprolegnia diclina* VS20
CYP5619B	XP_012203946.1	*Saprolegnia parasitica* CBS 223.65	92.98	CYP5619B1	*Saprolegnia diclina* VS20
CYP5619B	OQR84819.1	*Achlya hypogyna*	75.29	CYP5619B2	*Saprolegnia diclina* VS20
CYP5619B	XP_012203942.1	*Saprolegnia parasitica* CBS 223.65	90.69	CYP5619B2	*Saprolegnia diclina* VS20
CYP5619A	OQS07110.1	*Thraustotheca clavata*	65.67	CYP5619A1	*Saprolegnia diclina* VS20
CYP5619A	2OQR84828.1	*Achlya hypogyna*	77.71	CYP5619A1	*Saprolegnia diclina* VS20
CYP5619A	XP_012203945.1	*Saprolegnia parasitica* CBS 223.65	95.88	CYP5619A1	*Saprolegnia diclina* VS20

Table 2.2. Homolog CYP5619 P450 family members from NCBI blast results. All hit proteins were sorted into P450 family domains using the NCBI CDD database (see Appendix) and proteins with N-terminal P450 and C-terminal dioxygenase, the same as CYP5619A1, CYP5619B1, CYP5619B2, CYP5619C1, CYP5619D1 and CYP5619D2, were selected for the study.

Hit protein ID	Species name	Nature of species
1OQR84828.1	*Achlya hypogyna*	Oomycete
1OQR84833.1	*Achlya hypogyna*	Oomycete
2OQR84828.1	*Achlya hypogyna*	Oomycete
2OQR84833.1	*Achlya hypogyna*	Oomycete
AIG56100.1	*Achlya hypogyna*	Oomycete
AIG56283.1	*Achlya hypogyna*	Oomycete
AIG56338.1	*Achlya hypogyna*	Oomycete
OQR84819.1	*Achlya hypogyna*	Oomycete
OQR84821.1	*Achlya hypogyna*	Oomycete
OQS03666.1	*Thraustotheca clavata*	Oomycete
OQS07110.1	*Thraustotheca clavata*	Oomycete
OQS07119.1	*Thraustotheca clavata*	Oomycete
XP_005778763.1	*Emiliania huxleyi* CCMP1516	Phytoplankton
XP_005786468.1	*Emiliania huxleyi* CCMP1516	Phytoplankton
XP_008878127.1	*Aphanomyces invadans*	Oomycete
XP_008879406.1	*Aphanomyces invadans*	Oomycete
XP_009834503.1	*Aphanomyces astaci*	Oomycete
XP_012194083.1	*Saprolegnia parasitica* CBS223.65	Oomycete
XP_012203939.1	*Saprolegnia parasitica* CBS223.65	Oomycete
XP_012203940.1	*Saprolegnia parasitica* CBS223.65	Oomycete
XP_012203942.1	*Saprolegnia parasitica* CBS223.65	Oomycete
XP_012203945.1	*Saprolegnia parasitica* CBS223.65	Oomycete
XP_012203946.1	*Saprolegnia parasitica* CBS223.65	Oomycete

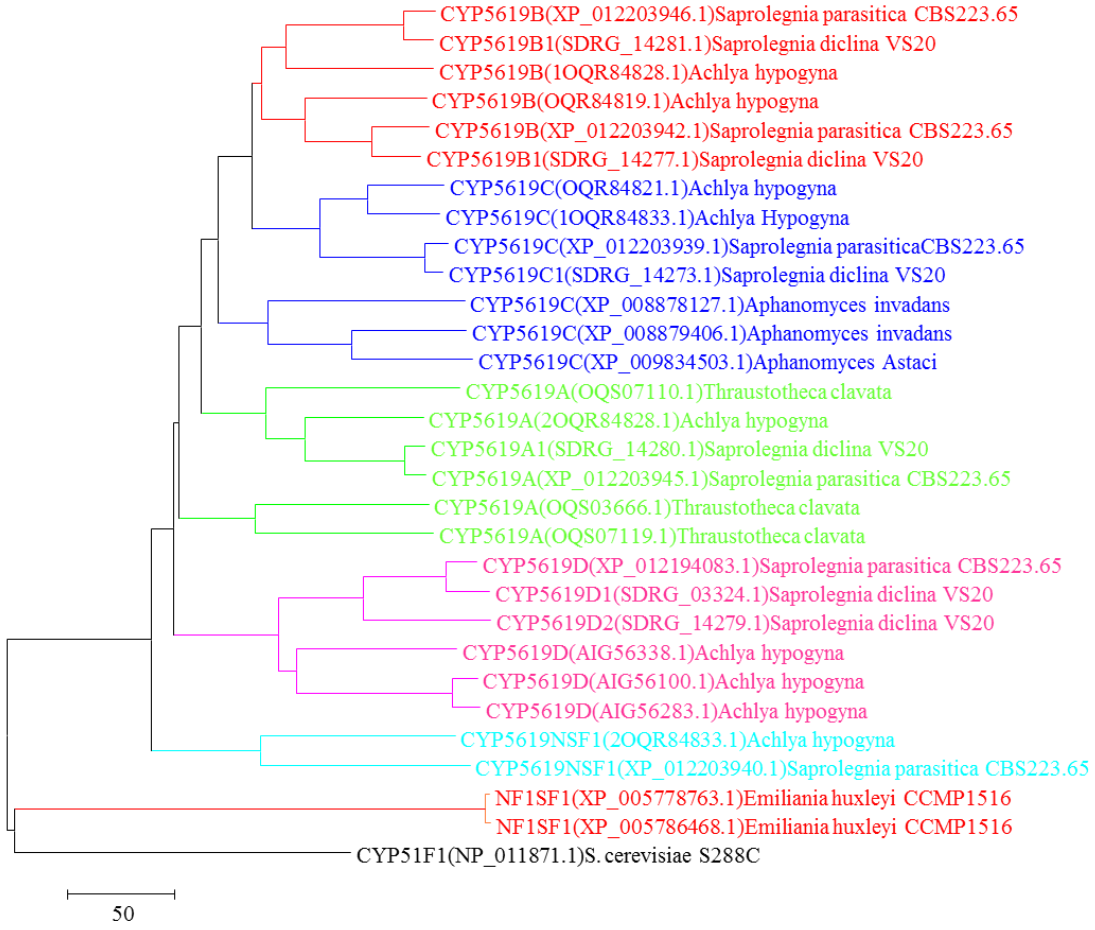


Figure 2.3. Evolutionary analysis of CYP5169 family members. The evolutionary history was inferred using the neighbour-joining method (Saitou & Nei, 1987). The optimal tree with a sum of branch length = 2397.51806641 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method (Nei & Kumar, 2000) and are in the units of the number of amino acid differences per sequence. The analysis involved 30 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 547 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011). CYP51F1 of *Saccharomyces cerevisiae* is used as outgroup. Different CYP5619 subfamilies were indicated with different colours.

2.3.3. Conservation of the secondary structure

The detailed secondary structure analysis of the CYP5619 family members was shown in Figure 2.4. The helices and beta-sheets follow the P450 fold pattern as described by some authors (Graham & Peterson, 1999; Sirim *et al.*, 2010), with a few exceptions such as the CYP5619D subfamily members not showing the supposedly conserved helix L, or the CYP5619C1 lacking the supposedly conserved sheets β1_1, β1_2, β1_4, β1_5 and β2_1. Nevertheless, a high conservation index (conservation index 9) was observed in helix K with the ExxR motif (shaded cyan in Figure 2.4), predicted to be involved in the stabilisation of the core structure. Also, prior to the L-helix, which forms part of the heme-binding region, lies the conserved Cysteine, which is responsible for the 450 nm Soret absorbance observed in CO-bound proteins. It is part of the CxG motif (highlighted in cyan in Figure 2.4), but not present in CYP5619D subfamily.

```
Conservation:                                6     6     6                   6 6    6  6  96 6
CYP5619D1_       1  ----------------------MVSLPLLVIVIVGQVAGAPQGLGSVLQGVINDVKHSVAGVRYVFESLV   48
CYP5619D2        1  ----------------------MVALVPLLLTAVGVVGTQENTLKGFFQGVISDIKHAVTDVRYTFESLV   48
CYP5619A1_       1  MGNLTST---GATHGD-------VHMDKMVMYLDGDRS-------AMM-----D------GDLFVLEKAL   42
CYP5619C1        1  MGNQPST---EAGVAPLPDSKRANSIFSLLAFAKDPKA-------AMA-----ESR-DTLGNLFLIESAV   54
CYP5619B2_       1  MGNEASTVHADGAATDLPASHRAMNILKMIEFSKDPRA-------GML-----ESR-DQFGDLFLLESHL   57
CYP5619B1_       1  MGSQAST---PAGAA--PSLRRAASLKKMIMFMKDPRT-------AMM-----DCR-DHYGDVFLMESSL   52
Consensus_aa:                ....................hhsl..hlbh..ss.t.......thh.....-s+.s.hGs.@lhEShl
Consensus_ss:                                    ehhhhhhhhh hhh     hhhhhhhhh hhhhhhhhheeeeeeee

Conservation:                  966 999696 9  69  6  9   9  9 9 9   9 96   9 96  6  66   6 966   9
CYP5619D1_      49  DAEPVTGFCSPEALRAFDDALASGALERRTAYPTGILELTGPTLSTIDGPAFLKRQDAFLNALSGAALST  118
CYP5619D2       49  NAEPVTGFCSPDALRAFDDALSSGALSRHAAYPAGLLDLSGPTLSTLDGQAFAIRQESLLNALSGPSLAA  118
CYP5619A1_      43  ATEKVVGFCGPEALKVFDANLRDGTFVRHGALPSGLNELLGAVLPTTDGDAHARKKKLVLAAFSGAQLAA  112
CYP5619C1       55  VSEKIAGFCGPEMLSQYDAHVAAGHIVRENALPAGIVELLGPILATLDGDVHDSRKEAIMGAFSKEMLAS  124
CYP5619B2_      58  VSEKIAGFCGPELLAAFDDKLRDGSIVREGAFPPGVLALLGPIMSTIDGEEHDARKAAALEALTPARLDL  127
CYP5619B1_      53  VNEKIMGFCGPEALLAYDTQVKEGKIVRASAFPTGILELLGAVVSTLDDDAHAKRKAALLVAFTPEKLDA  122
Consensus_aa:               ssE.lhGFCtPEhL.h@Ds.l.sG.lsRcsAhPsGll.LhGshhtTlDGp.@s.Rp.thL.Ahos..Lsh
Consensus_ss:                 eeeeee  hhhhhhhhhh  hhh       hhhhhhh    eee   hhhhhhhhhhhhhhhhhhhhh

Conservation:                  9 9 6   66 699 99969   996   6  69  66   69               9 666 6
CYP5619D1_      119  YQPRIQRRIQEDHATWAARGSTFSLALYAKTSTFKVFLDVVYGIDDPEKY------TGHRAQLDEYLFYL  182
CYP5619D2       119  YQPRIQQLIQDDHATWAARGGTFSLALQAKTTTFKVFLAVVYGVTQPDEY------VGYRAQLDEYLEYA  182
CYP5619A1_      113  YKPLIRTTIQNEHAKWAAHGASMSLVANAKVLVFKLSLLLILGLED--NY------DNSRELLDTYMLAL  174
CYP5619C1       125  YAPIVFEIVQKEHAAWAAHGGEISLALSCKKTVFKVFLAILYGITNLTPAEYDAKFDPFRDLLDSFIRAI  194
CYP5619B2_      128  YAPIIREIVEAEHASWAARGGAISLACLTRDMVFRIFLKVLYGVER--HDG-----NKFRVLLDDFIVSI  190
CYP5619B1_      123  YKPKIREIIQHDHAAWAARGGSLSLALSCKKMVFHVFMATLLGLENVD--------DEYRELVEAFVSSI  184
Consensus_aa:               Y.P.Ipphlp.-HAsWAA+GtshSLAh.h+phhF+lFL.llhGlpp..p.......s.@R.bLDp@l..l
Consensus_ss:               hhhhhhhhhhhhhhhhhhh   eehhhhhhhhhhhhhhhhhhhhh   hh      hhhhhhhhhhhhhh

Conservation:                  6   66  6   6       6 99 6 6 696    6 6   6  66    6   6  9 96 6 666
CYP5619D1_      183  SKTSSRAPSDAAKIREHLLAAIVRPAIASSLARVRSGAPLTCVLDTVVAQGTVSEADLALESFQLLAMGL  252
CYP5619D2       183  KKTLSRAPSDAIKIRDRLLATLVRPAIAASHARVRAGAAPTCVLDALVAQNTMSDSDLATEGFQLMAMGL  252
CYP5619A1_      175  RNSVRRADPAGVRSRDELIRTMINPALATSHDRVHTGKPKPCALDHLVAAGVLSDDDLRAELFHLLCMSL  244
CYP5619C1       195  PKSSKGADAEGLVCKQRLLDELVAPALAASQARVEAKAPVPCFLDYMLGQTELTPDVVHLEAFHALFAGL  264
CYP5619B2_      191  RRSSKHADPHGVRCRTQILDELIRPAIANAQARASNKTPVPSVIDCLVANGKMTPDVLETEAFHFLFAGF  260
CYP5619B1_      185  RKSARKPDTTGMDARTQVVEELIRPAVREAKARVAAQKPLPTVVEVLVADGRLSDEELNLELFHALFAGL  254
Consensus_aa:               p+osp+AssphchRppll.pllpPAlAstpARhps..P.sthlDhlVApsphosssL.hE.FphLhhGh
Consensus_ss:               hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh    hhhhhhhhhh     hhhhhhhhhhhh
```

```
Conservation:          9  666 6  99              9  6  99   6  9 96 69   6 6  9969 6 969   6
CYP5619D1_      253  PGLEGLVVHTITAMVSLDDVRGQMATARDAYTAKYPGGA-FWSHLDDLDAVNQYVNEVQRVCGASPRHTF  321
CYP5619D2       253  LGLEGLVVHTITAMVSVDGVRGQLGSARDAYVSKYPNGA-HWRHLDDLSVVNAYVNEVQRVYNASPRHTF  321
CYP5619A1_      245  GGLECWVANCITAAASSTDVLAQLTAGRDAFITKYPAEADRWSHLGDLGYVNNYIQEVKRTYVAGPSHMY  314
CYP5619C1       265  GGTQCLVVNTITALAQYPTVAEKVHASRAKFVIKY--HDDRWRHFDNLGYCNRFLLEVKRFYSAGPAQLF  332
CYP5619B2_      261  GGVACLATNILTAVATHPSARKDLLDARAEYVTKY-DGDARWAHFHDLGYVNLFILEVKRFYVAGPTAVF  329
CYP5619B1_      255  GGVTCLVINAVTACIELPAIREKVSAAREAFLAKYPNEDDRWSHFADLGYMHHFILEVKRFYVAGPTQLY  324
Consensus_aa:        sGl.tLhhphlTAhhphssh..pl.stRs.@lhKY.s.s.+W.Hh.DLthhN.@l.EVpRhYsAtPp.h@
Consensus_ss:        hhhhhhhhhhhhhhhh hhhhhhhhhhhhhh         hhhhh  hhhhhhhhhhhhh   hhhh

Conservation:          6966 6       9     99  9  6 96  6  6 69  9  9 6 99            6
CYP5619D1_      322  ARATKDFSVPSGSGATVAVPKNRLTVVLLDCINNDPKRWPSPEQFQPARF----------AAANTS---  377
CYP5619D2       322  ARATKDFVVTNSS----SVPKHSLTAALLDCLNYNAARWPSPAQFQVARF----------AGANPS---  373
CYP5619A1_      315  ARATKDTDVRTSEG-TFHVPKGCLVAAALDGTNKHPSVWANPTKFDPSRF----------STAKVDM--  370
CYP5619C1       333  GRTTQELTFTTPDG-EFAIPKGVLAVAGLDATNRHPDVWTDPSVFNPDRF----------DNGFSEASD  390
CYP5619B2_      330  GRTKTDLEIPTKNG-VYKLPKGCLAAAGLEATNRHPDVWTDPNLFNPNRFR---------DLGHVRTTK  388
CYP5619B1_      325  GRATDDLEISTANG-SFKVPKGCLATAGLEVTSKHPDVWSDPHTFNPDRFAPQDASTTPVDPDAFKDGAR  393
Consensus_aa:        tRhppDhplso.sG.sh.lPKssLhhA.L-thNpcPs.WssPp.FpPsRF..........s.t.sp...
Consensus_ss:        eeeeeeeee       ee     eeeeehhh    hhh

Conservation:                66966        6 69  6 9 69699   999 99 6 6 9 9   9     69 9 6
CYP5619D1_      378  -------AYGFAPFAIDDL--VHRAEGRREGLSRLILQSHVVSLLDFVAVMAPLQSFALGDG-VNPLPID  437
CYP5619D2       374  -------AYEFAPFALNDL--VDRRAGRREGLSRLILQTHVVSLLDFAAVMAPLQSYALDDG-LNPLPVD  433
CYP5619A1_      371  -------AFGFCPHAIGA---DRRCAG--EELSTLILQSFMVSLFDFMWKMLPHQDYTLDTTLVNPMPKG  428
CYP5619C1       391  -------LYKLCPHAIGKTTGGRKCAG--RDLATLVLQASLVSLFDFKWTLVPNQDLSLEEGKSTPMPKG  451
CYP5619B2_      389  -------PHAFCPHAFGESS-HRRCAG--EDLTTLILQSTVVSLYDFVWQMVPNQDYKLAVGSSTPTPVG  448
CYP5619B1_      394  DVTAPDMMYKFCPHSIGI---ARRCAG--EGLTTLVLQCFVVSLFDFIWQMVPGQNYQLEEKSSTPTPIG  458
Consensus_aa:        ........@.FtP@Ahs.....+RtAG..EsLopLlLQthlVSLhDFhh.MhP.Qs@.L..s.ssPhP.s
Consensus_ss:                eeee                hhhhhhhhhhhhhhhhhhhhhh              ee

Conservation:          96 669             66     69  9  9 99 6    6   6   9 969669669669  6
CYP5619D1_      438  LLTTVSFRYVPG------VVQGD---IDAWRRLHHPSAKLYNGSLENPLLAASDKRLDFWTHSMIQLFNV  498
CYP5619D2       434  LLTTVGFHYAPG------VAHSSNAYDDAWRRLRQPSAKLYNSSIESPL--SSDKRLDFLTHSMIQLLNV  495
CYP5619A1_      429  GLMVVGFHRRTDLSASMVEVAGS---EEDWKFLSLPEAKVYRDDKEALHDMFADERLDLWTHLMLKLLAK  495
CYP5619C1       452  LLMASSFTHRHSESETECDVA-------DWHLLNLPEAKALVGIAGTVSDDEDDARLDLWTRLMIKLIAK  514
CYP5619B2_      449  QLMAVGFHRRTDDAVEIIGTVGS---KADWKFLNLPEAKELV---GTAMDLYDDARLDLWTRLMIKLIGK  512
CYP5619B1_      459  QLMAVGFHRRTLDDVVTFGTAGS---DEDWHFLSLPQAKELVG-SGT-ADLYDDARMDLWTRLMIKLIGK  523
Consensus_aa:        bLhhVtF+..ss......shhGS...c.sW+.LphPpAK.hss...ssh.h.sD.RLDhWT+.MIpLls.
Consensus_ss:               hhhhhhhhh                   hhhhhhhhhhhhhhhhhhhhh   hhhhhhh h
```

```
Conservation:          66  9  9 6   6  9  99 969 99 66 9 66669996   96      66969 696 969   9
CYP5619D1_      499    RFETWVTPTAAASIKVPTTQKNLPKRTLYGTSIQIPTEDEDVA-IPKVILESAKLLQDTAPFVDNFDAKW   567
CYP5619D2       496    RFATWVTPTAAASITVPKSQKPLAKQTLHGTSIQIPVDDEDVS-IPKVLLDGAKLLQDTAPFVDNFDDSW   564
CYP5619A1_      496    KQSMWNKPFANQAITAPKYQKTLPKITLYGLKIQIPTEDEDWPSDPWNEVATVKFLRDSCPLGDDFEHTW   565
CYP5619C1       515    KQARWNKPVANEVLTVPQFQKELPKMTLIQTNIQVATEDEDWPNQPWLEIQQSNFLRDYAPFVDNFEHTW   584
CYP5619B2_      513    KQAVWDRPYANQILRIPQHQKPLPKITLIQTNIDIATEDEDWPNQPWLEIQQSNFLRDHAPFVDNFEHTW   582
CYP5619B1_      524    KQATWDRPFVESCLTIPKHQKVLPKLTLIQTSIEIPTEDEDWPKQPWLEIKQSNFLRDHAPFIDDFKHTW   593
Consensus_aa:          +btWspPAs..lplPphQKsLPKbTLh.TsIpIsTEDEDhs.bPblblpptphLpDhAPFlDsF-coW
Consensus_ss:          hh     hhhhhhh   hhhhhhhhhhhh   hhhhhh             hhh         hhhh

Conservation:           99999966969969 99999699969999999999669996999966 99  996999996  6 66  6
CYP5619D1_      568    APGEDMEGCVLSKVGRMWPRVRVHWDDRYSDRALELLVFNGLGQHMVQKLATAHDDGSYYTVATNYLASI   637
CYP5619D2       565    VPGEDMEGYVLSKVGRMWPRVRVHWDDRYSDRALELFVFNGLGQHMVTKLSAAHSDGSYYTATTSFLETL   634
CYP5619A1_      566    LPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLVFNGLGQHLVTKLRTAHDDGSYYGICLDFMQAL   635
CYP5619C1       585    LPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLAFNGFGQHLLTKLPEAHDDGSYYGICLNFLKGL   654
CYP5619B2_      583    LPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLAFNGFGQHLLTKLPEAHDDGSYYGICLNFMKSL   652
CYP5619B1_      594    LPGEDMERYVMSKLGHMWPRVNVHWNDRYSDRALELLAFNGLGQHLLMKLPEAHDDGSYYGICLDFMNVL   663
Consensus_aa:          lPGEDME.YVhSKVGpMWPRVpVHWsDRYSDRALELLhFNGhGQHhlhKLspAHDDGSYYslthsFhpsL
Consensus_ss:               hhhhhhhhhhhh         hhh   eeeeeee    eeeeehheee           hhhhh

Conservation:          6996999 6999969   9999996966 6 69 6  969696999696  9699699969 9 9969 6
CYP5619D1_      638    EVRTGYAITGADAFFDKNGKVTKIVRLGKTIRPIDASWEYVKMCFRSSLVSKITAVDHLIGLHVTVGNYM   707
CYP5619D2       635    DVRPGYAVTGADAYFDKNGKVTKIVRLGKTFRPADAQWEYVKMCFRSSVANKVTAVDHLIGLHVTVGNYM   704
CYP5619A1_      636    DVRPGYAKYGADAYFNAKGKVTKIVRLGKTVHPGDEDWEYAKLCFRGSLQTKVTALDHLLGIHITVANGL   705
CYP5619C1       655    EVRPGYAKYGADAFFSAEGKVTKIVRGDVTVRPGDDNWAYAKLCFRGSLQTKITAVDHLLGVHATVANIM   724
CYP5619B2_      653    EVRPGYAKYGADAFFTSKGKVTKIIRGDIASRPGDSGWEYAKLCFRGSLQTKVTAVDHLLGIHATVANIM   722
CYP5619B1_      664    EVRPGYAKYGADAYFTAKGKVTKIIRGGVTSRPGEDGWEYAKLCFRGSLQTKVTAVDHLLGIHATVANYM   733
Consensus_aa:          -VRPGYAbhGADA@Fs.pGKVTKIlR.s.h.RPtDssWEYhKhCFRtSL.oKlTAVDHLlGlHhTVtNhM
Consensus_ss:               eeeeeee         eeeee        hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

Conservation:          6   99999969969996999999666996 96  99699969699  66 6966696  66  666969
CYP5619D1_      708    TTGSREQLPPTHPLRRLIKPFTFRAVAINYDASIALFAPKGMLHRAFPYTEKGLKDTWAMALKSLTLEPF   777
CYP5619D2       705    TTASREQLPPTHPLRRLIKPFTFRAVAINYEASKLLFAPKGILHRAHPYSEKGLKDTWAMALQSLKLEPF   774
CYP5619A1_      706    VTSTREQLPPTHPLRRLLKPFTFRSVIINYNASYALFWPKGMLHRAFSLSVEGMQQTWELGLANFKYETF   775
CYP5619C1       725    VIANREQLPPTHPLRRLIKPFTFRSIAINYGAGRALFWPKGMLQRAYALTDKGMKQTWDIGLANFKYETF   794
CYP5619B2_      723    VVANREQLPPTHPLRRLIKPFTFRSVAINYGAGRALFWPKGMLQRAYALTDKGMKQTTQDAPAH------   786
CYP5619B1_      734    VTSIREQLPPAHPVRRLLKPFTFRSVAINFGAGRSLFWPKGMLQRAYALTDKGMKQTWEYGLANFKYETF   803
Consensus_aa:          hhtsREQLPPTHPLRRLlKPFTFRtVAINYsAt.tLFhPKGMLpRA@sho-KGhKpTh..t..phKhEsF
Consensus_ss:          hhhhh       hhhhhhhhhhhhhhhhhhhhhhhh     eeeeee   hhhhhhhhhhh
```

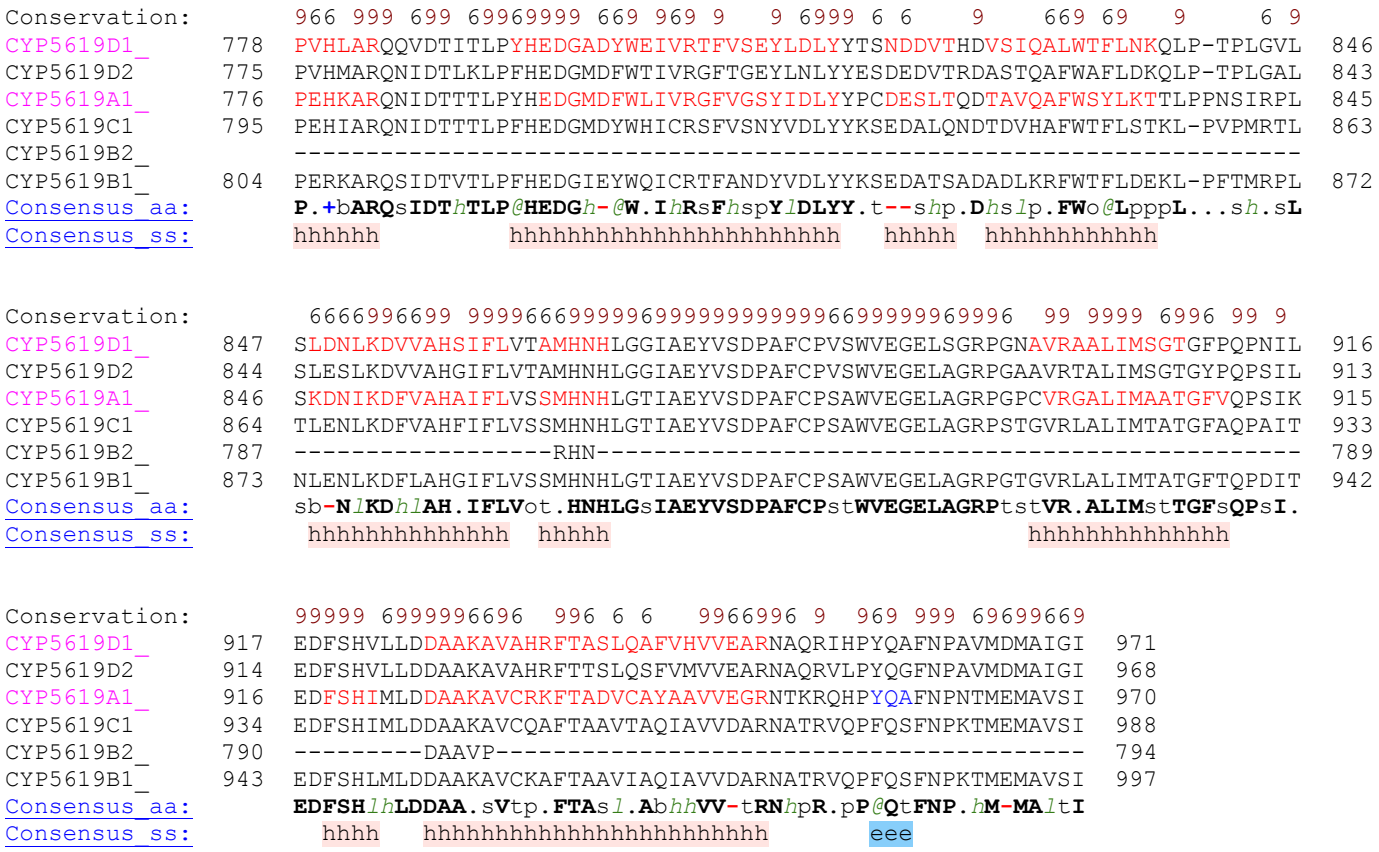


Figure 2.4. Structural alignment of CYP5619 family members from *S. diclina* using PROMALS3D (Pie *et al.*, 2008). P450 characteristic notations for α-helices (shown in red) and β-strands (shown in blue) were mapped. Residues highlighted in green appear in contact with the heme. The absolutely conserved ExxR motif is highlighted in cyan, as well as the amino acids matching the P450 heme signature consensus sequence according to the PROSITE database. The cysteine residue responsible for the co-ordination of heme iron is highlighted in cyan.

2.4. Conclusion

In conclusion, in this chapter phylogenetic and secondary structure analysis of CYP5619 P450 family members was carried out. Phylogenetic analysis revealed the presence of the CYP5619 family in other oomycetes and in phytoplankton. Secondary structure analysis revealed conservation of P450 characteristic motifs in all CYP5619 family members with the exception of CYP5619D members missing Cysteine in the CxG motif. The identified CYP5619 P450s in other oomycetes and phytoplankton will be submitted for naming to International P450 Nomenclature Committee headed by Prof DR Nelson, University of Tennessee, USA.

References

- Brodhun, F., Göbel, C., Hornung, E. and Feussner, I., 2009. Identification of PpoA from Aspergillus nidulans as a fusion protein of a fatty acid heme dioxygenase/peroxidase and a cytochrome P450. *Journal of Biological Chemistry, 284*(18), pp. 11792-11805.
- Chen, J.K., Wang, D.W., Falck, J.R., Capdevila, J. and Harris, R.C., 1999. Transfection of an active cytochrome P450 arachidonic acid epoxygenase indicates that 14, 15-epoxyeicosatrienoic acid functions as an intracellular second messenger in response to epidermal growth factor. *Journal of Biological Chemistry*, *274*(8), pp. 4764-4769.
- Graham, S.E. and Peterson, J.A., 1999. How similar are P450s and what can their differences teach us? *Archives of Biochemistry and Biophysics, 369*(1), pp. 24-29.
- Graham-Lorence, S., Truan, G., Peterson, J.A., Falck, J.R., Wei, S., Helvig, C. and Capdevila, J.H., 1997. An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio-and stereoselective (14S, 15R)-arachidonic acid epoxygenase. *Journal of Biological Chemistry*, *272*(2), pp. 1127-1135.
- Guengerich, F.P. and Munro, A.W., 2013. Unusual cytochrome P450 enzymes and reactions. *Journal of Biological Chemistry*, *288*(24), pp. 17065-17073.
- Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T. and Bozkurt, T.O., 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature*, *461*(7262), pp. 393-398.
- Hannemann, F., Bichet, A., Ewen, K.M. and Bernhardt, R., 2007. Cytochrome P450 systems — biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1770*(3), pp. 330-344.

- Hansen, B.G., Mnich, E., Nielsen, K.F., Nielsen, J.B., Nielsen, M.T., Mortensen, U.H., Larsen, T.O. and Patil, K.R., 2012. Involvement of a natural fusion of a cytochrome P450 and a hydrolase in mycophenolic acid biosynthesis. *Applied and Environmental Microbiology*, *78*(14), pp. 4908-4913.
- Hawkes, D.B., Slessor, K.E., Bernhardt, P.V. and De Voss, J.J., 2010. Cloning, expression and purification of cindoxin, an unusual Fmn-containing cytochrome P450 redox partner. *ChemBioChem*, *11*(8), pp. 1107-1114.
- Jackson, C.J., Lamb, D.C., Marczylo, T.H., Warrilow, A.G., Manning, N.J., Lowe, D.J., Kelly, D.E. and Kelly, S.L., 2002. A novel sterol 14α-demethylase/ferredoxin fusion protein (MCCYP51FX) from Methylococcus capsulatus represents a new class of the cytochrome P450 superfamily. *Journal of Biological Chemistry*, *277*(49), pp. 46959-46965.
- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and Syed, K., 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: Comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. *PloS one*, *9*(9), p. e107209.
- Kelly, S.L. and Kelly, D.E., 2013. Microbial cytochromes P450: Biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Phil. Trans. R. Soc. B*, *368*(1612), p. 20120476.
- Lawson, R.J., von Wachenfeldt, C., Haq, I., Perkins, J. and Munro, A.W., 2004. Expression and characterization of the two flavodoxin proteins of Bacillus subtilis, YkuN and YkuP: Biophysical properties and interactions with cytochrome P450. *BioI. Biochemistry*, *43*(39), pp. 12390-12409.
- 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. and Gwadz, M., 2010.

CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Research, 39*(suppl_1), pp. D225-D229.

- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I. and Lanczycki, C.J., 2014. CDD: NCBI's conserved domain database. *Nucleic Acids Research, 43*(D1), pp. D222-D226.
- Moncrieffe, M.C., Fernandez, M.J., Spiteller, D., Matsumura, H., Gay, N.J., Luisi, B.F. and Leadlay, P.F., 2012. Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *Journal of Molecular Biology, 415*(1), pp.92-101.
- Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C. and Page, C.C., 2002. P450 BM3: The very model of a modern flavocytochrome. *Trends in Biochemical Sciences, 27*(5), pp. 250-257.
- Munro, A.W., Girvan, H.M. and McLean, K.J., 2007. Variations on a (t) heme — Novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Natural Product Reports, 24*(3), pp. 585-609.
- Nei, M. and Kumar, S., 2000. *Molecular evolution and phylogenetics*. Oxford University Press.
- Phillips, A.J., Anderson, V.L., Robertson, E.J., Secombes, C.J. and van West, P., 2008. New insights into animal pathogenic oomycetes. *Trends in Microbiology, 16*(1), pp. 13-19.
- Pie, J., Kim, B.H. and Grishin, N.V., 2008. PROMALS3D: A tool for multiple sequence and structure alignment. *Nucleic Acids Res, 36*(7), pp. 2295-300.
- Rylott, E.L., Jackson, R.G., Sabbadin, F., Seth-Smith, H.M., Edwards, J., Chong, C.S., Strand, S.E., Grogan, G. and Bruce, N.C., 2011. The explosive-degrading cytochrome

P450 XplA: Biochemistry, structural features and prospects for bioremediation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1814*(1), pp. 230-236.

- Saitou, N. and Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), pp. 406-425.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific Reports, 5*, p. 11572.
- Sirim, D., Widmann, M., Wagner, F. and Pleiss, J., 2010. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Structural Biology, 10*(1), p. 34.
- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014a. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One, 9*(1), p. e86683.
- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and 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 and Genomics, 36*(3), pp. 321-333.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*(10), pp. 2731-2739.

- Urlacher, V.B. and Eiben, S., 2006. Cytochrome P450 monooxygenases: Perspectives for synthetic application. *Trends in Biotechnology*, *24*(7), pp. 324-330.
- Warman, A.J., Robinson, J.W., Luciakova, D., Lawrence, A.D., Marshall, K.R., Warren, M.J., Cheesman, M.R., Rigby, S.E., Munro, A.W. and McLean, K.J., 2012. Characterization of Cupriavidus metallidurans CYP116B1 – A thiocarbamate herbicide oxygenating P450–phthalate dioxygenase reductase fusion protein. *The FEBS Journal*, *279*(9), pp. 1675-1693.
- Wass, M.N., Kelley, L.A. and Sternberg, M.J., 2010. 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, *38*(suppl_2), pp. W469-W473.

APPENDIX

CYP5619 family member sequences along with CYP51 of Sacchaaromyces cerevisiae (used as an out group).

>CYP5619A1(SDRG_14280.1)Saprolegnia diclina VS20

MGNLTSTGATHGDVHMDKMVMYLDGDRSAMMDGDLFVLEKALATEKVVGFCGPEALKVFDANLRDGTFVRHGALPSGLNE

LLGAVLPTTDGDAHARKKKLVLAAFSGAQLAAYKPLIRTTIQNEHAKWAAHGASMSLVANAKVLVFKLSLLLILGLEDNY

DNSRELLDTYMLALRNSVRRADPAGVRSRDELIRTMINPALATSHDRVHTGKPKPCALDHLVAAGVLSDDDLRAELFHLL

CMSLGGLECWVANCITAAASSTDVLAQLTAGRDAFITKYPAEADRWSHLGDLGYVNNYIQEVKRTYVAGPSHMYARATKD

TDVRTSEGTFHVPKGCLVAAALDGTNKHPSVWANPTKFDPSRFSTAKVDMAFGFCPHAIGADRRCAGEELSTLILQSFMV

SLFDFMWKMLPHQDYTLDTTLVNPMPKGGLMVVGFHRRTDLSASMVEVAGSEEDWKFLSLPEAKVYRDDKEALHDMFADE

RLDLWTHLMLKLLAKKQSMWNKPFANQAITAPKYQKTLPKITLYGLKIQIPTEDEDWPSDPWNEVATVKFLRDSCPLGDD

FEHTWLPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLVFNGLGQHLVTKLRTAHDDGSYYGICLDFMQALDVRPG

YAKYGADAYFNAKGKVTKIVRLGKTVHPGDEDWEYAKLCFRGSLQTKVTALDHLLGIHITVANGLVTSTREQLPPTHPLR

RLLKPFTFRSVIINYNASYALFWPKGMLHRAFSLSVEGMQQTWELGLANFKYETFPEHKARQNIDTTTLPYHEDGMDFWL

IVRGFVGSYIDLYYPCDESLTQDTAVQAFWSYLKTTLPPNSIRPLSKDNIKDFVAHAIFLVSSMHNHLGTIAEYVSDPAF

CPSAWVEGELAGRPGPCVRGALIMAATGFVQPSIKEDFSHIMLDDAAKAVCRKFTADVCAYAAVVEGRNTKRQHPYQAFN

PNTMEMAVSI

>CYP5619A(XP_012203945.1)Saprolegnia parasitica CBS223.65

MGNLSSTGAMHGDVHMDKMVMYLDSDRSAMMDGDLFVLEKALATEKVVGFCGPEALKVFDANLRNGTFVRHGALPSGLNE

LLGPVLPTTDGDAHARKKKLVLAAFSGAQLAAYSPLIRTTIQKEHAKWAAHGASMSLVANAKVLTYKLSLLLILGLEDNY

DNSRELLDTYMLALRNSVRRADPTGVRSRDELIRTMINPALATSHDRVHAGTPKPCALDYLVAAGCLSDDDLRTELFHLL

CMSLGGLECWVANCITAAASSADVLAQLTAGRDAFMTKYPAEADRWSHLSDLGYVNSYIQEVKRTYVAGPSHMYARATKE

TDVHTSEGTFRVPKGCLVAAALDGTNKHPSVWANPTKFDPSRFSSTKVDMAFGFCPHAIGAVADRRCAGEELSTLILQSF

MVSLFDFMWKMLPHQDYTLDTTLVNPMPKGGLMVVGFHRRTDLSASMVEVAGSEEDWKFLSLPDAKVYRDDKETLHDMFA

DERLDLWTHLMLKLLGKKQSMWNRPFANQAITVPKYQKTLPKITLYGLKIQIPTEDEDWPADPWNEVAMVKFLRDSCPLG

DDFEHTWLPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLVFNGLGQHLVTKLPTAHDDGSYYGICLDFMQALDVR

PGYAKYGADAYFNAKGKVTKIIRLGKTVHPGDDDWEYAKLCFRGSLQTKVTALDHLLGIHITVANGLVTSTREQLPPTHP

LRRLLKPFTFRSVIINYNASYALFWPKGMLHRAFSLSVEGMQQTWELGLANFKYETFPEHKARQNIDTTTLPYHEDGMDF

WLIVRGFVGSYIDLYYPCDESLTQDTAVQAFWSYLKATLPPNSIRPLSKDNIKDFVAHAIFLVSSMHNHLGTIAEYVSDP

AFCPSAWVEGELAGRPGPCVRGALIMAATGFTQPSIKEDFSHIMLDENAKAVCRKFTADVCAYAAVVEARNSKRQHPYQA

FNPNTMEMAVSI

>CYP5619B(1OQR84828.1)Achlya hypogyna

MGSAASSSSTGEVPCLRRAVSLKKMIMFMKDPRTAMMDCRDHYGDFFLVESWLTDEKVMGFCGPEALRAFDAKVAEGLIV

RGGSFAPGVLELLGDILPTIDGEAHATRRASIDTAFASEKIGMYKPKIREIVQREHASWAAHGGSISLAQNSRRMVFDVF

LAVLFGIEGSFDEHRDLLDTFVSAIRKSARKADAAGLEARRRIVEDLVRPAIRDARGRASVGEAKPVAIDALIADGKLGD

NELELELFHALFAGFSGVACLVVNSITAAIEWPEARARVFEARDAFFAKYPTEDDRWSAVELGYIDMYLLEVKRFYVAGP

TQIYGRAKEEVALTTAEGTFTIPKGCLATAGLETTNKHPGVWADPHVFNPDRFAADKTAVAKDGTIDVTADEMAYKFCPH

SIGSARRCLGEGLTSLVLQCAFVSLLDFVWQMVPDQSYALDEKSATPTPTGQLMAVGFRRRSPGTEHSVAGSEADWKFLR

CPEAEALVGGATGDLFGDARLDLWTRLMIKIIGKKQATWNCPPVNSLLTVPKHQTTLPKITLIQTEIEIPTEDEDWPHQS
WFEVQQSNFLRDHAPFIDDFVHKWLPGEDMERYVLSKVGHMWPRVNVHWNDRYSDRALELLAFNGLGQHLLQKLPEAHSD
GSYYGIELDFMQVLEVRPGYAKYGASAYFNQKGKVTKIVRGGSTFVPGDAGWEYAKLCFRGSLQTKVTAVDHLLGIHVTV
ANYMVTSAREQLAPAHPLRRLLKPPRAFRPKGMLQRAYALTTDGMKQTWEYGLSHFKYETFPEHRARQNIDTTTLPFHED
GMDYWNIVRTFVNDYLDLYFKTDTDVTGDVHVNKFWSFLNDKLPFD
>CYP5619A(2OQR84828.1)Achlya hypogyna
MRPLTLENLKDFVAHGIFLVSSMHNHLGTIAEYV
SDPAFCPSAWVEGELAGRPGTAVRLALIMTATGFTQPAITEDFSHIMLDDAAKAVCHSFTKAVTDQIAVVDARNASRVQP
FQSFNPKTMEMAVTSYGDAHMDRMVMYLDADRGAMMDGDLFVLEKALATEKVVGFCGPEALKEFDAKVRDGSFVRQGALP
PGLLELLGPVLPTLDGTAHARKKAAVVAALSSSQLNKYKPLIRSVVQDEHARWAAHGASMSLVAYTKQLVFKLALLVLLG
LEDNYDHQREQLDTYMTALRNSTRRADPAGVTARAQLIAGLLNPALATAHDRVAAHAPKACVLDFLVGQGQLSDDDLRVE
LFHMLCMSLGGLECWATNCITAAASNPAVLKQLTAARDAFMTKHPTEDARWAHFQDLGYVNRYISEVKRVYVAGPSHLYA
RAAKTTDVHTSEGAFTVHAGVLVAAALDGTDKHPSVWPDPTKFNPDRFGAKVDMSYAFCPHAVGAVANRRCPGEELSTLV
LQSFLVSLFDFMWKMVPTQDYTLDTTLVNPMPKGGLMVVGFHRRTDLSASMVEVAGSEADWHFLSLPEASVYRNSSETLH
DVFADERLDVWTHLMLKLVAKKQSKWNRPFANSSITIPKYQKELPKITLFGLKIQVPTEDEDWPADPWVEVAMVKFLRDS
CPFVDNFTDTWLPGEDMERYVMSKVGHMWPRVNVHWNDRYSDRAFELLGFHGLGQHMLTKLPAAHADGSYYTIGLDFMQV
LEVRPGLAKYGADAFFDRNGKVTKIVRHGTTSRPGDDNWEYFKLCFRGSLQTKVTALDHLLGIHITVANQLVTSTREQLP
PTHPLRRILKPFTFRSVIINYNASYALFWPKGMLHRAYSLNEKGMQQTWDFGLANFKYETFPEHKARQNIDTLTLPYHED
GMDYWTIVRKFVSNYLDLYYKCDESLTQDTAVQAFWSYLKSTLPTGAVRPLNRENLKDFVAHAIFLVSSMHNHLGTIAEY
VSDPAFCPSSWVEGELAGRPGTGVRLALIMTATGFTQPAITEDFSHVMLDDAAKKIAKQFTADVTDFIAVVDKRNASRPQ
AYQSFNPKTMEMAVSI
>CYP5619A(OQS07110.1)Thraustotheca clavata
MGNLTSTRPLGHDCANHKMLLDLGPEALKEFDAKLQNGSFVRQGAFPQGLLDLMGPILPTLDGAAHHAKKAAILEALNGV
QVEKYKPVIRSMVQKAHARWSAQGGAMSLVANCKQLAFKLMLVVLLGLENDYDDHRDDHRELLDTYILSLRDSTHRADPD
GVRSRQHLLDDMINPALQTSHERLNQNSLKPCVLDFLVSQKKLSDADLRIELFHLLTMGVGGLECWLANCITAAASSPDV
LKQLTVARDTYLKKYSNEEDRWRRFDDLGYVNWYIQEVKRVYIAGPSHVYARSTAQVDIVTSDGTFRVPKGALVAAALDT
TNKHPKVWTNPGEFNPNRFNKWDETKGMYTFCPHSVGSDRRCPGEQLSTVVLQSFMVSLFDFMWKMIPKQDYSLDTKAVN
PMPRGGLMVVGFHRRTAASDDMVQVAGSEADWKFLSLPEAKVYADSKETMYEMFSDERLDVWTHLMIQLLSKKQERWNRP
YANISIKVPQKQVKLKKVTLDGTKVEIPTEDEDWPSDPWFEVKTVEFLRDSCPMDDDFKYQWVPGEDKERYVMSKVGHMW
PRVLVHWNDRYSDRALELLAFNGMGQHLVQKLEKAHDDGSYYSITLEFMQGIEVRPGYATYGADAFFNSKGKVTKIIRKG
VTYRPKDDGWEYAKLCFRGSLNTRVTAVDHLLGIHLTVANYLVTSSREQLPPNHPLRRLIKPFTFRSVIVNFAASWGLIW
PRAMLQRAFAVSEKGIDTLWKTGLASFKYEPFPEHMERQKVDTISMPFHEDGLDYWYICHTFVSDYLNLYYANDEALTQD
TAVRAFWNFLNEKLPTGVRPLSLANLKDFITHAIVLVSAMHNHLGTLAEYVPDPAFCPSSWVEGEMAGRPGTSVRAALLM
AATGFTQPAITEDISGIMLDDKAKAVCKRFSEALTKQIDVVNERNKHRVQIYQSMNPAVMEMAVSI
>CYP5619B(XP_012203946.1)Saprolegnia parasitica CBS223.65
MGSQTSTPAGAAPSLRRAASLKKMIMFMKDPRTAMMDCRDHYGDVFLMESSLVNEKIMGFCGPEALLAYDTQVQAGKIVR
AGAFPTGVLELLGSVIPTLDGDAHAKRKAALHVAFTPETLDTYKVKIREIIQHEHAAWAARGGSLSLALSCKKLVFHVFS
ATLLGLENVDDEYRELIETFVSSIRKSARKPDATGMDARTQVVEELIRPAIREAKARVAAEKPLPTVVDVLVADGRLSDE
ELGLELFHALFAGLGGVTCLAINSITVCIELPAIREKVSAAREAYLTKYPNEDDRWRHFADLGYMQHFLLEVKRFYVAGP

TQLYGRATDDLEISTADGSFKVPKGCLATAGLEATSKHPDVWSDPHTFNPDRFAPQDASAQPSGSVDPDELKDGARDVTA

PGLMYKFCPHSIGTARRCTGEGLTTLVLQCFVVSLFDFIWQMVPGQNYQLEEKSSTPTPVGQLMAVGFHRRALDRVVTFG

TAGSDEDWHFLSLPQARELVGCGAADLYDDARMDLWTRLMIKLIGKKQAAWDRPFVDSCLKIPKHQKVLPKLTLIQTSIE

IPTEDEDWPKQPWIEIKQSNFLRDHAPFIDDFTHTWLPGEDMERYVMSKLGHMWPRVNVHWNDRYSDRALELLAFNGFGQ

HLLMKLPEAHDDGSYYGICLDFMSVLEVRPGYAKYGADAYFNAKGKATKIVRGGVTSRPGEDGWEYAKLCFRGSLQTKVT

AVDHLLGIHATVANYMVTSIREQLPPAHPVRRLLKPFTFRSVAINFGAGRSLFWPKGMLQRAYALTDKGMKQTWEYGLAN

FKYETFPERKARQNIDTLTLPFHEDGIEYWQICRTFANDYINLYYKSEDAISADADLKRFWTFLDEKLPFAMRPLNLENL

KDFLAHGIFLVSSMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPGTGVRLALIMTATGFTQPAITEDFSHIMLDDDAEAV

CQAFTAAVTAQIAVVDARNATRVQPFQSFNPKTMEMAVSI

>CYP5619B(OQR84819.1)Achlya hypogyna

MGNDASVHAEGHAQALPSSNRATSLLKMIAFSKDPRAGMLDARDHYGDLFLLESKVVSEKIAGFCGPELLEAFDSKLAAG

EIVREGAFPAGILALLGPILSSLDGAAHTSRKAAVLEALSQAKLETYKPSIRAIVQTEHAAWAARGGAISLALLTRNLVF

RIFLQVLYGVEMLDDRHRVALDEFIASIRRSSKAPDPHGVSCRTRILEELIRPAIVKARARIAADAPAPCVLDNLITAAK

LDADALEVEAFHFLFAGFGGVACLATNVLTACATHPGVLPKLLEARAEFVTRYPTEDARFAHLDDLGYVNDFLLEVKRYY

VAGPTTVFGRAAVDLEVKTSNGVYHLPKGCLAAAGLEATNKHPDVWANPHDFNPDRFKDLDMASHAHRFCPHAFGEASHR

RCAGETLTTVILQTIVVSLFDFVWQMVPGQNYALQEGVATPTPVDQLMAVGFHRRTDDAVEFGVAGSQGDWKFLNLPEAK

ALVGGASDLYDDARLDLWTRLMIKLIGKKQAAWDRPYADQILSIPKFQKVLPKITLIQTNIEIATEDEDWPNQPWIEIQQ

SNFLRDHAPFVDNFNAKWVPGEDMERYVLSKVGHMWPRVNVHWNDRYSDRALELLAFNGLGQHLLQKLPEAHSDGSYYGI

ELDFMQVLEVRPGYAKYGASAYFNQKGKVTKIIRAGVTSHPGDKDWEYFKLAFRGSLQTKVTAVDHLLGIHATVANIMVI

ANREQLPPTHPLRRLIKPFTFRSVAINYGAGRALFWPKGMLQRAYALTSNGMKQTWEYGLSHFKYETFPERRARQNIDTT

TLPFHEDGMDYWNIVRTFVNDYLDLYFKTDANVGGDANVVQFWGFLRSKLPADAMRELTLENLKDFVAHFIFLVSSMHNH

LGTIAEYVSDPAFCPSAWVEGELAGRPGTAVRLALIMTATGFTQPAITEDFSHVMLDDAAKKVAKQFTKAVTEQIAVVDA

RNASRVQPFQSFNPKTMEMAVSI

>CYP5619C(OQR84821.1)Achlya hypogyna

MGNSHSVQTSDAPPLPASKRSNSIFSLLNFAKNPNAAMAQGRDTLGDLFLLESAVLSEKIIGFCGPDMLAQYDGQVEAGG

IVRAGALPSGIVELLGPILPVLDGNVHAIRKKFVMAAFTEDQLTAYAPTIFSIVQNEHAAWAAHGGSISLGLLSKKLVFK

VFLAVLFGLTNIPPIEYDTKYDQYRDEVDGFIAGISKSATAPDAHAVACKQRLITELIGPAIVASQARVKAGAPRPCVLD

ALVAGDGLSDAQLRLEGLHMLFAGLGGVQCLVVNSLTVMAKFPDISEKLQEARAAFVTRCPTPADRWRHFDQLGYANQFL

LEVKRFYTAGPTQLFGRTATELTFQTPDGTYSVPKGALAVAGLNATNKHPEVWADPSVFNPDRFANFDTTTDLYTLCPHS

IGKMVGGRRCAGQDLATAVMQASLVSLFDFKWTFAPGQDFTLETGKSTPMPVGNIMVTAFQHRHEVGEDGCDVANWHLLN

MPEAKALAGVAAEVSEDEDDARLDLWTRLMIKLIGKKQSRWNKPVANEVLTIPKSQVTLPKITLIQTDIQVATEDEDWPN

QPWLEIQQSNFLRDYAPFVDNFEHTWLPGEDMERYVMSKVGKMWPRVNVHWNDRYSDRAVELIAFNGFGQHLLTKLPEAH

DDGSYYGIELNFMRTLEVRPGFAKYGANAYFNKKGKVTKIVRGGVTSRPGDATWEYAKLCFRGSLQTKITAVDHLLGIHA

TVANIMVIANREQLPPTHPLRRLIKPFTFRSVAINYGAGRALFWPKGMLQRAYALSTLGMKQTWDYGLSHFKYETFPERR

VRQNIDTVTLPFHEDGMDYWNIVRTFVSNYVDLYYKADSAIANDEHVRKFWSFLDDKLPFDMRPLTLENLKDFVAHGIFL

VSSMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPGNAVRLALIMTATGFAQPAITEDFSQIMLDDAAKAVCKKFTADVTA

FIDVVDTRNLSRPQAYQSFNPKTMEMAVSI

>CYP5619C(XP_012203939.1)Saprolegnia parasiticaCBS223.65

MGNQPSTEAGAAPLPDSKRANSIFSLLAFAKDPKAAMAESRDTLGNLFLIESAVVSEKIAGFCGPEMLSQYDAHVAAGHI

VRENALPAGIVELLGPILATLDGEVHDSRKEAIMGAFSKDMLASYAPIVFGIVQKEHAAWAAHGGKISLALSCKKTVFKV
FLAILYGITDMTPAEYDATYDPFRDLLDGFIRAIPKSSRGADAEGLVCKQRLLDELVAPALAASQARVAAKTPVPCFLDY
MLGQTELTPDVVHLEAFHALFAGLGGTQCLVVNTITALAQYPTVAEKVHASRAKFVTKYHEDRWRHFDNLGYCNRFLLEV
KRFYNAGPAQLFGRTTQELTFTTPDGEFAIPKGVLAVAGLDATNRHPDVWTDPSVFNPDRFDNGFNEATDLYKLCPHAIG
KTTGGRKCAGRDLATLVLQASLVSLFDFKWTLVPNQDLSLEEGKSTPMPKGLLMASAFTHRHSADETECDVSDWHLLNLP
EAKALVGIAGTVSDDEDDARLDLWTRLMIKLIAKKQARWNKPAANEVLTVPRFQRELPKMTLIQTNIQVATEDEDWPNQP
WLEIQQSNFLRDYAPLVDDFEHTWLPGEDMERYVMSKVGHMWPRVNVHWNDRYSDRALELLAFNGFGQHLLMKLPEAHDD
GSYYGICLGFMKGLEVRPGYAKYGADAYFNAEGKVTKIVRGDITARPGDDSWAYAKLCFRGSLQTKITAVDHLLGVHATV
ANIMVIANREQLPPTHPLRRLIKPFTFRSIAINYGAGRALFWPKGMLQRAYALTDKGMKQTWDIGLANFKYETFPEQIAR
QNIDTATLPFHEDGMDYWHICRSFVSNYVDLYFKSEDALQSDTDVHAFWTFLSTKLPVPMRTLTLENLKDFVAHFIFLVS
SMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPGTGVRLALIMTATGFAQPAITEDFSHIMLDDDAKAVCQAFTAAVTAQI
AVVDARNATRVQPFQSFNPKTMEMAVSM

>CYP5619A(OQS03666.1)Thraustotheca clavata

MWNCFSGSGDEAFPSGKVPYITDEQVQHMHGVIYLLEYAVGEENVAVLKGSNLIQQFDTHRQNGNLSRQDALPIGLVDLA
GKTLSTLDNATFAKRQSALLDAFSIEQVAKYQSKIDAIVQSRHSAWAARGGSFSIAVETKKLVFHIFVGVILGLEDQYDA
VFNLVNQYRELLPKSLRRPHAKAITLRQEILSKLITPAVTSSRTRVANKQTNDSVVDYLIRKGQLSDADITIELFQALID
GTDGISSLVINCVNAWVNQPGLSDKLASVRDAPDTFVNQFIDEVERVYTAGPSHEYARVVKNTTFTTPKGSFTLPKGQLV
VAFTESINEDASVWPNPTLFDPSRFENDTPDPYKFTAFSLLQLVNRVQNVREAFTKAVLRSNMSSLLNCMWQMVPLQSYE
LTEHTVTNPTPVGQLTVVNFHKRHEQSANSVATAGTPEDWKFLEQPEAKQYADCSESLDELFADKRLDVWTNLTLKLLER
KQAKWNRPFANSAITVPKYQAELPKIQLYGTNINVPTEDEDWPSNPWIEVKTVEFLRDSCPVNDNFDDMWLPGEDMESYV
MSKVGKIWPRVNVHWNDRYSDRALELLAFNGLGQHMLEKLPDAHDDSSYYGIFMNYMDGLDVRPGYAKYGADAYFDRDGI
ITKIVRQGVTYQPKDAGWEYAKLCFRGSLITKVTAVDHLLGIHVTVANSLVTSSREQLPPNHPLRRLIKPFTFRTVIINH
AASYALFWPKGMLHRAFALSLDGMQQTWEFGLANFKYETFPEHKARQNIDTATLPYHEDGIDYWNIVHNFVSEYLDLYYK
SDDSLLQDDSVVAFWEYLKSTLPKDSIRPLTLVNLKDFIAHSIFLVSSMHNHLGTIAEYVPDPAFCPSAWVEGELAGRPG
TSVRASLIMGATGFAQPSIKEDFSHIMLDDEAKAICKKFTAEVSAFSAVVIDRNTRRKQPYQSFNPDTMEMAVSI

>CYP5619C(1OQR84833.1)Achlya Hypogyna

MGNQPSTETGAPPLPDTKRANSIFSMLAFAKNPREAMAESRDTLGNLFLIESAIVSEKIVGFCGPEMLAQYDAQVEAGGI
VRDGAFPAGIAELLGPILPALDGEIHAARKAAVMTAFSKEQLALYVPLIFGITQKEHAAWAAHGGAISLALLSKKLVFKV
FLAVLYGIETDTPAEYEAKYDHFRDVVDGYIHAIPKSAKAPDADGLRYKARAIDELIAPALAASQARIEAGTPRPCVLDY
WVQHSGMQPDDICLEAFHALFAGLGGVQCLVVNTITAMATNPGAAEKLHHTRAEYVLKYHSAEDRSSHFDQLGYANQFLL
EVKRFYMAGPSQLFGRTTAELTFQTPDGTYSVPKGALAVAGLNATNKHPEVWADPRVFNPDRFADFDADADLYKLCPHAI
GKTNGGRRCAGQDLATAVMQASLVSLFDFKWTFVPGQDFTLETGKSTPMPVGNIMVTAFQHRHEVGEDGCDVANWHLLNM
PEAKALAGVAAQVSDDEDDARLDLWTRLMIKLIGKKQSRWNKPVANEVLTIPKSQVTLPKITLIQTDIQVATEDEDWPNQ
PWLEIQQSNFLRDYAPFVDNFEHTWLPGEDMERYVMSKLGHMWPRVNVHWNDRYSDRALELLAFHGFGQHLLQKLPESHD
DGSYYGIELDFMRTLEVRPGFAKYGADAYFNENGKVTKIVRGGVTSRPGDATWEYAKLCFRGSLQTKITAVDHLLGVHAT
VANIMVMANREQLPPTHPLRRLIKPFTFRSIAINYGAGRALFWPKGMLQRAYALTDKGMKQTWDFGLSHFKYETFPEHIA
RQNIDTTTLPFHEDGMDYWTIVRTFVSNYVDLYYKAEADVENDADLHAFWSYIGSMLPVPMRKLTLENLKDFVAHFIFLV
SSMHNHLGTIAEYVSDPAFCPSSWVEGELAGRPGTAVRLALIMTATGFAQPAITEDFSHIMLDDAAKAVCLAFTKAVTDQ
IAVVDARNASRVQPFQSFNPN

>CYP5619NSF1(2OQR84833.1)Achlya hypogyna

MTSRNKVMALPDNASVSLRDLVADATAMAELRLLLSNQKLMGAYGARLLAVLEEQAYIA

QQPAVQRRPPPATEWTGGLHKLQTFAKDPCGTSLALHAKYGDIFFLDSVWSSTMIAGVAGPSLLMAFDDHWNAGRLGSAV

PSGVLPLLGPVLPTLDGSKHRARKAALLAGVVPATHAVVIANMVADELEAWAAAECTFSFVVRAQSLALKLLLRILLGIT

GASPLLLGNCQHWIDTLVAAVPASTVAPVPQGLLAKERLLADLCRPAVAASRTRFASKAAVTCVLDALVERNELPDEVLA

LELLHCLCTGVAPLGSLLANTVTASHKFPAVWSKLQRSASAYTQSKIDGALWEYGSHFAMEVQRFYSAGSSLRYGRAKTD

LIFTAGEVVYTLPKDSLVVAGVRATHVRAASWAVPHHFNPDRFAAGVEKGAWQPLRLGGFCDALSTRIVEAWALALTNYS

WHLVPGQDFAVDRAVPSSVPAGKLVASHFRRIRPAVAASPAQCLALPSASEYTQLIAVAGDVVSKHDPRLDFWTREMYKL

VVLKLSRWSRPEASKALTIPATHGPIDKITLVQTSIQVPLEDEDWPNQPWIEIKFANGLRDYAPFVDNFAADWLPGEDKE

RYVMQRFGHIWPRVQVHWDDRYSDRALELIAFNGLGQHMITKLPEAHTDGSYYSVATNFMYGLEVRPGFAKYGADAYFDS

NGKVTKIVRGSLTFRPDDPDWEYAKLCFRGSLQIKVTALDHLLLVHSTVANHVTVLHREQLPPAHPLRRLIKPFTFRSAA

INFSAGRALFAPKGMLQRTVALTTAGMKQAWDYGLASFAYEPFPAMIARQNIDTVSLPFHEDGMDYWCIVEQFVDAYIAL

YFHTDVDVTGDDAIVGFWAALNATMPYDLPALSLAALKEFITYFVFTVSSMHNHIGAIAEYVSDPAFCPAAWVEGELAGR

PGTAIRLALLMIVTGFDQPQITEDFSHVMLDDDAKAVARSFTRAVTEQIAVVDARNQKRVQPFQSFNPSTMEMAVGI

>CYP5619B(XP_012203942.1)Saprolegnia parasitica CBS223.65

MGNEASTVHADGAATDLPASQRAMNILKMIEFSKDPRAGMLESRDQYGDLFLLESHLVSEKIAGFCGPELLAAFDDKLRD

GSIVREGAFPPGIVALLGAIMPTIDGEEHHARKAAALEAFTPARLDLYAPLVREIVQAEHASWAARGGAISLACLTREMV

FRIFLKVLYGVERHDANKFRKRPDPHGVSCRTQILDELIRPAIADARARAATKTPAPSVIDCLVTNGKMASDVLETEAFH

FLFAGFGGVACLATNILTAVATHPSARKDLLDARAEYVTKYDGEARWAHFHDLGYVNLFILEVKRFYVAGPTAVFGRAKT

DLEIPTKNGVYKLPKGCLAAAGLEATNRHPDVWTDPNLFNPNRFRDLGHVRTTKPHAFCPHAFGALSHRRCAGEDLTTLI

LQSTIVSLFDFVWQMVPNQDYKLAVGVSTPTPVGQLMAVGFHRRTDDAAEIIGTVGSNADWKFLNLPESKELALWDRPYA

NQILSIPQHQKTLPKITLIQTHIEIATEDEDWPSQPWIEIQQSNFLRDYAPFVDNFELTWLPGEDMERYVMSKVGHMWPR

VNVHWNDRYSDRALELLAFNGFGQHLLMKLPEAHDDGSYYGICLGFVKGLEVRPGYAKYGADVYFTAKGNVTKIVRGDIT

SRPGDAGWEYAKLCFRGSLQTKVTAVDHLLGIHATVANIMVIANREKLPPTHPLRRLIKPFTFRSVAINYGAGRALFWPK

GMLQRAYALTDKGMKQTWDFGLANFKYETFPEHKARQNIDTTTLPFHEDGMDYWQICRSFVSNYVDLYFKSEDALQNDTD

NLKDFVAHFIFLVSSMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPGTGVRLALIMTATGFAQPAITEDFSHIMLDDAGK

AVCQAFTAAVTAQIAVVDARNATRVQPYQSFNPKTMEMAVTVAVDATGSSKRR

>CYP5619C(XP_008879406.1)Aphanomyces invadans

MGASASSFVAKESTSLSTLIAFSKDPRTALLSARDHFGDIFLVESAFVTTRIAGLCGPEALQQFEEKMLEGALVREGAFP

PSILALLGPILVTMDGEVHRTKKNALLRAVSPVQLDVYKPIIRRIIQAEHSKWAAHGGAISFALNTKILVFKILLAVLYG

MEGEFDSFRTYIEDYVAAIKQSAKTTSAHGVTCRATFIAEILEPAIAAAKARQQVNATSSGPLESVLDVLVASGELNDDD

LKNEGFHIMFAGFGGLSAAATNLITAAVVFPEIRAQVFAARDKYLSKFGDDRWGHLDDLGYLNKYILEVKRFFLAGPTQV

YAKAARDVDLVTSKGVFHLSKGSLVMAGLEATNHDPDVWAAPNTFDPSRFNDADIDAAHGVKNTRKYSFCPHGFGDVRNR

RCAGEELSTIVMQSLLVSYFDFTWKMVPGQNYTLQPHSVTAVPIGLLMAMGFQRQQDDGSGNLDYGVVGSHADWKFLRRP

DVQELTGHNAAEYFDDSRLDLWTRLMIKLISKKQSVWNRPYAQSALSLPMEQVVLDKITLIQTQIEIPTVDEDWPSQPWL

EIQQSNLLRDHAPFVDDFDHPWLPAEDGERYVMSKVGHMWPRVNVHWNDRYSDRALELLVFHGLGSHLVQKLPQEHDDGS

YYGLLLNVMQGLEVRPGFAKYGADAFFDKHGHVVKIKRGDQTYTKTDAAWEYVKMCFRGSLQTKVTAVDHLLGVHATAAN

YLVTSSREKLPVNHPLRRLIKPFVFRSVAINYSAGRALFWPNGMLQRAYALTTAGMKSTWEFGLSQFEYATFPDRIARQQ

IDTLTIPFHEDGLDYWNIMIKFVSSYVDLYYPDDASIQHDDDVVAFWSNLTAVSPAPLPDLNKSNLKDFLAEGFFLVSSM

HNHLGTIAEYVSDPAFCPSAWVEGELSARPGNAVRLALIMTATGFTQPSITEDFSHVMLDDAAKAIVRTFTTDVKAQIKV

VDARNATRVQPFQSFNPKTMEMAVSI

>CYP5619C(XP_008878127.1)Aphanomyces invadans

MGNVTGHVQREREYVKTVIGFMKDPRTFMSASRNTYGDVFLFQSSLVNQKIAGLSGPEALQAFEARLADGSLVKTGALPS

GVSDLLGPIMSVLDGEDHHRKKAGIMTAFTPQQLAKYLLVVRRIIQTEHARWAARGGVISITASSKELVFKLLLAVLYGI

EGDFDEYRPLVDEFVASIRKSAVKASPEGKAARDTIMNDLVIPAIEAAKVRVAGGTPSPSALDHLVGLNQLADDDLGVEM

FHVLFAGFGGLSCLATNLVTPLVTMPDVREKILDARDQFLSKYTGDTKWDHLEDLGYINQYILEVKRFFVAGPTQSFAKA

AVAFDVVTSKGTFHIPKGCLVAAGLETTAFDAEVWPNPDNFDPSRFDNNDDLSALQFKLCPHGIGSTSNRRCAGETLTTL

VCQALVVSLFDFTWNMVPGQDYELDENTSIPTPRGGLKAVGFRRRDAVTSYGVAGTDDDWTFLKLPEAKAIVSVHGGWGD

SDGLFADPRLDLWTELMIKLIGKKQAKWNRPYADTALMLPKNKQPLVKLTLAQTSIQVPTEDEDWPTQSWVEVKQANFLR

DHAPFKDDFVHKFLPGEDGERYVMSKVGHMWPRVNVHWNDRYSDRALELLVFNGLGSHLVQKLPTEDPTDGSYYGVLLNF

MQVLDVRPGFAKYGADAFFDKQGKLIKIIRGDKTYTKTDVEWEYVKMCFRGSLQTKVTAVDHLLGIHVTVANYLVTASRE

QLAVNHPLRRLFKPFTFRTVSINFSAGRALFWPNGMLQRAYALTNSGMKQTWEYGLSHFVYAPFPDRVKAQQIDTFTLPF

HQDGLDYWAIVFSFVSKYIDLYFADDAAIAGDTDVVNFWTYVTSVSPVPLPPVSKASLKDFIAQGIFLVSSMHNHLGTIA

EYVSDPAFCPSAWVEGDHAAPPGNAVRLALIMTATGFTQPAITEDFSHVMLDNAAKDLVRTFTADLFKLIDVIDARNTTR

VQPFQSFNPKTMEMAVSI

>CYP5619C(XP_009834503.1)Aphanomyces Astaci

MGAAASNYVYNEATSLSTLIGFSKDPRTALLNARDHYGDIFLVESAFVSTKIAGLCGPEALKEFEAKLQDGSLVKQGAFP

PSILALLGPILVTLDGDVHHAKKAALLKALSPAQLDVYKPIIRRIVQTEHSKWAAHGGAISFAVNTKILVFKVLLAVLYG

VEGEFDTYRRYVDDYVTAIKQSAKVTDEHGVTCRAKFIAEIIAPAIAAAKANQTKRQQQPLNSVLDVLVATGDLTDDDDL

QNEMFHFMFAGFGGVSAAATNLITAVCVFPDIRAKVLRARDDFLRQYDGRDESPWNHLDEMGYLNLFVLEVKRYFVAGPT

QVYAKAARDLDLVTSTGVFRIPEGALVMAGLEATNRDPDTWPSPDSFDPTRFTQADVDGMHMTRPFSFCPHGFGSHRRCA

GEQLTTVIMQSVLVSLFDFTWKMIPGQEYALQPHSVTAVPIGQLMGVNFHRRLNEDDPSTPEVETYGIVGTQDDWKFLRR

PDVQELTGVNAAEYFDDSRLDLWTRLMIQLISKKQTLWNRPYATTALSVPQHQQVLDKITLIQTNIQIPIVDEDWPCQPW

LEIQQTNLLRDHAPFVDDFSHLWLPAEDGERYVMSKVGHMWPRVNVHWNDRYSDRALELLVFHGLGSHLVQKLPQAHADG

SYYGVLLNVMQGLEVRPGFAKYGADAFFNKHGKLVKIQRGDKTYTNTHDDWAYIKMTFRGTLMTKVTAVDHLLGVHVTAA

NYLVTASREKLPVRHPLRRLLKPFTFRSVSINYGAGRALFWPNGMLQRAFALTTAGMKQTWEFGLTQFEYATFPETMAKQ

EIDTLTLPFHQDGLDYWHIVYKFVANYVDLYYPSDDDVAMDVDVGKFWRYMGELSPAPLPDLTKSHLKDFMSQGIFLVSS

MHNHLGTIAEYVSDPAFCPSAWVEGELSARPGNAVRLALIMSATGFTQPSITEDFSHIMLDDKAKALVKTFTADLYAQIK

VVDARNANRVQPFQSFNPKAMEMAVSI

>CYP5619D(XP_012194083.1)Saprolegnia parasitica CBS223.65

MVSIPLLLITIVGQAAGAPQGLGSVLQGVINDVKHSVAGVRYAFESLVDAEPVTGFCSPEALRAFDDALISGALERRTAY

PAGILELSGPTLSTIDGPAFSKRQDAFLNALSGPALAAYQPRIQRRIQEDHAMWAARGSTFSLALHAKTSVFKVFLDVVY

GVSDPEKYTGYRAQLDEYLFYVSKTSRRAPADATKIRERLLAAIVRPAIASSVARVRSGASTTCVLDAVVAQGSVSEADL

VVESFQLLAMGLPGLEGLVVHTITAMVSHDDVRGQMATARDAYTAQYPNGAHWSHLEDLDAVNQYVNEVQRVYGASPRHT

FARATKDLTVPDGSGAMVAVPKNRLTVALLDCINHDPKRWPSPEQFQPARFATANTSAYGFAPFAIDDLVHRVEGRREGL

SRLILQSHVVSLLDFVAVMAPLQSFALGDGVNPLPIDLLTTVSFRYAPGVVQQDVDAWRRLHHPNAKLYNGSLENPLLAA

SDKRLDFWTHSMIQLFNVRFETWVTPTAAASIKVPTTQKTLPKRTLYGTSIQIPTEDEDVAIPKVVLESAKLLQDSAPFV

DNFDAKWAPGEDMEGYVLSKVGRMWPRVRVHWDDRYSDRALELLVFHGLGQHMVQKLATAHDDGSYYTVATNFLASIEVR

AGYAITGADAFFDAKGKVTKIVRLGKTIRPTDAAWEYAKMCFRSSLVSKITAVDHLMGLHVTVGNYMTTASREQLPPAHP
LRRLIKPFTFRAVAINYDASIALFAPKGMLHRAFPFTEKGLKDTWAMALKSLTLEPFPVHLARQQVDTITLPYHEDGADY
WKIVRTFVSEYLDLYYKSDDDVTRDASIQALWAFLNKQLPTPLGVLSLENLKDVVAHSIFLVTAMHNHLGGIAEYVSDPA
FCPVSWVEGELSGRPGNAVRAALIMSGTGFPQPNILEDFSHVLLDDAAKAVAHRFTASLQAFVQVVEARNAQRILPYQAF
NPLVMDMAIGI

>CYP5619D(AIG56338.1)Achlya hypogyna

MVAVSWLWCFGPTLVAAEPQGLGSFFQGIISDIKHAVNDLTFRFEHLVDAEPVTGFCGPDALRAFDNYLATGALVRHDAY
PKGVLDLVGSTLATLDGSAFATRQAAFLNALSPAAVQRYKSTVHNIVQADHATWAARGGTFSLANAAKVTTFKVVLAVVL
GLDNPEAYTGYRSQIDEYLALLAQTEWRAPADAVTIRSRLLAALIRPAVVAAHARATPKSCVVDALVEAGTVSDEDLATE
LFQLLVHGIPGLEGLVVHSLTAIASVDGVRAHLASARDVYMAKYYGAARWDHFDDLGYGNQFLLEVQRTYTASPRQEYAR
ATVDLKVLTPTGTTIVPKNRLKVGVLECLNKDAKRWPNPTSFDPTRFASANTSAYAFAPYAMNLLADRRRGVGEALSQLV
LQTHLVSLWDFAWTMAPRQSYALADSPNPSPVDALTTDGFFVAPGAVVDTEAWRRLHQPDVQLYNASIENPLLAAGDKRL
DFFTHSAIQLFNTRYNLWVKPSASAITVPKVQKVLPKRKLYGTAIQIPTEDEDVDIPKALLEAAKLIQDTAPFVDNFDAK
WLPGEDMEDYVLSKVGHMWPRVRVHWDDRYSDRALELLVFHGLGQHLVTKLPHAHDDGSYYTVALDFLGALEVRSGFAKL
GADGFFTKDGKVTKIVRQGVTYLPGAAKWEYAKLCFRGSLNAKITAVDHLIGLHVTVGNYMTTATREQLPPKHPVRRLLK
PFTFRAVAINYEASNVLFAPKGLLQRAFPLTEKGMAQTWVTALKDLKLETFPQHIARQQVDTMTLPFHHDGTDYWNIVRR
FTSNYLDLYYKDDTAVTSDASLQSFWRTLSAQLPMPLPPLGLAVLKDTTAIGIFLVTAMHNHLGGIAEYVSDPAFCPTAW
VEGEIAGRPGSCVCAAVLMAGTGYLQPNVMEDFSHVLLDDAAKAVARNFTTSLQAFTDIVRSRNAQRLLAYRAFDATIMD
MAIGI

>CYP5619D(AIG56100.1)Achlya hypogyna

MVSITRLLHLSLAAATVAGAPQGFFQDLISDIRHGIAETLFGLEQLVAAEPVVGFCSPEAIRAFDELIAAGALQRQSAYP
KGVQNLVGSTTLTLDGPAFAARQAALLAALSPMAVQTYAPTIRAIVQADHATWAARGGLFSLDDAARTMTFKVFVAVVLG
LESPERYTGYRAQLDDYLSYLRVTASVAPPEAVAIRKRLLDTLVRPAITAARARSTPKPSVVDILVAMGSVADADLADEI
FALLANGLPGLEGLVVHTLSTMASVEGVVANLATARDVYLAKYPGAARWQHLDELGYANQFLLEVQRTYGAKPSHVFARA
TKELSVAGVKVPKNRLTAVLLECLNQDPGRWPEPARFDPSRFAVANTSAYEFAPFAMNLLTDRPHGIREALTTTVLQTHV
VSLFDFVWSMAPHQNYTVEAGVNAGPVDGLMTVSFRAAPGAVVDTEAWRRLTRPYPEAFNSSLDNPLAADPRLDFLTHSL
IQLGNTRFTLWVKPSAATAITIPTTQGVLPKRTLYGTTIQIPTVDEDVKIPKELLEAVKLLQDTAPFVDNFDATWRPGED
MEAYVLSKVGRMWPQVRVHWDDRYSDRALELLVFHGIGQHMVTKLPQPHADGSYYTVALNFMDALEVRAGYAKAGADAFF
TSKGKVTKIVRQGVTYVPGDAGWEYAKLCFRGSVIIKITAVDHLIGLHVTAGNYLTTASREQLPPAHPLRRLLKPFTFRA
AAINYDASSALFAPKGILHRAFALSEKGMAQTWAAAQTMIRLETFPQHIARQGVDSLSLPFHEDGLAYWDIVHSFASDYL
GLYFPSDAAVTGDASVVAFWKALAAVTPLPALSRTALVDATATAIFLVTAMHNHLGGIAEYVSDPAFCPAAWVEGELAGR
PGTSVRSAIIMSGTGYLQPNVMEDFTHVLLDDKAKAVARRFTAALRGLVGVVQSRNAKRVLPYRGFDPEIIDMAIGI

>CYP5619D(AIG56283.1)Achlya hypogyna

MVSITRLLRLSLAAATVAGAPQGFFQDLISDIRHGIAETLFGLEQLVAAEPVVGFCSPEAIRAFDELIAAGALQRQSAYR
KGVQNLVGSTTLTLDGPAFAARQAALLAALSPAAVQTYAPTIRAIVQADHATWAARGGLFSLADAARTMTFKVFVAVVLG
LESPERYTGYRAQLDDYLSYLRVTASVAPPEAVAIRKRLLDTLVRPAITAARARSTPKPSVVDSLVAMGSVAEADLADEI
FALLANGLPGLEGLVVHTLTTMASVEGVVANLATARDVYLAKYPGAARWQHLDELGYANQFLLEVQRTYGAKPSHAFARA
TKELSVAGAKVPKNRLTAVLLECLNQDPGRWPEPGRFDPSRFAIANTSAYAFAPFAMNLLTDRPHGMREALTTTVLQTHV
VSLFDFVWSMAPHQNYTVEAGVNAGPVDGLMTVGFRAAPGAVVDTEAWRRLTRPYPEAFNSSLDNPLAADPRLDFLTHSL

IQLINTRFNLWVKPSAATAITIPTTQGVLPKRTLYGTTIEIPTVDEDVTIPKALLEAGKLLQDTAPFVDNFDAKWRPGED

MEAYVLSKVGHMWPQVRVHWDDRYSDRALELLVFQGLGQHMVTKLPQPHADGSYYTVALNFMDALEVRAGYAKAGADAFF

TSKGKVTKIVRQGVTYVPGDAGWEYAKLCFRGSVNIKITAVDHLIGLHVTAGNYLTTASREQLPPAHPLRRLLKPFTFRA

AAINYEASNSLFAPKSVLHRAFAFSEKGMAQAWAAAQSMIRLETFPQHIARQGVDSLSLPFHEDGLAYWDIVHSFASDYL

GLYFPSDAAVTGDASVVAFWKALAAVTPLPALSRTALVDATATAIFLVTAMHNHLGGIAEYASDPAFCPTAWVEGELAGR

PGTSVRSAIIMAGTGYLQPNVMEDFTHVLLDDKAKAVARRFTAALRSLVGVVQSRNAKRVLPYRGFDPEIIDMAIGI

>CYP5619NSF1(XP_012203940.1)Saprolegnia parasitica CBS223.65

MGNQPSGRTKVMALPPPDAKLHDLATDPVIMADLRKLLSNRSIAGAYGPLLLAAIEEHVGATPQPVAMVQRRPAPSTEWS

GGLSKLRAFAAAPVASFEALHATYGDLFYIESVWTSDKIAGVAGPTLVAAFEDHMDACRLARSVPSGVTHLLGPVLATLN

GPSYKARWTNLASAFAPGHQFEPVVQRLFRDELAAAHAAGRTFSFTVLAQHLVLKLLLSLLLGVTASSQLELANVQHWID

TMVAALPRSTVAPHNDALQAKEQLLATLLQPALLASRRRVDAKAPVACVLDNLVLKNDLSDDVILLELLHALYTGAGPLA

ALLANTISASHAYPAVWAKLVADTRAHKQQSPGAWKFGRAFAKEIQRFYRVGSGLRFARATSDITFAVNDVVYTVPKHTV

VVAGIDATHKHAASWSAPADFIPNRFLDDAESTKNALHLFQLGGVSDALPTMVLESWLLAVADYTWFLTPGQETSLDKAS

VTSPLPVGKLIASHMERRVGVSAIPDASALVRLPTTDEYAALIAVANEQLLTRDPRLDFWTHQMYKLVLIKLSRWTRPEA

AKALTIPATMGPVDKMTLAQTNIQVPLDDEDWPNQPWIEIKFANAIRDYAPFIDNFDENWLPGEDKERYVMRFYAHIWPR

IQVHWNDRYSDRALELMAFNGLGQHMLQKLPTTHSDGSYYTIATNFMQALDVRKGYAKYGADVFFDDKCKVTKIVRGNIT

YRPDDAEWEYVKMCFRGSLQTKVTAIDHLLLIHSTIANHVTVVHREQLPPTHPLRRLLKPFTFRSAAINYGAGRALFWPQ

GMLQRAIALTTRGMKQAWDIGLGSFGYETFPALVERQQIDTTTLPLHEDGIDYWHIVSRFVSSYLDLYYAADAEVTADAS

VVAFWTMLDATLPFALPPLSLKSLHEFVTYFIFMVSSMHNHVGAIAEYVSDPAFCPSAWVEGELAGRPGTSVRLALIMII

TGFDQPQITEDFSHVMLDDAAKCVARAFTTDVKAQIPVVNRRNATRVQSFQSFNPSTMEMAVGI

>CYP5619A(OQS07119.1)Thraustotheca clavata

KDCVVDYLVQQAQITDADITIELFQALIDGTDGISSLIINCVTAWVKQPGMSDKLASIRDSPDAFVDQFINEVERVYTAG

PNHEYARVLTKTTFTTPKSSFSLTKGQLVVVFTESINEDVTVWSNPTSFNPSRFENGTPEPYKFTSFNLLQLVNRAQGVR

EEFTKAVLRSNMLSLLTCMWQMVPLQSYELSEHTVSNPTPVGQLMAVSYHKRHGLSANSVTTAGNPQDWKFLEQPEAKEY

RQSVESLGEAFEDCRLDFWTHAMIQAVKNRSLVWRQPTARAEITLPKYQKVLEKVTLSGTNIEVPVEDEDTGNDLNFAQA

HTFNLLRDLAPLIDNMDATWLPGEDMEGYVMGKVGMMWPRVNVHWNDRYSDRALELIAFNGVGQHLLTKLPEAHEDGSYY

TIALEFMYGLAVREGFANYGGDAFFTQEGKVVKIVYGGEEYLSDNEQWEHIKMAFRGSLLARVTALDHLLGTHVTVANYL

TTASREQLPPDHPLRRLIKPFTFRSVAINYSAASVLFWPKGMVDRAFAFTHESLENVWAYGLKHFSYEPFPEFVANQKID

TVELPFHQDGMDYWTICHAFVSKYVDLYFSSEEQLIRDAAVASFWQFLVEKVPVPKFPTLSLDNLKNFLAHGIFLVSAIH

NHVGSIAEYVSDPAFCPSAWVKGELAGRPPTCVRAALIMAAAGLPQPSILEDFSHVMLDDDAKSICQDFTAALVKHQNVV

DERNAKRVQPFQSFNPKMMEIAVSI

>NF1SF1(XP_005786468.1)Emiliania huxleyi CCMP1516

MELGLEQSCLVSKNKLRWQTAIARLIGQRDICSTRICSTSAAHAVFARLDLALRSSLVDTAGADLEAGLPTAARLALLAA

FEVPVAVATGMDPAAHLDVYVRASGRQPFLFGKGIAVPGYDDVSTLVSSPQQERRAMVLAHPVLIADPVPPACMGGGTLI

YLSTGAKHTALRRAIGRAVTGFALKRGRGPPLLFPRGAAPAEWDSRAVRETAPLLGAALDSQANATLETRLALKAAVLAS

PLGGRLRKANRADKLDADELAQQVADGLLFAGGYGTTHLTLAALERISSDPALYADPDAFLVESARLDPPVTSVSAIAPK

GGQQLQGPGGGQLRVAQGVPMQLLLSHANRDPAVFERPYAFDPSRRNLDKVLSWNGVDAAGSCDYSASDRPSEDLELRNE

HDDEAMYAYYGAGLVNIVLGVLALIVLGWFCLSQRLAFGALYVKGDAGIIRVQSWLGLLHYLGQAIAAFSWLRLHAVVTH

ADDARARAESASALRLCFGVMAIFVGAFAVFGTAAACVLFVPAAAKSWIAIAMFWWRHGGVLVLAACAFIAFWVGQVIAE

QRGMSLIDYEAGSGIGVTLLARGLLGAEAAPHAWKYEVFARGFLAPSVYFAAAAHIDRVQGRGLGRAYPSELFMRRALRF
KHSAPVLLALALSAIHSFAIPRSFGDITGCGEGGGEAAGPECAVDPTGGLDQYTKVYFSIIHLLDDGSQPGPSFVQAPNR
TVQPLPKEQVVPGLVLPSYDEDEGLVTSRALANQAFSGYVKDGRLYPLEDLDLPWPEKDGAIEALMGRLSGSLAPAELYD
YDANIGGDSLIGDAGNPDFPPPASRRWFRTQFPRDADFTYLAPLTVRPGFERYGAKATFDAAARPVSIWWSHGEKEVRPD
DAAWTHAKFAFRSSLLTGVTLKDHLAATHLTIAATLVSASRDHLPATHPLRRLLKPFTYRTIALSLLHHTVALDAAGLEA
GFAFAFNRTRDTFDPANPARFLEYPLEYPLSSAAECAATPPSSQAACADEADELAAFAADGTRYRAAVREYVGRYVGIYY
ADDAAVGSDVVLASFWAALVAHFPRIPPLSSREALVEVLTGFVFHVTAGHRHVGAAYSAVKDPRYAGAKIRPGRDMSDVQ
AAVQVLAIALVTGFKQPMLLGDYSHVFLRDGHRNATRALWSDFQAKLLQVSAEVDERNRGPRRFKVRAFDPREMATSVSI

>NF1SF1(XP_005778763.1)Emiliania huxleyi CCMP1516

MRATRHLLDAHLLDGTLAVSRARQRRRSHMPSCCFSALVISGTSAILLAGALSSLDGALKFARLDLALRSSLVDTAGADL
EAGLPTAARLALLAAFEVPVAVATGMDPAAHLDVYVRASGRQPFLFGKGIAVPGYDDVSTLVSSPQQERRAMVLAHPVLI
ADPVPPACMGGGTLIYLSTGAKHTALRRAIGRAVTGFALKRGRGPPLLFPRGAAPAEWDSRAVRETAPLLGAALDSQANA
TLETRLALKAAVLASPLGGRLRKANRADKLDADELAQQVADGLLFAGGYGTTHLTLAALERISSDPALYADPDAFLVESA
RLDPPVTSVSAIAPKGGQQLQGPGGGQLRVAQGVPMQLLLSHANRDPAVFERPYAFDPSRRNLDKVLSWNGVDAAGSCDY
SASDRPSEDLELRNEHDDEAMYAYYGAGLVNIVLGVLALIVLGWFCLSQRLAFGALYVKGDAGIIRVQSWLGLLHYLGQA
IAAFSWLRLHAVVTHADDARARAESASALRLCFGVMAIFVGAFAVFGTAAACVLFVPAAAKSWIAIAMFWWRHGGVLVLA
ACAFIAFWVGQVIAEQRGMSLIDYEAGSGIGVTLLARGLLGAEAAPHAWKYEVFARGFLAPSVYFAAAAHIDRVQGRGLG
RAYPSELFMRRALRFKHSAPVLLALALAAIHSFAIPRSFGDITGCGEGGGEAAGPECAVDPTGGLDQYTKVYFSIIHLLD
DGSQPGPSFVQAPNRTVQPLPKEQVVPGLVLPSYDEDEGLVTSRALANQAFSGYVKDGRLYPLEDLDLPWPEKDGAIEAL
MGRLSGSLAPAELYDYDANIGGDSLIGDAGNPDFPPPASRRWFRTHSLEMCARHADFTYLAPLTVRPGFERYGAKATFDA
AARPVSIWWSHGEKEVRPDDAAWTHAKFAFRSSLLTGVTLKDHLAATHLTIAATLVSASRDHLPATHPLRRLLKPFTYRT
IALSLLHHTVALDAAGLEAGFAFAFNRTRDTFDPANPARFLEYPLEYPLSSAAECAATPPSSQAACADEADELAAFAADG
TRYRAAVREYVGRYVGIYYADDAAVGSDVVLASFWAALVAHFPRIPPLSSREALVEVLTGFVFHVTAGHRHVGAAYSAVK
DPRYAGAKIRPGRDMSDVQAAVQVLAIALVTGFKQPMLLGDYSHVFLRDGHRNATRALWSDFQAKLLQVSAEVDERNRGP
RRFKVRAFDPREMATSVSI

>CYP5619B1(SDRG_14281.1)Saprolegnia diclina VS20

MGSQASTPAGAAPSLRRAASLKKMIMFMKDPRTAMMDCRDHYGDVFLMESSLVNEKIMGFCGPEALLAYDTQVKEGKIVR
ASAFPTGILELLGAVVSTLDDDAHAKRKAALLVAFTPEKLDAYKPKIREIIQHDHAAWAARGGSLSLALSCKKMVFHVFM
ATLLGLENVDDEYRELVEAFVSSIRKSARKPDTTGMDARTQVVEELIRPAVREAKARVAAQKPLPTVVEVLVADGRLSDE
ELNLELFHALFAGLGGVTCLVINAVTACIELPAIREKVSAAREAFLAKYPNEDDRWSHFADLGYMHHFILEVKRFYVAGP
TQLYGRATDDLEISTANGSFKVPKGCLATAGLEVTSKHPDVWSDPHTFNPDRFAPQDASTTPVDPDAFKDGARDVTAPDM
MYKFCPHSIGIARRCAGEGLTTLVLQCFVVSLFDFIWQMVPGQNYQLEEKSSTPTPIGQLMAVGFHRRTLDDVVTFGTAG
SDEDWHFLSLPQAKELVGSGTADLYDDARMDLWTRLMIKLIGKKQATWDRPFVESCLTIPKHQKVLPKLTLIQTSIEIPT
EDEDWPKQPWLEIKQSNFLRDHAPFIDDFKHTWLPGEDMERYVMSKLGHMWPRVNVHWNDRYSDRALELLAFNGLGQHLL
MKLPEAHDDGSYYGICLDFMNVLEVRPGYAKYGADAYFTAKGKVTKIIRGGVTSRPGEDGWEYAKLCFRGSLQTKVTAVD
HLLGIHATVANYMVTSIREQLPPAHPVRRLLKPFTFRSVAINFGAGRSLFWPKGMLQRAYALTDKGMKQTWEYGLANFKY
ETFPERKARQSIDTVTLPFHEDGIEYWQICRTFANDYVDLYYKSEDATSADADLKRFWTFLDEKLPFTMRPLNLENLKDF
LAHGIFLVSSMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPGTGVRLALIMTATGFTQPDITEDFSHLMLDDAAKAVCKA
FTAAVIAQIAVVDARNATRVQPFQSFNPKTMEMAVSI

>CYP5619B2(SDRG_14277.1)Saprolegnia diclina VS20

MGNEASTVHADGAATDLPASHRAMNILKMIEFSKDPRAGMLESRDQFGDLFLLESHLVSEKIAGFCGPELLAAFDDKLRD

GSIVREGAFPPGVLALLGPIMSTIDGEEHDARKAAALEALTPARLDLYAPIIREIVEAEHASWAARGGAISLACLTRDMV

FRIFLKVLYGVERHDGNKFRVLLDDFIVSIRRSSKHADPHGVRCRTQILDELIRPAIANAQARASNKTPVPSVIDCLVAN

GKMTPDVLETEAFHFLFAGFGGVACLATNILTAVATHPSARKDLLDARAEYVTKYDGDARWAHFHDLGYVNLFILEVKRF

YVAGPTAVFGRTKTDLEIPTKNGVYKLPKGCLAAAGLEATNRHPDVWTDPNLFNPNRFRDLGHVRTTKPHAFCPHAFGES

SHRRCAGEDLTTLILQSTVVSLYDFVWQMVPNQDYKLAVGSSTPTPVGQLMAVGFHRRTDDAVEIIGTVGSKADWKFLNL

PEAKELVGTAMDLYDDARLDLWTRLMIKLIGKKQAVWDRPYANQILRIPQHQKPLPKITLIQTNIDIATEDEDWPNQPWL

EIQQSNFLRDHAPFVDNFEHTWLPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLAFNGFGQHLLTKLPEAHDDGS

YYGICLNFMKSLEVRPGYAKYGADAFFTSKGKVTKIIRGDIASRPGDSGWEYAKLCFRGSLQTKVTAVDHLLGIHATVAN

IMVVANREQLPPTHPLRRLIKPFTFRSVAINYGAGRALFWPKGMLQRAYALTDKGMKQTTQDAPAHRHNDAAVP

>CYP5619C1(SDRG_14273.1)Saprolegnia diclina VS20

MGNQPSTEAGVAPLPDSKRANSIFSLLAFAKDPKAAMAESRDTLGNLFLIESAVVSEKIAGFCGPEMLSQYDAHVAAGHI

VRENALPAGIVELLGPILATLDGDVHDSRKEAIMGAFSKEMLASYAPIVFEIVQKEHAAWAAHGGEISLALSCKKTVFKV

FLAILYGITNLTPAEYDAKFDPFRDLLDSFIRAIPKSSKGADAEGLVCKQRLLDELVAPALAASQARVEAKAPVPCFLDY

MLGQTELTPDVVHLEAFHALFAGLGGTQCLVVNTITALAQYPTVAEKVHASRAKFVIKYHDDRWRHFDNLGYCNRFLLEV

KRFYSAGPAQLFGRTTQELTFTTPDGEFAIPKGVLAVAGLDATNRHPDVWTDPSVFNPDRFDNGFSEASDLYKLCPHAIG

KTTGGRKCAGRDLATLVLQASLVSLFDFKWTLVPNQDLSLEEGKSTPMPKGLLMASSFTHRHSESETECDVADWHLLNLP

EAKALVGIAGTVSDDEDDARLDLWTRLMIKLIAKKQARWNKPVANEVLTVPQFQKELPKMTLIQTNIQVATEDEDWPNQP

WLEIQQSNFLRDYAPFVDNFEHTWLPGEDMERYVMSKVGSMWPRVNVHWNDRYSDRALELLAFNGFGQHLLTKLPEAHDD

GSYYGICLNFLKGLEVRPGYAKYGADAFFSAEGKVTKIVRGDVTVRPGDDNWAYAKLCFRGSLQTKITAVDHLLGVHATV

ANIMVIANREQLPPTHPLRRLIKPFTFRSIAINYGAGRALFWPKGMLQRAYALTDKGMKQTWDIGLANFKYETFPEHIAR

QNIDTTTLPFHEDGMDYWHICRSFVSNYVDLYYKSEDALQNDTDVHAFWTFLSTKLPVPMRTLTLENLKDFVAHFIFLVS

SMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPSTGVRLALIMTATGFAQPAITEDFSHIMLDDAAKAVCQAFTAAVTAQI

AVVDARNATRVQPFQSFNPKTMEMAVSI

>CYP5619D1(SDRG_03324.1)Saprolegnia diclina VS20

MVSLPLLVIVIVGQVAGAPQGLGSVLQGVINDVKHSVAGVRYVFESLVDAEPVTGFCSPEALRAFDDALASGALERRTAY

PTGILELTGPTLSTIDGPAFLKRQDAFLNALSGAALSTYQPRIQRRIQEDHATWAARGSTFSLALYAKTSTFKVFLDVVY

GIDDPEKYTGHRAQLDEYLFYLSKTSSRAPSDAAKIREHLLAAIVRPAIASSLARVRSGAPLTCVLDTVVAQGTVSEADL

ALESFQLLAMGLPGLEGLVVHTITAMVSLDDVRGQMATARDAYTAKYPGGAFWSHLDDLDAVNQYVNEVQRVCGASPRHT

FARATKDFSVPSGSGATVAVPKNRLTVVLLDCINNDPKRWPSPEQFQPARFAAANTSAYGFAPFAIDDLVHRAEGRREGL

SRLILQSHVVSLLDFVAVMAPLQSFALGDGVNPLPIDLLTTVSFRYVPGVVQGDIDAWRRLHHPSAKLYNGSLENPLLAA

SDKRLDFWTHSMIQLFNVRFETWVTPTAAASIKVPTTQKNLPKRTLYGTSIQIPTEDEDVAIPKVILESAKLLQDTAPFV

DNFDAKWAPGEDMEGCVLSKVGRMWPRVRVHWDDRYSDRALELLVFNGLGQHMVQKLATAHDDGSYYTVATNYLASIEVR

TGYAITGADAFFDKNGKVTKIVRLGKTIRPIDASWEYVKMCFRSSLVSKITAVDHLIGLHVTVGNYMTTGSREQLPPTHP

LRRLIKPFTFRAVAINYDASIALFAPKGMLHRAFPYTEKGLKDTWAMALKSLTLEPFPVHLARQQVDTITLPYHEDGADY

WEIVRTFVSEYLDLYYTSNDDVTHDVSIQALWTFLNKQLPTPLGVLSLDNLKDVVAHSIFLVTAMHNHLGGIAEYVSDPA

FCPVSWVEGELSGRPGNAVRAALIMSGTGFPQPNILEDFSHVLLDDAAKAVAHRFTASLQAFVHVVEARNAQRIHPYQAF

NPAVMDMAIGI

>CYP5619D2(SDRG_14279.1)Saprolegnia diclina VS20

MVALVPLLLTAVGVVGTQENTLKGFFQGVISDIKHAVTDVRYTFESLVNAEPVTGFCSPDALRAFDDALSSGALSRHAAY

PAGLLDLSGPTLSTLDGQAFAIRQESLLNALSGPSLAAYQPRIQQLIQDDHATWAARGGTFSLALQAKTTTFKVFLAVVY

GVTQPDEYVGYRAQLDEYLEYAKKTLSRAPSDAIKIRDRLLATLVRPAIAASHARVRAGAAPTCVLDALVAQNTMSDSDL

ATEGFQLMAMGLLGLEGLVVHTITAMVSVDGVRGQLGSARDAYVSKYPNGAHWRHLDDLSVVNAYVNEVQRVYNASPRHT

FARATKDFVVTNSSSVPKHSLTAALLDCLNYNAARWPSPAQFQVARFAGANPSAYEFAPFALNDLVDRRAGRREGLSRLI

LQTHVVSLLDFAAVMAPLQSYALDDGLNPLPVDLLTTVGFHYAPGVAHSSNAYDDAWRRLRQPSAKLYNSSIESPLSSDK

RLDFLTHSMIQLLNVRFATWVTPTAAASITVPKSQKPLAKQTLHGTSIQIPVDDEDVSIPKVLLDGAKLLQDTAPFVDNF

DDSWVPGEDMEGYVLSKVGRMWPRVRVHWDDRYSDRALELFVFNGLGQHMVTKLSAAHSDGSYYTATTSFLETLDVRPGY

AVTGADAYFDKNGKVTKIVRLGKTFRPADAQWEYVKMCFRSSVANKVTAVDHLIGLHVTVGNYMTTASREQLPPTHPLRR

LIKPFTFRAVAINYEASKLLFAPKGILHRAHPYSEKGLKDTWAMALQSLKLEPFPVHMARQNIDTLKLPFHEDGMDFWTI

VRGFTGEYLNLYYESDEDVTRDASTQAFWAFLDKQLPTPLGALSLESLKDVVAHGIFLVTAMHNHLGGIAEYVSDPAFCP

VSWVEGELAGRPGAAVRTALIMSGTGYPQPSILEDFSHVLLDDAAKAVAHRFTTSLQSFVMVVEARNAQRVLPYQGFNPA

VMDMAIGI

>CYP51F1(NP_011871.1)S. cerevisiae S288C

MSATKSIVGEALEYVNIGLSHFLALPLAQRISLIIIIPFIYNIVWQLLYSLRKDRPPLVFYWIPWVGSAVVYGMKPYEFF

EECQKKYGDIFSFVLLGRVMTVYLGPKGHEFVFNAKLADVSAEAAYAHLTTPVFGKGVIYDCPNSRLMEQKKFVKGALTK

EAFKSYVPLIAEEVYKYFRDSKNFRLNERTTGTIDVMVTQPEMTIFTASRSLLGKEMRAKLDTDFAYLYSDLDKGFTPIN

FVFPNLPLEHYRKRDHAQKAISGTYMSLIKERRKNNDIQDRDLIDSLMKNSTYKDGVKMTDQEIANLLIGVLMGGQHTSA

ATSAWILLHLAERPDVQQELYEEQMRVLDGGKKELTYDLLQEMPLLNQTIKETLRMHHPLHSLFRKVMKDMHVPNTSYVI

PAGYHVLVSPGYTHLRDEYFPNAHQFNIHRWNKDSASSYSVGEEVDYGFGAISKGVSSPYLPFGGGRHRCIGEHFAYCQL

GVLMSIFIRTLKWHYPEGKTVPPPDFTSMVTLPTGPAKIIWEKRNPEQKI

CHAPTER 3

IN SILICO STRUCTURAL AND FUNCTIONAL ANALYSIS OF NOVEL P450-FUSION PROTEIN CYP5619A1 FROM *SAPROLEGNIA DICLINA*

3.1. Introduction

Cytochrome P450 monooxygenases (CYPs or P450s) form a very divergent family, with highly variable sequences. They catalyse a wide variety of oxidative reactions and are therefore of great relevance in drug development and biotechnological applications (Urlacher & Eiben, 2006; Bernhardt, 2006, 2013). Despite their differences in sequence and substrate specificity, the structures of P450s are highly similar. They all share a common fold and conserved catalytic machinery (Graham & Peterson, 1999; Sirim *et al.*, 2010). Genome-wide identification and annotation of P450s revealed the presence of a moderate number of P450s in the genome of 13 oomycetes belonging to two different classes and three different orders and led to the discovery of a novel P450 family, CYP5619 (Sello *et al.*, 2015). The novel P450 family members are fusion proteins found in *Saprolegnia diclina*, a fish pathogen, with an N-terminal P450 domain fused to a heme-dioxygenase/peroxidase domain at the C-terminal. In order to use this novel P450 as drug target, it is necessary to predict its structure, which would consequently enable researchers to screen for potential inhibitors.

Protein structure prediction is the evaluation of a protein that results in the elaboration of a three-dimensional structure of that protein, based on its amino acid sequence. In other words, one predicts the folding and the secondary, tertiary, and quaternary (if applicable) structures of the protein from its primary structure. This is one of the most important goals pursued in bioinformatics and theoretical chemistry and it is fundamentally different from the inverse problem of protein design. The importance of protein structure prediction tools is well demonstrated in medicine and pharmacy (for example, in drug design) and in biotechnology

(for example, in the design of novel enzymes) (Dunbrack *et al.*, 2000). Proteins are chains of amino acids linked by peptide bonds. The chain's ability to rotate around each carbon-α atom allows a multitude of conformations. Predicting the structure of the protein occurs through determining secondary structures and assigning them as α-helices, β-strands, and turns (loops and coils), which form the overall three-dimensional configuration of the chain. This is achieved by using a set of techniques in bioinformatics (Dubey, 2014).

The bends and folds of the secondary structures of the polypeptide chain determine the tertiary structure of the protein, which is necessary for it to assume its function. In some cases where there are more than one polypeptide, a quaternary structure is required in order for the proteins to assume their biological function. Examples of this type of conformation include those of haemoglobin, deoxyribonucleic acid (DNA) polymerase and ion channels (Clugston, 2000; Dubey, 2014). A set of methods is followed for the prediction of the tertiary structure of proteins, including *ab initio* or *de-novo* protein modeling (Hardin *et al.*, 2002), comparative protein modeling (Sánchez & Šali, 2002), side-chain geometry prediction (Keating *et al.*, 2001) and statistical prediction of structural classes (Metfessel *et al.*, 1993). Nonetheless, protein structure prediction remains an extremely difficult and unresolved undertaking, the two main problems being calculating the protein free energy and finding the global minimum of that energy. A protein structure prediction method must explore the space of possible protein structures, which is astronomically large (Perdomo-Ortiz *et al.*, 2012). These problems must explicitly be resolved when using the *ab initio* or *de novo* protein structure prediction. On the other hand, they can be partially bypassed in comparative or homology modeling and fold recognition methods, in which the search space is pruned by the assumption that the protein in question adopts a structure that is close to the experimentally determined structure of another homologous protein (Fiser, 2010). The progress and challenges in protein structure prediction have been reviewed by Zhang (2008).

The method used in this study is that of homology modeling, and it is based on the observation that the protein tertiary structure is better conserved than the amino acid sequence (Marti-Renom *et al.*, 2000). Because a protein's fold is more evolutionarily conserved than its amino acid sequence, a target sequence can be modeled with reasonable accuracy on a very distantly related template, provided that the relationship between target and template can be discerned through sequence alignment. The procedure can be fragmented into four major steps, namely template selection, target-template alignment, model construction and model assessment.

This chapter aims to employ bioinformatics tools to understand the structural and functional aspects of the novel P450 family member, CYP5619A1, from the deadliest aquatic pathogen *S. diclina*. Functional analysis of CYP5619A1 is carried out with different predicted substrates based on homolog protein with the same motifs found in fungi (Brodhun *et al.*, 2009) to identify possible substrates based on binding affinity.

3.2. Methodology

3.2.1. Homology modeling and validation

The amino acid sequence of CYP5619A1 was retrieved from the database mentioned in the literature (Sello *et al.*, 2015) and was aligned against protein structures deposited in Brookhaven Protein Data Bank (PDB) using the protein-protein BLASTp. Alternatively, the Molecular Operating Environment (MOE, 2016) was used to predict the best homolog template for CYP5619A1, for comparison purposes. The crystal structure of retinoic acid-bound CYP120A1 was used as template, since its structure was determined by both methods described earlier, to be the closest to CYP5619A1 in terms of primary sequence identity. Homology modeling of CYP5619A1 was performed using a restrained-based approach implemented in MOE. The amino acid sequence of CYP5619A1 was aligned with that of CYP120A1. A set of 10 models

was constructed for the target enzyme. The coordinates of the heme in the model were obtained from the crystal structure of CYP120A1 and the homology model was constructed along with those coordinates. The resulting three-dimensional models were optimised and a final model was obtained. The structure was validated with the Protein Structure Analysis (ProSA-web) tool (Sippl, 1993; Wiederstein & Sippl, 2007), ERRAT (Colovos & Yeates, 1993) and the VERIFY 3D program (Bowie *et al.*, 1991; Lüthy *et al.*, 1992) on the Structural Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES).

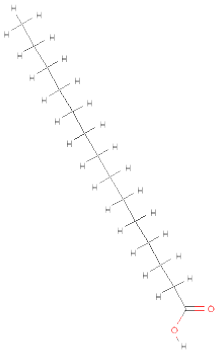
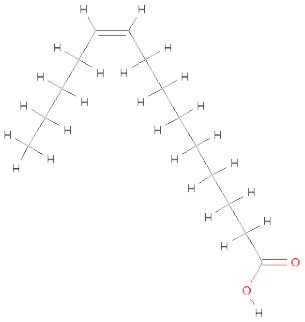
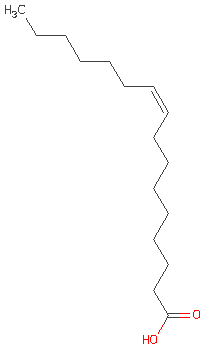
3.2.2. Binding site analysis

The software MOE was used on the final model, to assess the binding sites. A set of sites were found to be likely to bind with substrates. The site with more residues, which appeared to contain the heme group, was selected for docking studies.

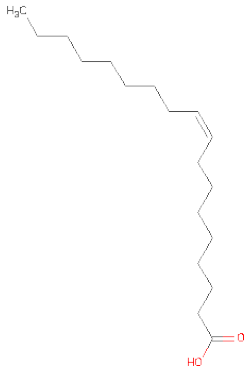
3.2.3. Ligand database

Three-dimensional structures of fatty acids of different lengths and saturation states alongside the organic compound malachite green (MGR), shown in table 3.1, were obtained from PDBeChem: Ligand Dictionary at www.ebi.ac.uk/pdbe-srv/pdbechem/PDBEntry and used in the docking of the target model.

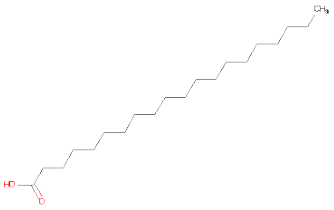
Table 3.1. Substrates used for docking (obtained from PDBeChem: Ligand Dictionary at www.ebi.ac.uk/pdbe-srv/pdbechem/PDBEntry).

Saturated fatty acids	Unsaturated fatty acids
C14:0 Myristic acid 	**C14:1 Δ9Z Myristoleic acid** 
C16:0 Palmitic acid 	**C16:1 Δ9Z Palmitoleic acid** 
C18:0 Stearic acid	**C18:1 Δ9Z Oleic acid**

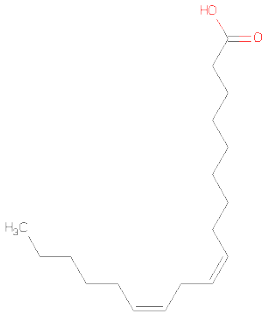
C20:0 Icosanoic acid



C18:2 Δ9Z,12Z Linoleic acid

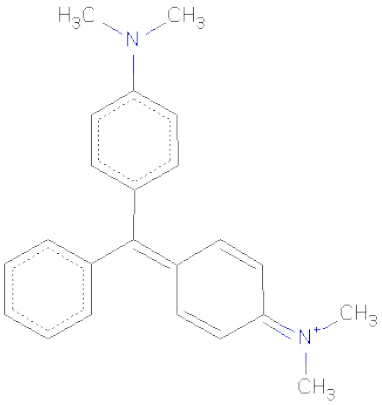


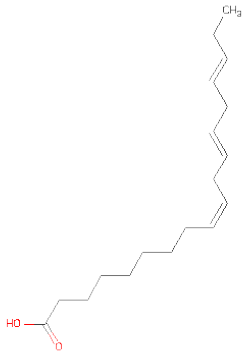
ORGANIC COMPOUND



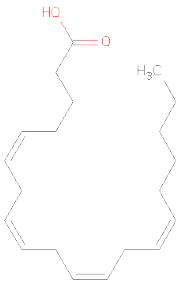
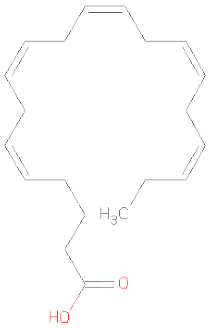
C18:3 Δ9Z,12Z,15Z alpha-Linolenic acid

Malachite Green





C20:4 Δ5Z,8Z,11Z,14Z Arachidonic acid

	
	C20:5 Δ5Z,8Z,11Z,14Z,17Z **Eicosapentaenoic acid** 

3.2.4. Molecular docking

The CYP5619A1 model was prepared for docking in MOE and AutoDockTools 1.5.6 (Goodsell & Olson, 1990). MOE was used to correct the protonation and remove the solvent. The different ligands were all prepared for docking in AutoDockTools, following the same steps as the target protein: protonation, addition of charges, merging of non-polar H+ and assignment of atom types. Partial charges of ligands and protein were generated using the Gasteiger method with the aid of AutoDockTools. Non-polar hydrogens were merged and atom types were assigned. A cubic grid having 60 × 60 × 60 grid points per side and spacing of 0.375 Å was set, which corresponds to the substrate recognition site of the target P450 model. The grid was positioned onto the substrate access channel extending into the binding pocket of the model. Affinity maps of the grid were calculated using the AutoGrid program. The AutoDock 4.0 program was used to dock 12 ligands into the active-site cavity of the target model using the Lamarckian genetic algorithm, consisting of 200 runs and 270 000 generations, with the maximum number of energy evaluations set to 2.5 × 106. The resulting docked conformations within 2.0 Å root mean square deviation (RMSD) tolerance were clustered and analysed using AutoDockTools. Conformations with the lowest interaction energy and closest interaction to heme iron were selected for each ligand and rendered.

3.3. Results and discussion

3.3.1. Sequence alignment

The sequence alignment of CYP5619A1 and the template 2VE3 obtained from BLASTp result has been carried out. The alignment (Figure 3.1) shows a highly conserved secondary structure despite being poorly conserved at amino acid sequence level.

3.3.2. Structural analysis of CYP5619A1

Multiple sequence alignment of CYP5619A1 with CYP120A1 demonstrated that the two proteins are poorly homologous (28%), which was confirmed by the BLASTp homology search using the *S. diclina* CYP5619A1 sequence against the PDB database. Nevertheless, this is comprehensible, since CYP5619A1 from *S. diclina* belongs to a novel protein family that has not yet been characterised. The template for homology modeling of CYP5619A1 is the crystal structures of substrate-free and all-trans-retinoic acid-bound CYP120A1 from *Synechocystis* sp. PCC 6803, the first structural characterisation of a cyanobacterial P450, which was determined at 2.4 and 2.1 Å resolution, respectively (Kuhnel *et al.*, 2008).

The resulting modeled enzyme is a monomer, folded into a α/β domain consisting of a seven-stranded β-sheet and 14 α-helices (Figure 3.1 and 3.2). The β-sheet tends to form the hydrophobic substrate channel. The residues in the Glu287-Arg290 (EXXR) motif are found in the K helix. Literature suggested that they might be involved in stabilising the core structure of the protein and are on the proximal side of the heme (Graham & Peterson, 1999). Furthermore, the heme (displayed in sticks in Figure 3.2) is bound to the absolutely conserved cysteine at position 371, which is the fifth ligand of the heme iron, and the reason for the typical 450 nm soret absorbance found in CO-bound P450s (Omura & Sato, 1962).

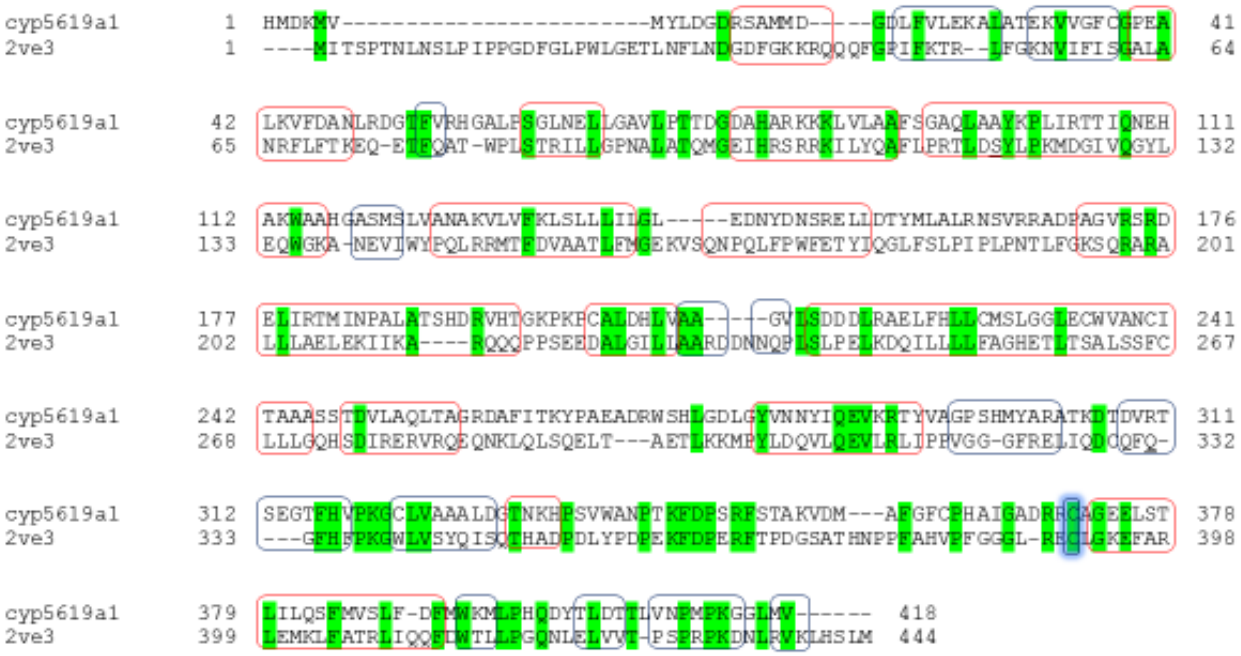


Figure 3.1. Sequence alignment of protein CYP5619A1 with template CYP120A1 (PDB ID: 2VE3). Red rectangles represent α-helices, blue rectangles represent β-sheets. The Cys-pocket is shown in a blue rectangle. Conserved residues are highlighted in green.

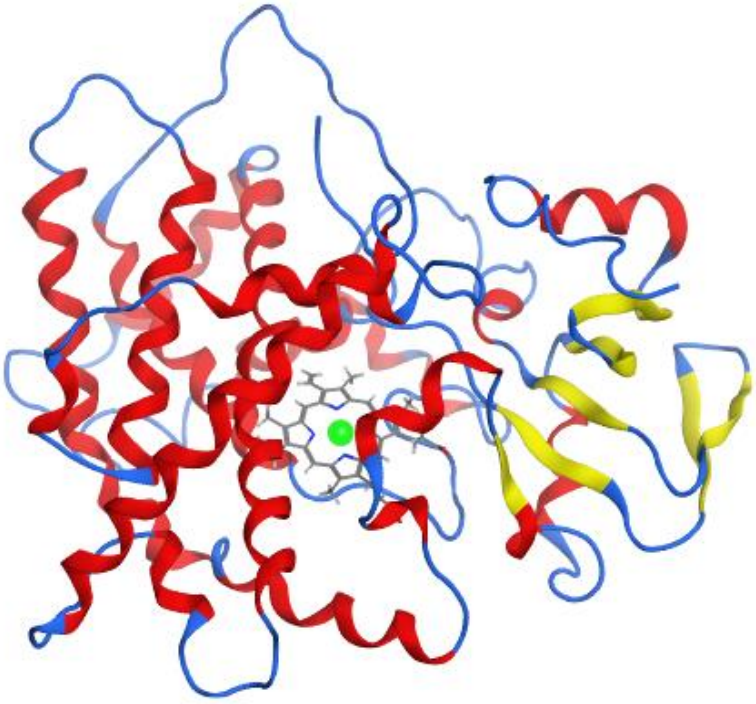


Figure 3.2. Homology model of CYP5619A1 with heme prosthetic group. Secondary structures are displayed in red (helices), yellow (sheets) and blue (coils and turns). The heme prosthetic group is shown in sticks at the center of the molecule with the iron as green ball.

Homology modeling usually results in the production of protein models with quite unfavourable bond lengths, bond angles, torsion angles and contacts. In that case, it is essential to minimise the energy in order to regularise local bond and angle geometry, and to relax close contacts in the geometric chain. The model of CYP5619A1 from *S. diclina* was optimised using the *tleap* and *sander* programs of the AMBER suite. *Sander* is the main simulation engine of the entire suite; it takes two input files describing the molecular system to be simulated; one control file specifies the conditions of the simulation and computes a classical molecular dynamics trajectory based on this information. *tleap* is the helper program that takes predetermined coordinate files such as pdb and generates a topology file and a restart file, to feed to *sander*. Energy minimisation was performed to minimise stearic collisions and strains without significantly altering the overall structure. Energy computations and minimisation were carried out using the Amber14 force field. After optimisation the 3D model of CYP5619A1 was verified using ProSA-web (Sippl, 1993; Wiederstein & Sippl, 2007), as well as the ERRAT (Colovos & Yeates, 1993) and VERIFY 3D (Bowie *et al.*, 1991; Lüthy *et al.*, 1992) programs available from the Structural Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES). ProSA-web was used to calculate the Z-score, while the Verify3D program analysed the compatibility of an atomic model (3D) with its own amino acid sequence (1D) to assess the 3D protein structure. ERRAT verifies crystallography-determined protein structures and plots error values as a function of the position of a sliding nine-residue window, based on the statistics of non-bonded atom-atom interactions in the reported structure (compared to a database of reliable high-resolution structures). Validation results are displayed in Figures 3.3 to 3.5.

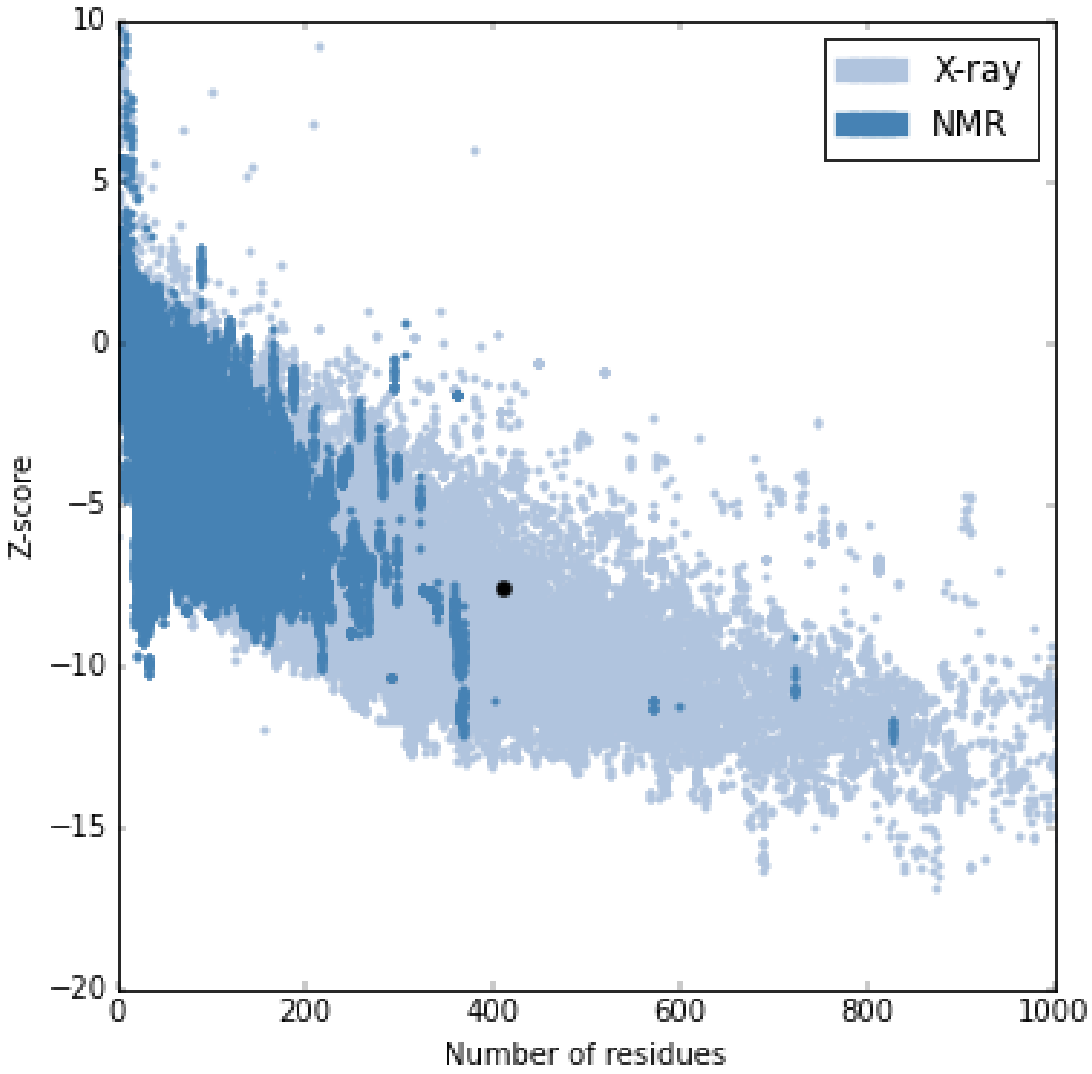


Figure 3.3. Z-score estimation for CYP5619A1 refined model on the ProSA-web server. The z-score of the modeled protein (-7.61) is represented as a black dot.

CYP5619A1 from *S. diclina* had a z-score of -7.61, indicating a good overall model quality. The score is represented by a dot displayed in a plot that contains the z-scores of all experimentally determined protein chains in current PDB (Figure 3.3). In the plot, groups of structures from X-ray crystallography and NMR are displayed in dark and light blue respectively and serve as a basis for comparison with the modeled protein. As shown, the z-score of the input structure (that of CYP5619A1) is within the range of scores typically found for native proteins of similar size.

ERRAT has been termed an "overall quality factor" for non-bonded atomic interactions, with higher scores indicating higher quality. The generally accepted range is >95 for a high-quality model. For the current 3D model, the overall quality factor predicted by the ERRAT server was 96.226 (Figure 3.5). The Verify 3D server predicted that 86.36% of the residues in CYP5619A1 would have an average 3D-1D score > 0.2 (Figure 3.4), thereby confirming the good quality of the model, since the minimum percentage for good quality is 80.

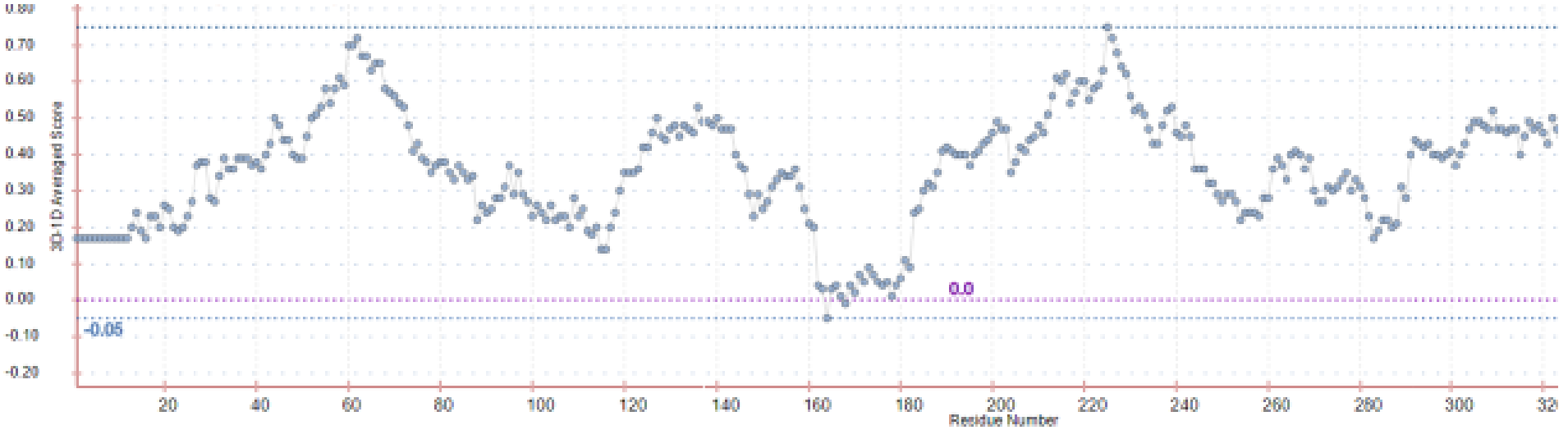


Figure 3.4. Verify 3D result for CYP5619A1 refined model. 86.36% of the residues had an averaged 3D-1D score ≥ 0.2. Pass: At least 80% of the amino acids scored ≥ 0.2 in the 3D/1D profile.

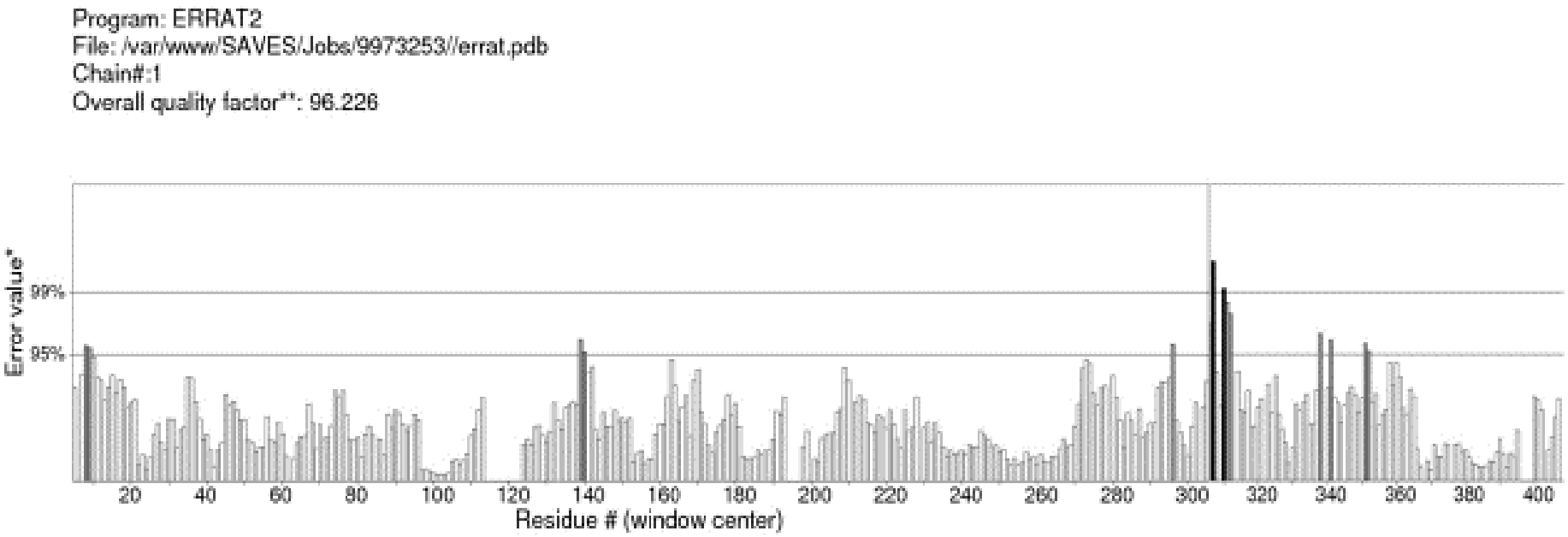


Figure 3.5. ERRAT result for CYP5619A1 refined model.

*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.
**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high-resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3 Å), the average overall quality factor is around 91%.

3.3.3. Model-template superimposition

For more assurance on the quality of the model, the structures of CYP5619A1 and CYP120A1 were superimposed and compared based on the distance between their Cα backbones. The two superimposed structures have been coloured by RMSD and are displayed in Figure 3.6 below.

The green colour indicates a very narrow gap between the two structures, while lime-green and white colours indicate a wide gap and a total mismatch between the query sequence and the template, respectively. Mismatches are observed particularly around loops and also at the template structure's N- and C- termini, which are observed in the alignment (Figure 3.1) protruding from the CYP5619A1 sequence.

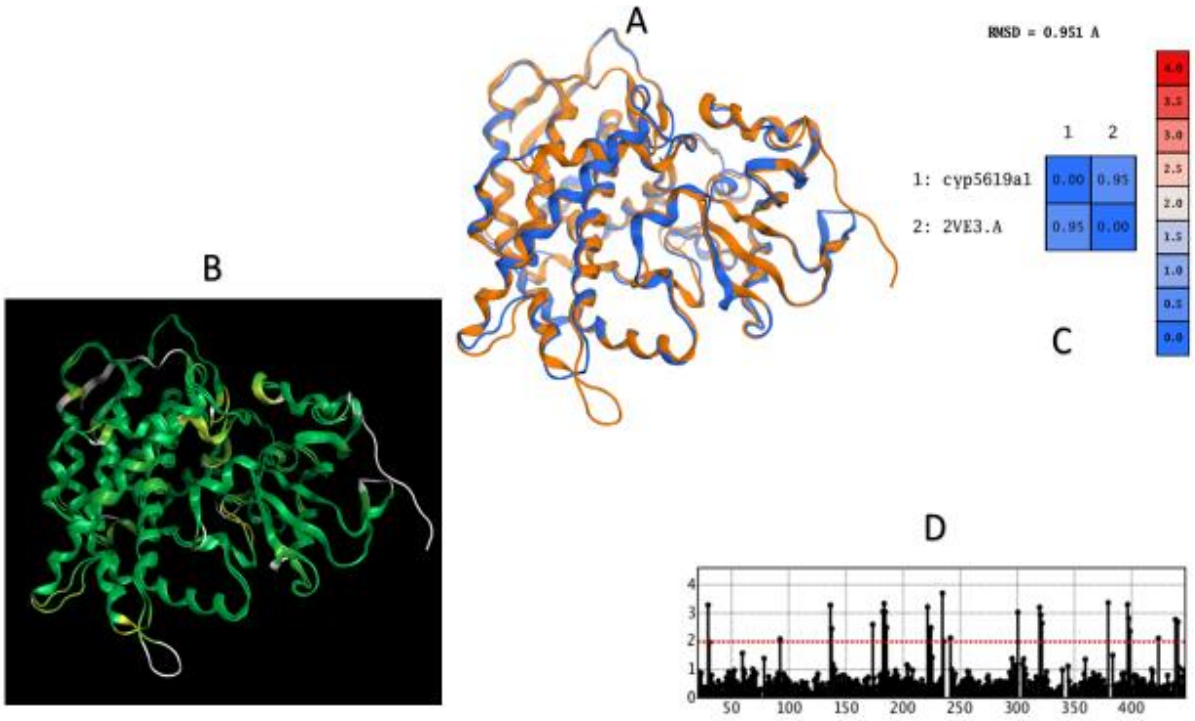


Figure 3.6. Superimposed structures of the target protein CYP5619A1 and its template CYP102A1 (PDB ID: 2VE3). ***A)*** The superimposed structures are displayed in blue (CYP5619A1) and orange (CYP102A1). ***B)*** The superimposed structures are coloured by RMSD, with the green colour indicating close proximity between α-carbons. ***C)*** RSMD calculation; the overall RMSD is 0.951 Å. ***D)*** RMSD plot of all aligned residues showing only a few aligned residues with RMSD over 2 Å.

3.3.4. Active site mapping

After the final model had been obtained and its quality confirmed, different potential binding sites of CYP5619A1 were searched using MOE, and structural comparison of the template and the model was done. In this study, the sites were searched in order to determine the protein active sites and binding sites by locating cavities in the CYP5619A1 structure that will allow access to the heme group. When the search was complete, the largest site was automatically displayed on the structure, as shown in Figure 3.7.

Furthermore, the binding pocket was viewed and displayed (Figure 3.8). As shown in the Figure 3.8, the heme is in the core of the pocket, which appears to be highly hydrophobic, suggesting a very high affinity with the docked fatty acids, as shown in the docking results in the following section (Figure 3.21)

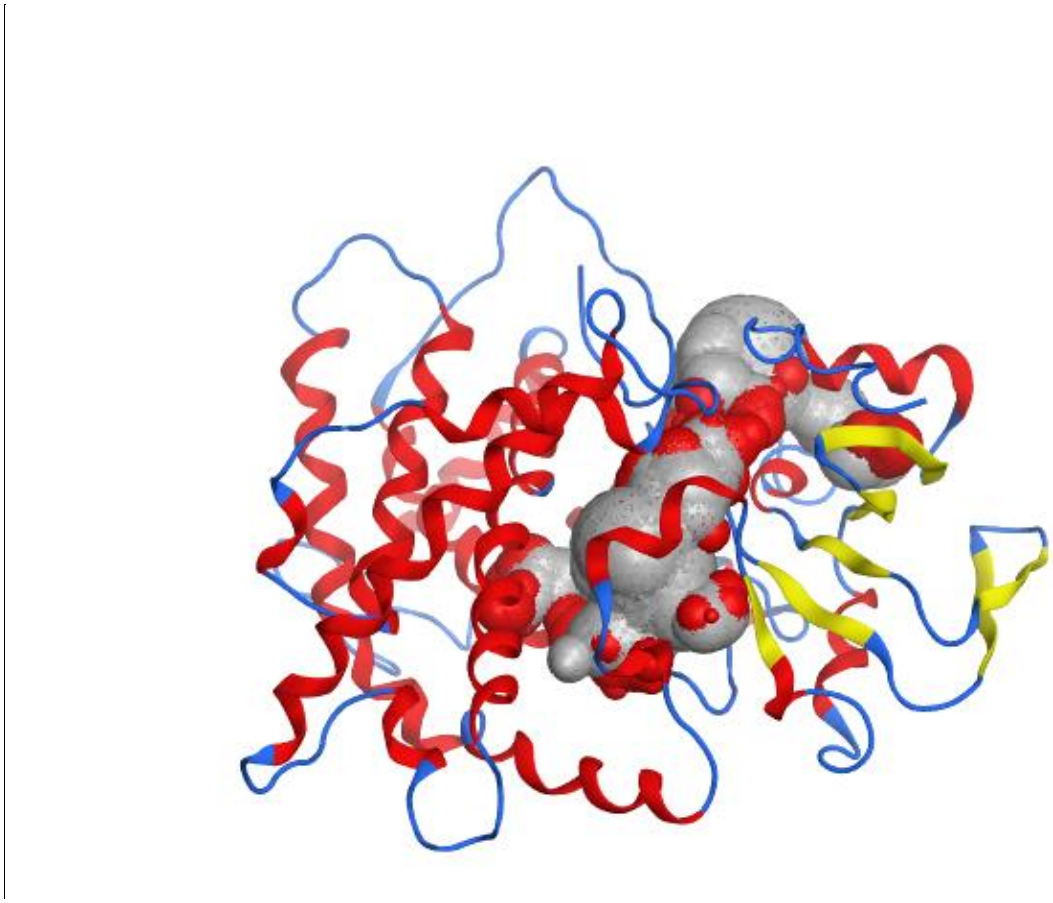


Figure 3.7. Active site cavity of CYP5619A1. Secondary structures are displayed in red (helices), yellow (sheets) and blue (coils and turns). The active site cavity is shown with the substrate access channel in grey (hydrophobic site) and red (hydrophilic site) spheres.

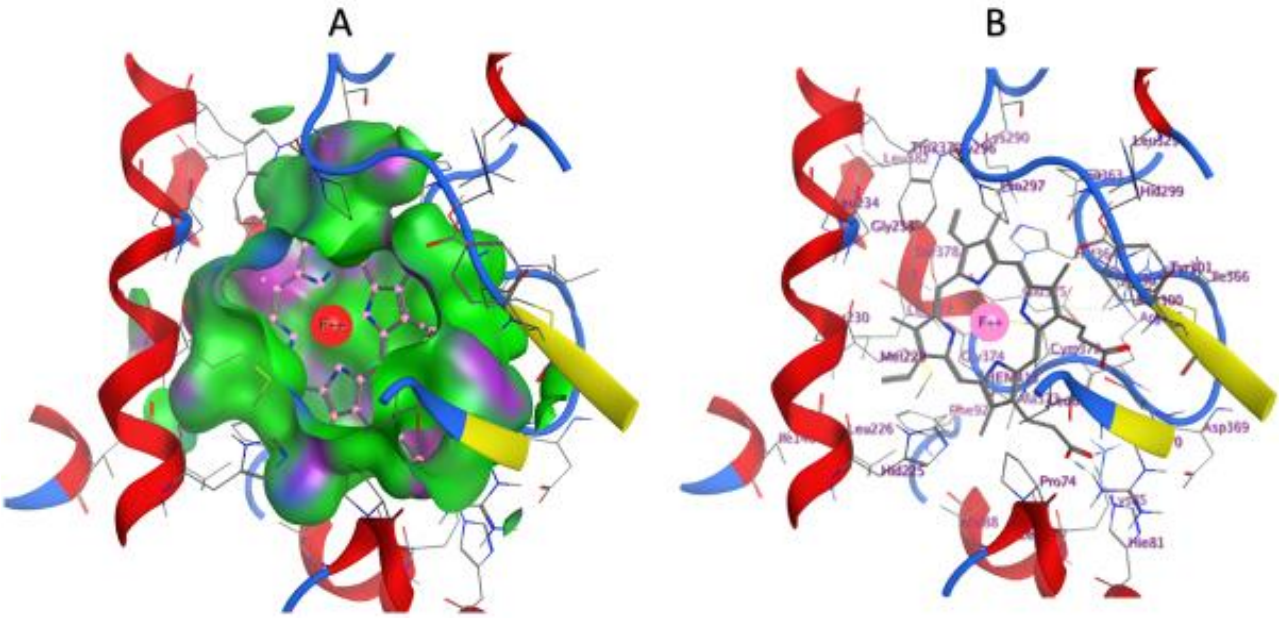


Figure 3.8. Active site view of the binding pocket of CYP5619A1. ***A)*** The pocket is displayed with MOE ActiveLP colour coding (Blue: mild polar; Green: Hydrophobic; H-Bonding: Pink) and shows a pattern of high hydrophobicity. ***B)*** Residues forming the pocket are labelled. Secondary structures in ***A*** and ***B*** are displayed in red (helices), yellow (sheets) and blue (coils and turns). The heme prosthetic group appears at the centre of the active site in grey sticks.

3.3.5. Substrate binding analysis

The docking of model CYP5619A1 with 12 ligands (predicted substrates): myristic acid (MYR), palmitic acid (PLM), stearic acid (STE), icosanoic acid (DCR), myristoleic acid (MYZ), palmitoleic acid (PAM), oleic acid (OLE), linoleic acid (EIC), alpha-linolenic acid (LNL), arachidonic acid (ACD), eicosapentaenoic acid (EPA) and MGR, was performed with AutoDockTools. The 11 ligands were selected as possible substrates for CYP5619A1 based on homolog protein with the same motif found in fungi (Brodhun *et al.*, 2009). The best results were selected according to the output clustering histogram and thus the lowest binding energies. The representative conformation for each cluster was chosen as the best pose for each ligand and was viewed in AutoDockTools. Interactions between residues are displayed in Figures 3.9 to 3.20.

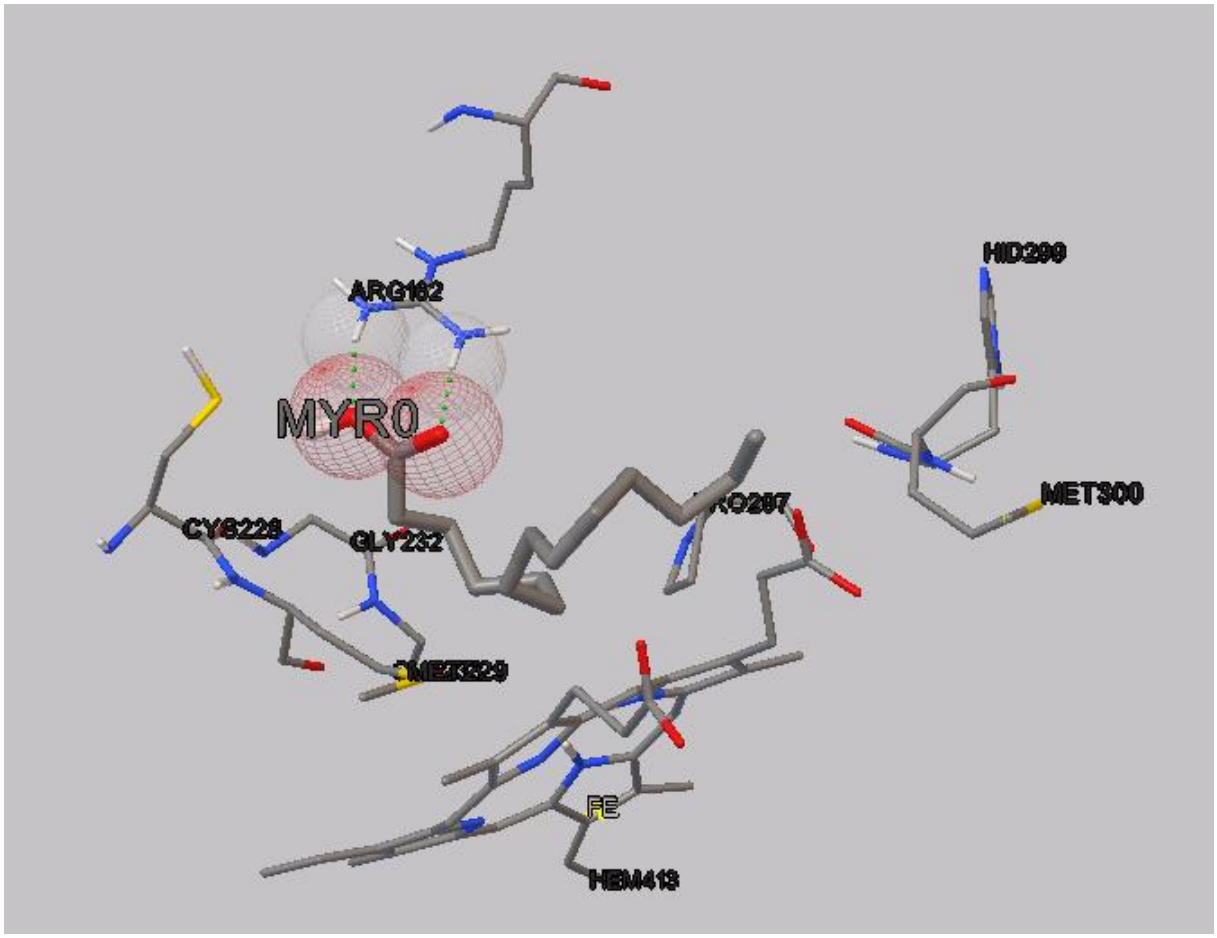


Figure 3.9. Interaction of myristic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

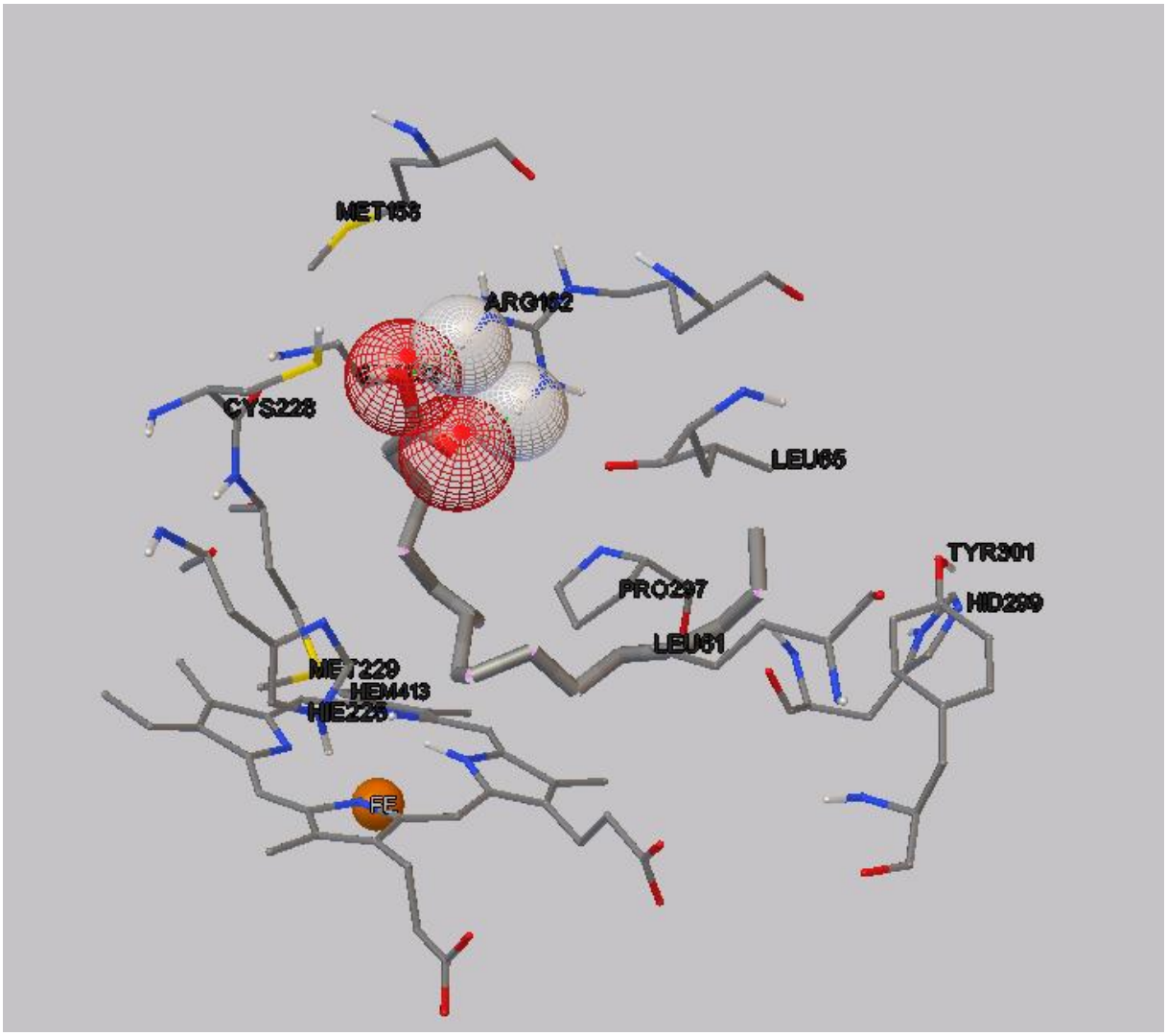


Figure 3.10. Interaction of palmitic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

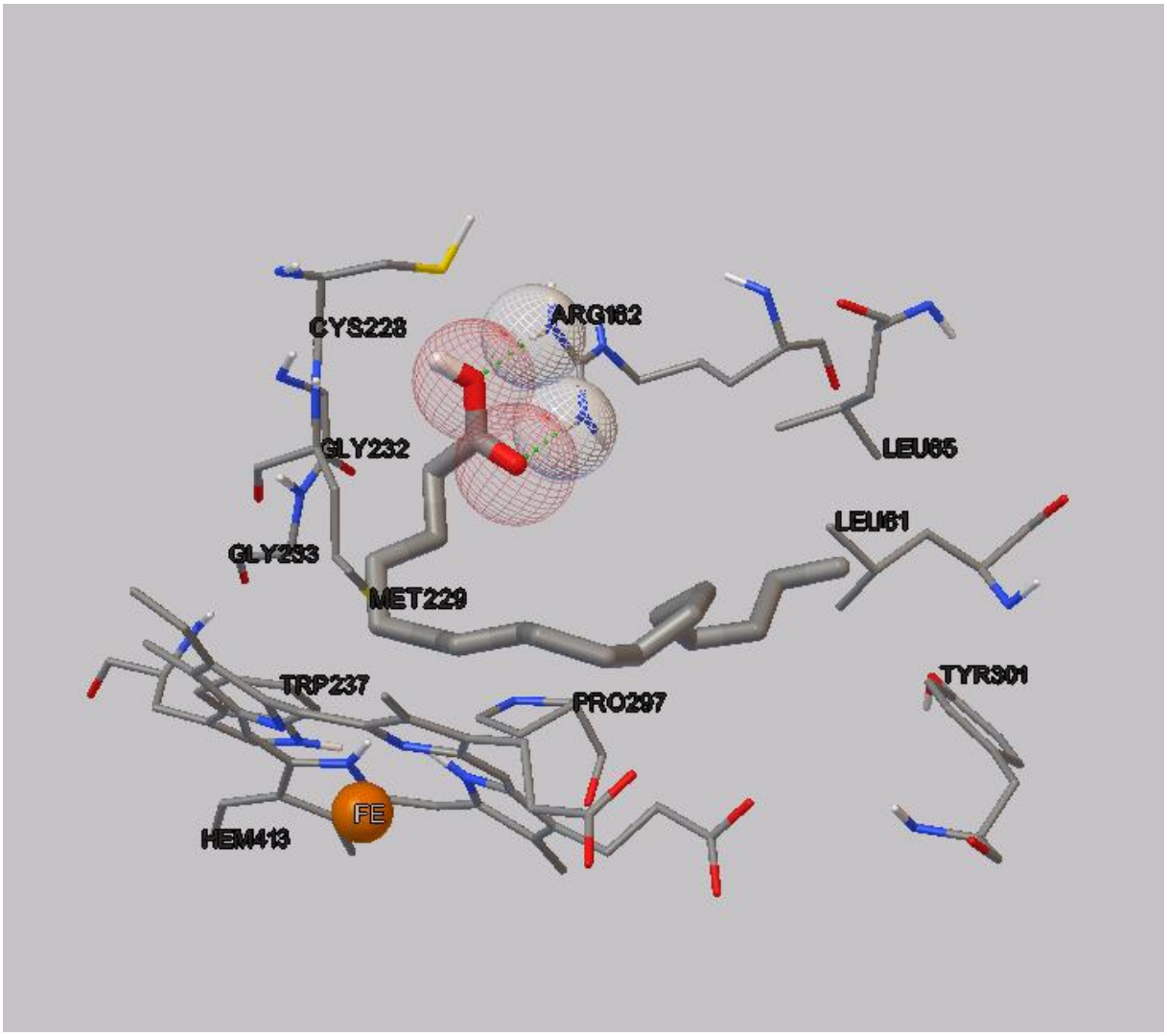


Figure 3.11. Interaction of stearic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

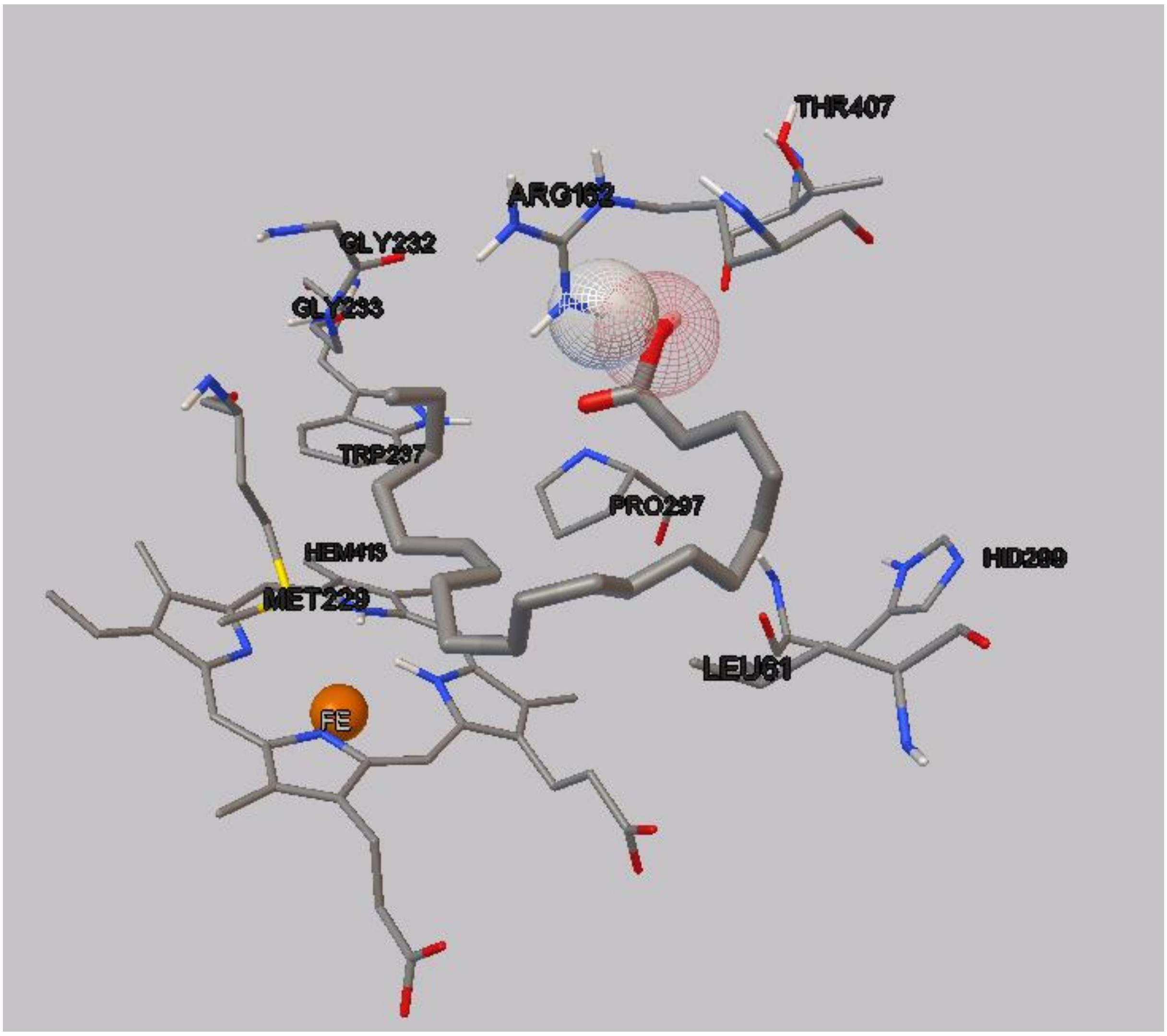


Figure 3.12. Interaction of icosanoic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

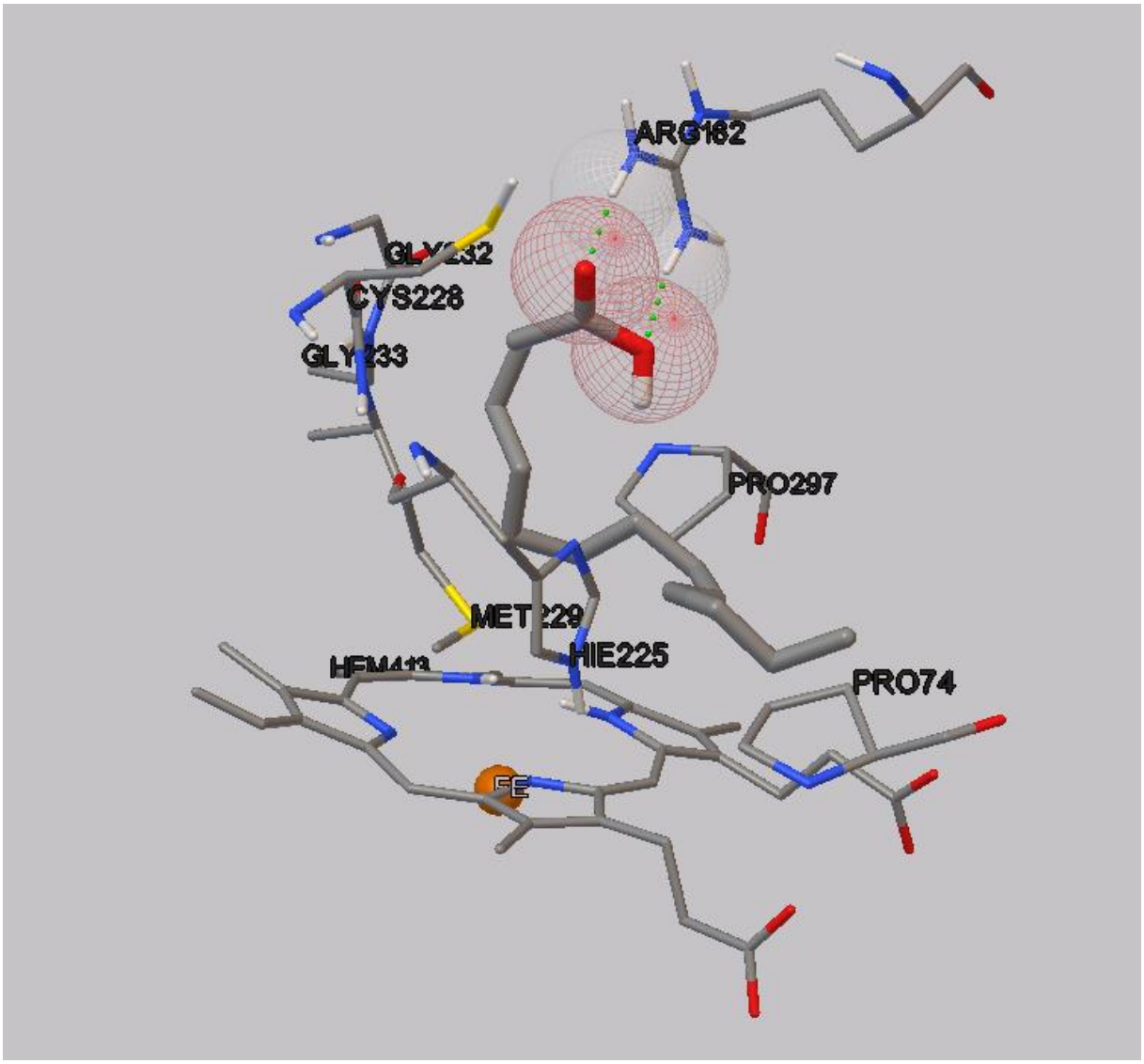


Figure 3.13. Interaction of myristoleic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

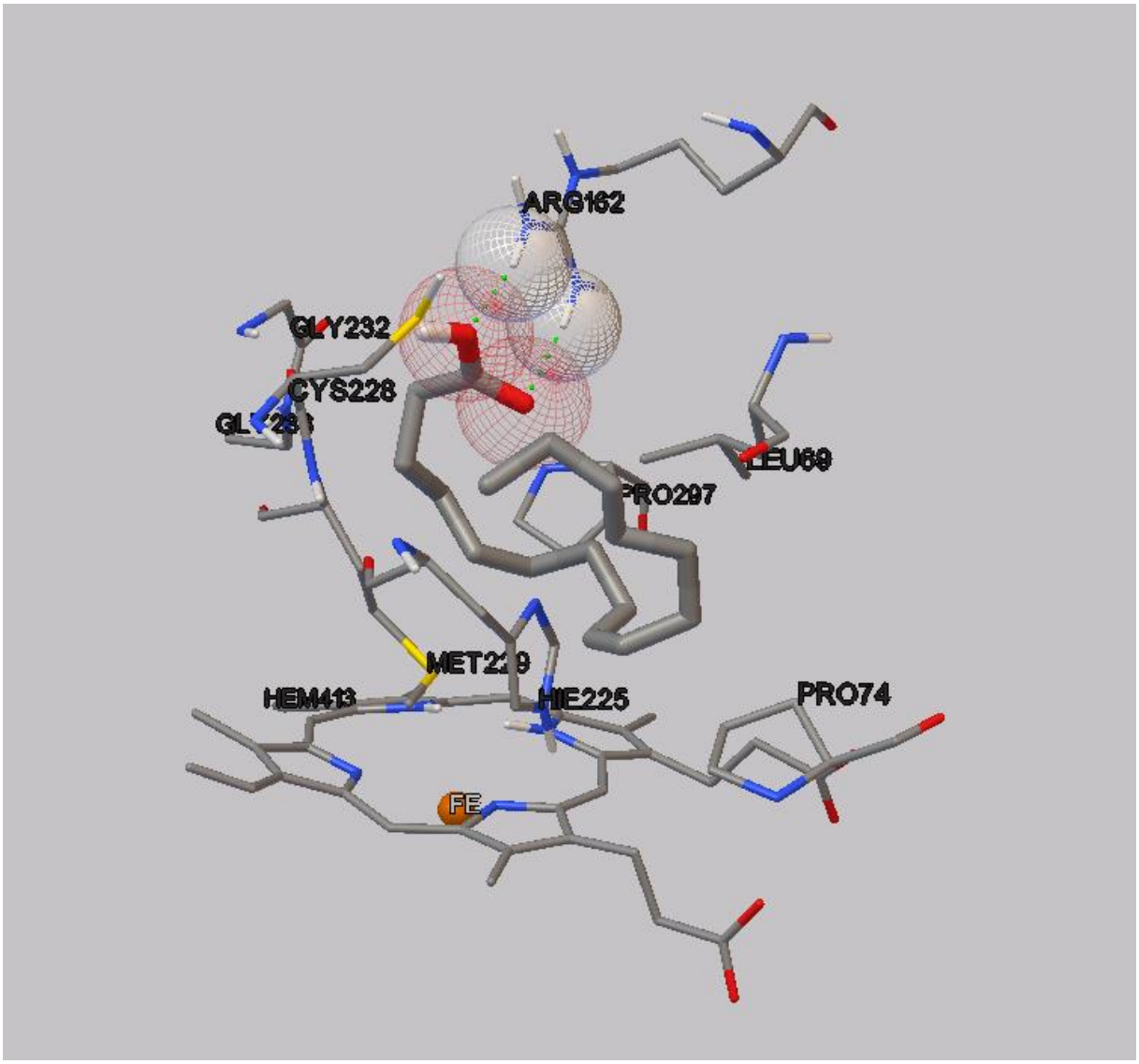


Figure 3.14. Interaction of palmitoleic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

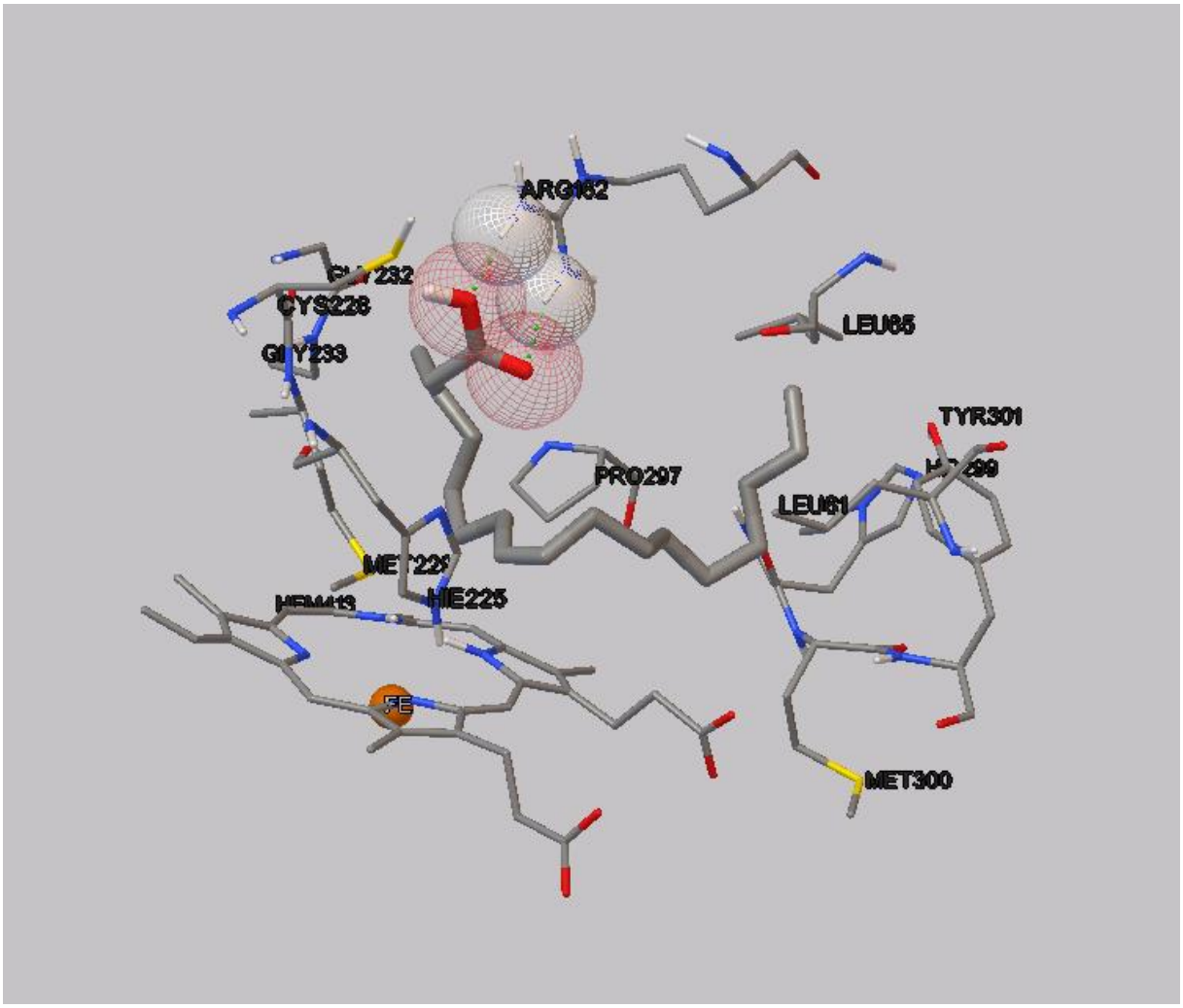


Figure 3.15. Interaction of oleic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

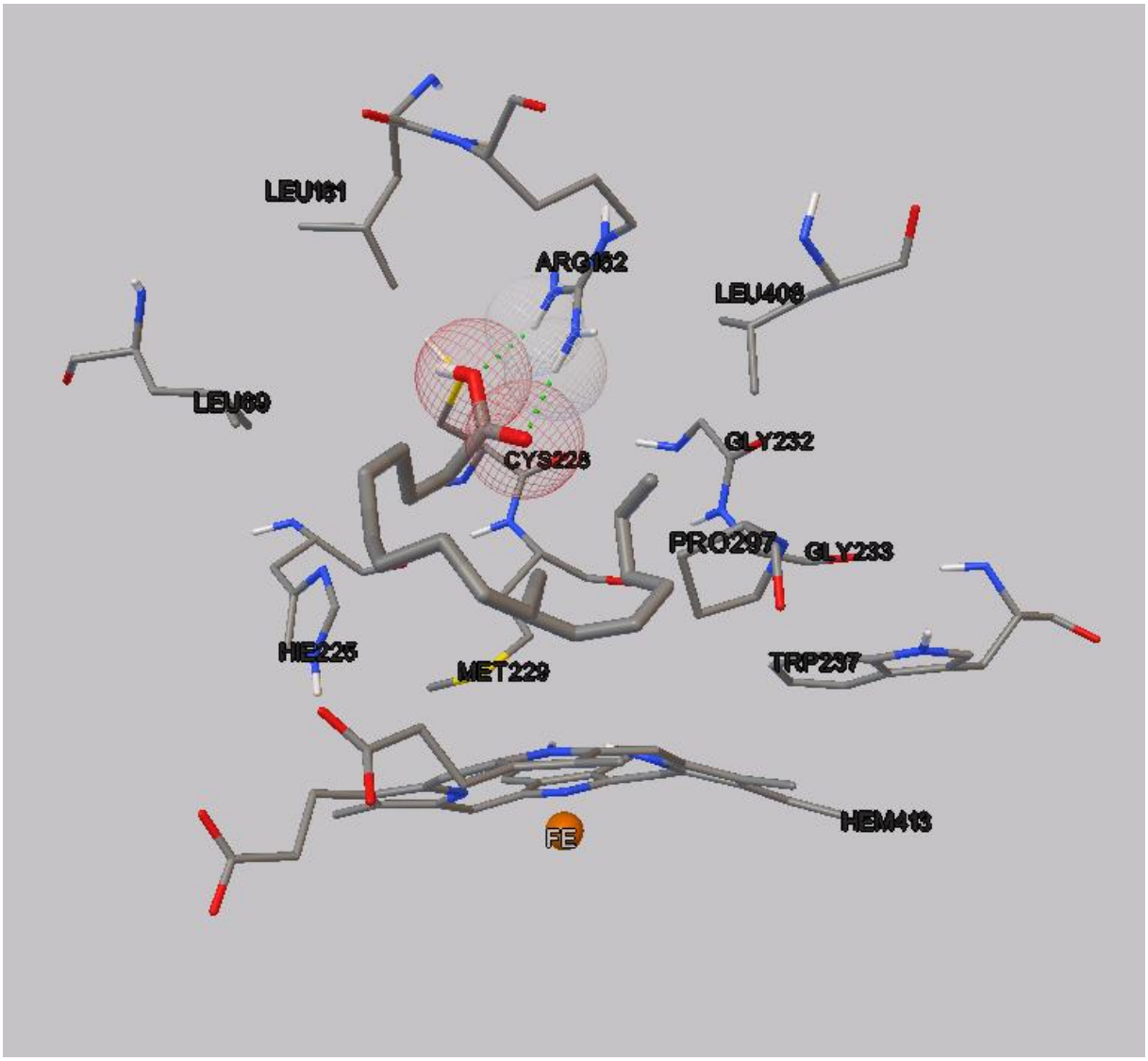


Figure 3.16. Interaction of linoleic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

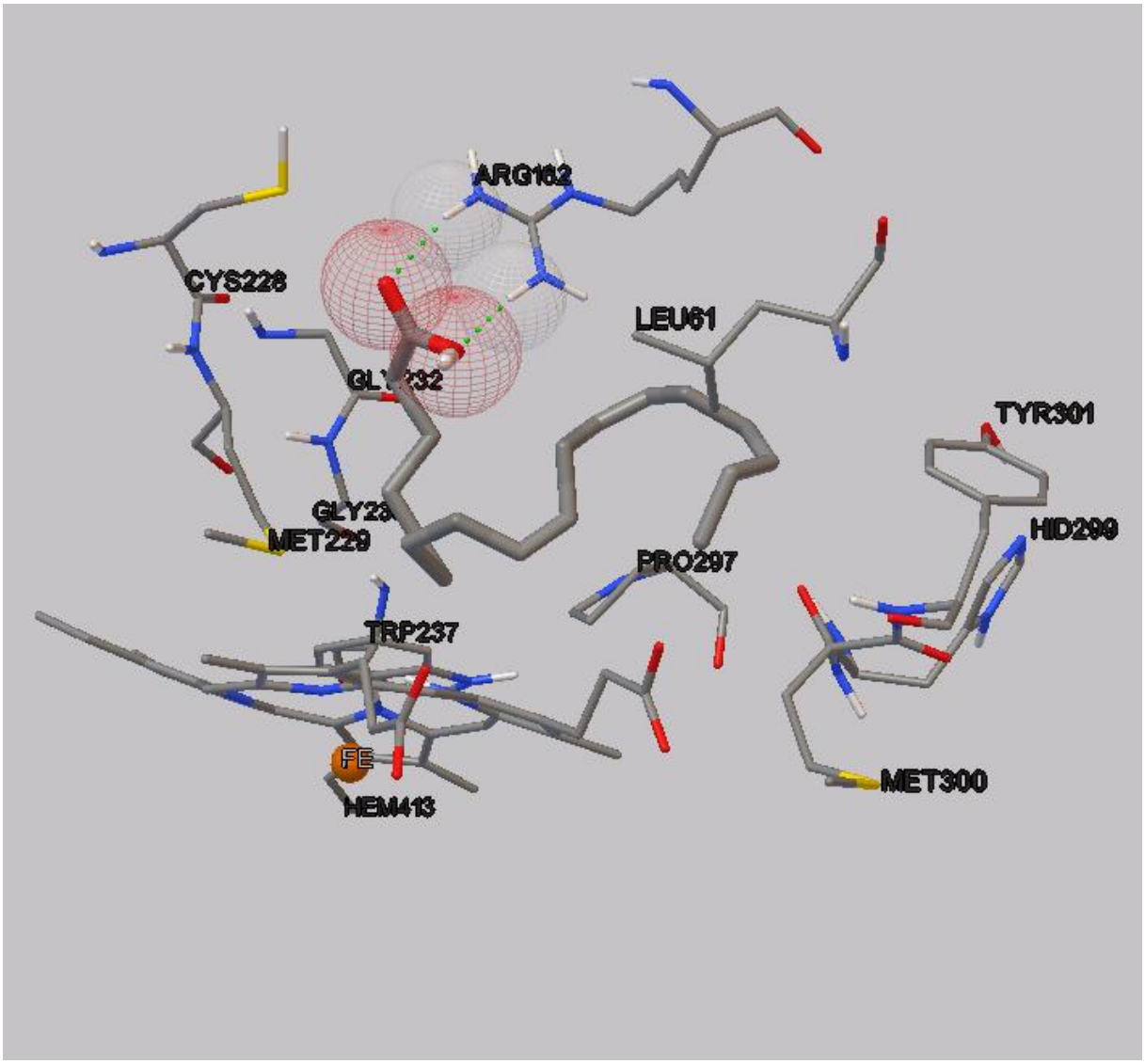


Figure 3.17. Interaction of alpha-linolenic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

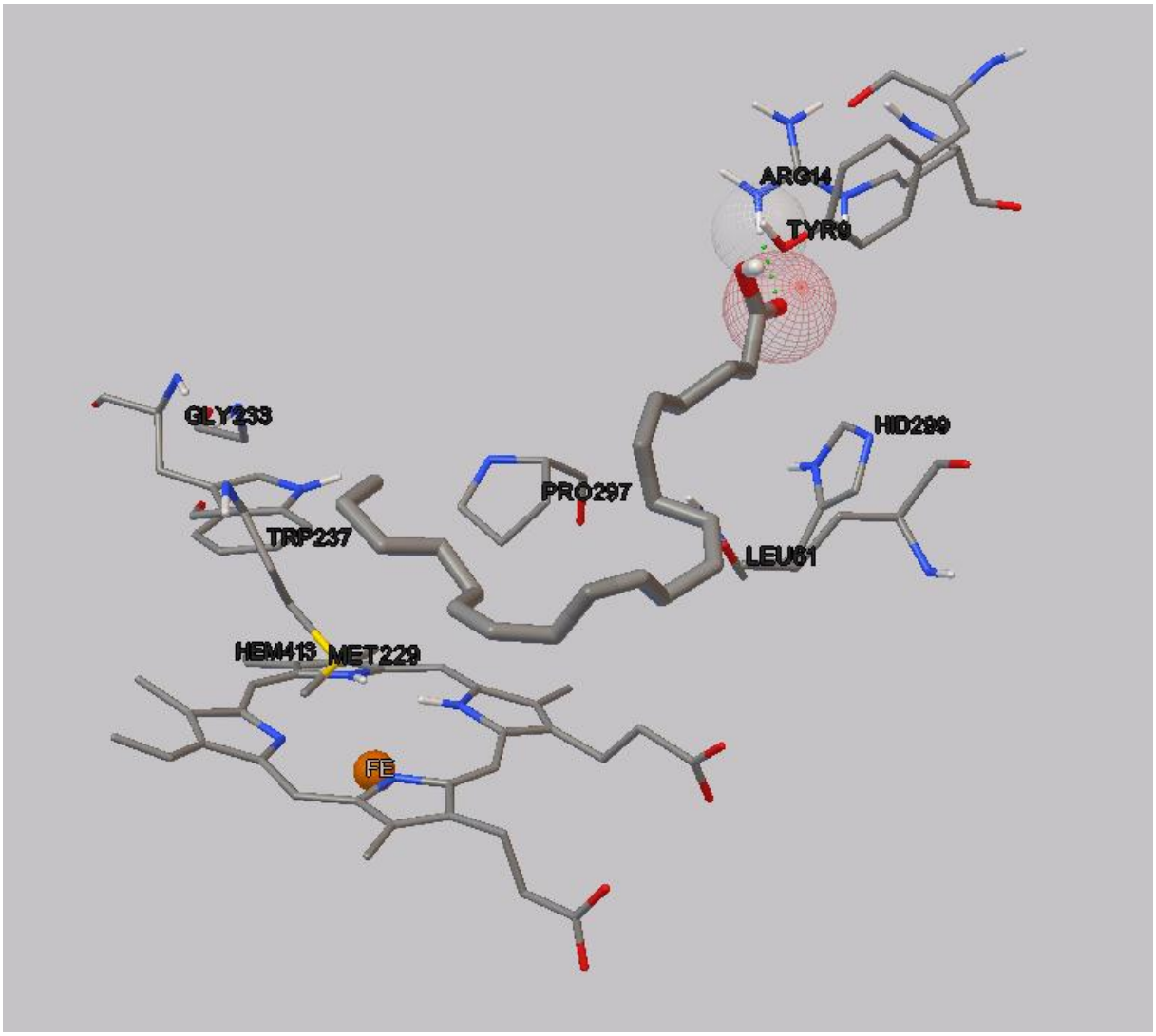


Figure 3.18. Interaction of arachidonic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

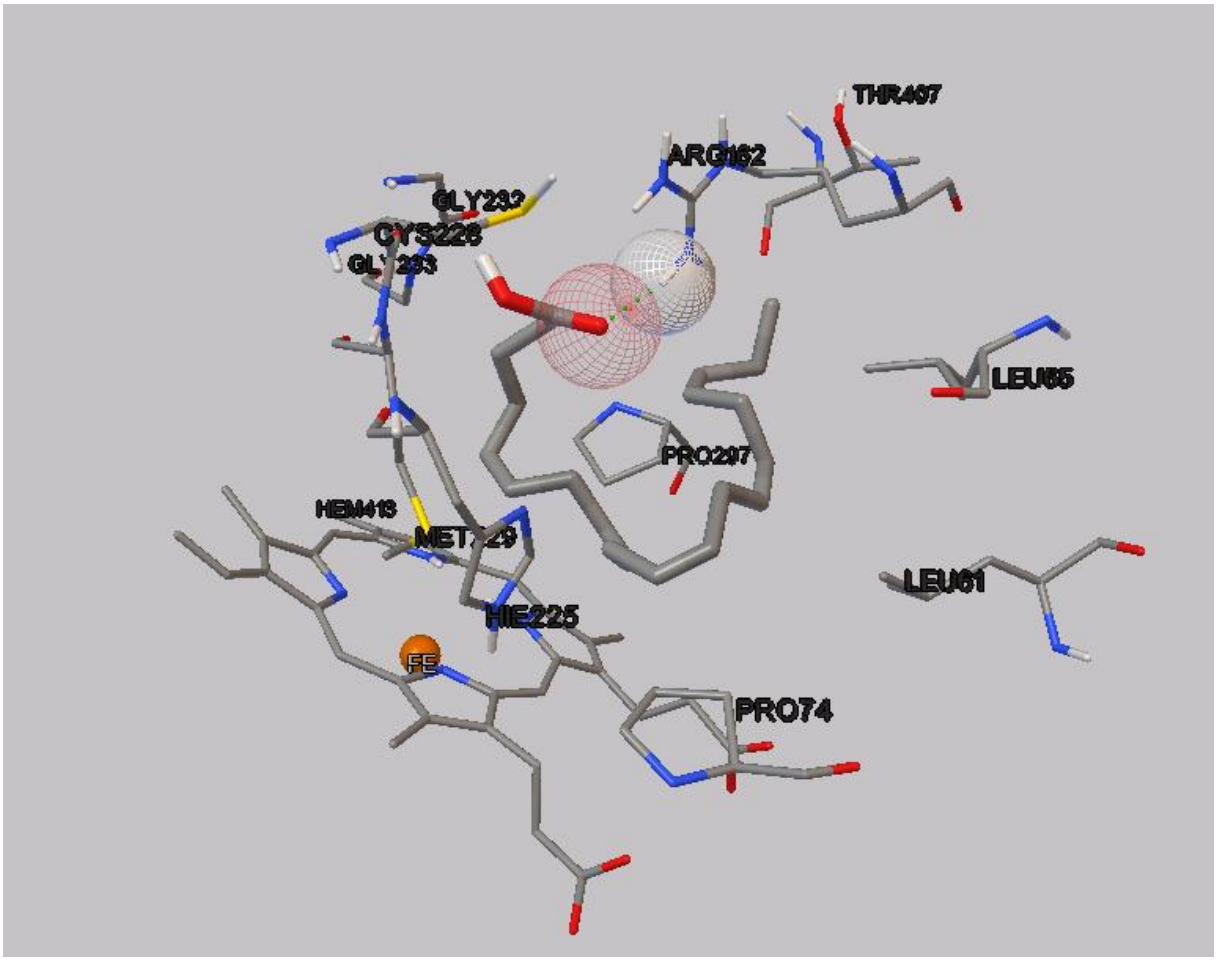


Figure 3.19. Interaction of eicosapentaenoic acid with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

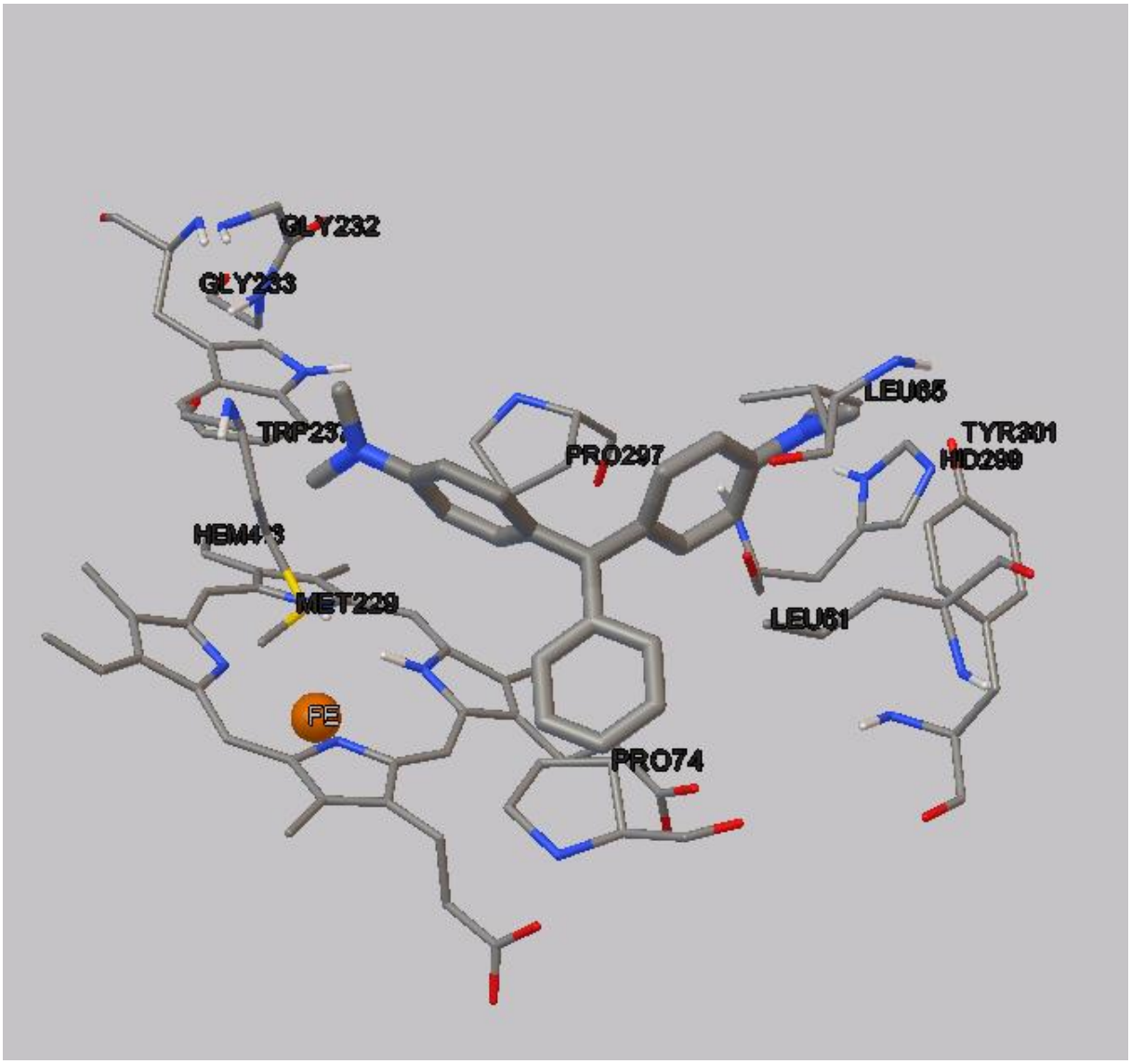


Figure 3.20. Interaction of malachite green with CYP5619A1 model. The ligand is displayed in thick lines and the interacting residues in thin lines. The residues are labelled by three-letter codes and position in the protein sequence. The heme iron is also labelled. Colour coding is as follows: white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur. Hydrogen bonds are displayed as green dots and the atoms involved are accentuated by spheres coloured as per the respective atoms.

The molecular docking studies showed that EIC and MGR form a better complex than others with the lowest free binding energy of -6.70 kcal/mol (Figure 3.21). EIC forms a tight hydrogen bond with residue Arg162 of the CYP5619A1 model (Figure 3.16).

However, EIC conformations show a pattern of instability that is observed when comparing the binding energies of its different poses to those of MGR (Figure 3.22), suggesting that despite the same binding energy of their best conformations and MGR's life-threatening properties, MGR remains the best compound that possibly inhibits the metabolic action of CYP5619A1. Nonetheless, this comparison is limited to a small set of docked ligands and might not constitute a tangible argument to draw conclusions.

Furthermore, analysis of amino acid residues interacting with different ligands (11 possible substrates and MGR) revealed that most of the residues interacting with different ligands are the same, suggesting that all ligands are interacting with almost the same amino acids and the binding patterns are conserved (Table 3.2).

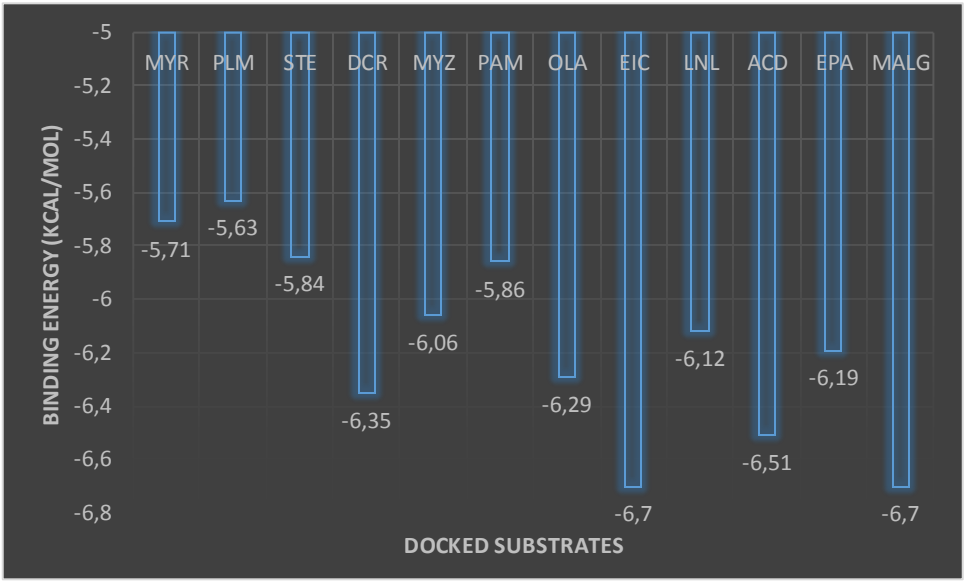


Figure 3.21. Graphic representation of the free binding energies of the docked possible substrates and malachite green. Abbreviations: MYR: myristic acid; PLM: palmitic acid; STE: stearic acid; DCR: icosanoic acid; MYZ: myristoleic acid; PAM: palmitoleic acid; OLE: oleic acid; EIC: linoleic acid; LNL: alpha-linolenic acid; ACD: arachidonic acid; EPA: eicosapentaenoic acid; MALG: malachite green.

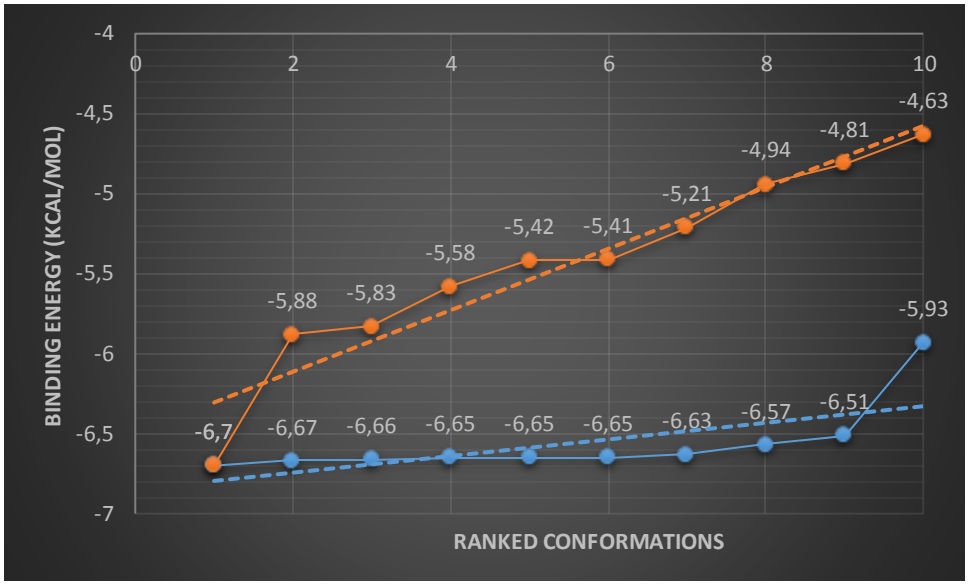


Figure 3.22. Graphic comparison of binding energies for each conformation of the best two ligands. The binding energy for each of the 10 conformations of linoleic acid is plotted in orange. The binding energy for each of the 10 conformations of malachite green is plotted in blue. The trend for each ligand is plotted as a dashed line of the same colour.

Table 3.2. Amino acids residues interacting with the different ligands.

Ligand code	Interacting residues
MYR	Arg162 (2HB), Cys228, Met229, Gly232, Pro297, His299, Met300, HEM413
PLM	Leu61, Leu65, Met158, Arg162 (2HB), Cys228, His225, Met229, Pro297, His299, Tyr301, HEM413
STE	Leu61, Leu65, Arg162 (2HB), Cys228, Met229, Gly232, Gly233,Trp237, Pro297, Tyr301, HEM413
DCR	Leu61, Arg162 (1HB), Met229, Gly232, Gly233, Trp237, Pro297, His299, Thr407, HEM413
MYZ	Pro74, Arg162 (2HB), His225, Cys228, Met229, Gly232, Gly233, Pro297, HEM413
PAM	Leu69, Pro74, Arg162 (2HB), His225, Cys228, Met229, Gly232, Gly233, Pro297, HEM413
OLA	Leu61, Leu65, Arg162 (2HB), His225, Cys228, Met229, Gly232, Gly233, Pro297, His299, Met300, Tyr301, HEM413
EIC	Leu69, Leu161, Arg162 (2HB), His225, Cys228, Met229, Gly232, Gly233, Trp237, Pro297, Leu408, HEM413
LNL	Leu61, Arg162 (2HB), Cys228, Met229, Gly232, Gly233, Trp237, Pro297, His299, Met300, Tyr301, HEM413
ACD	Tyr9, Arg14 (1HB), Leu61, Met229, Gly233, Trp237, Pro297, His299, HEM413
EPA	Leu61, Leu65, Pro74, Arg162 (1HB), His225, Cys228, Met229, Gly232, Gly233, Pro297, Thr407, HEM413
MGR	Leu61, Leu65, Pro74, Met229, Gly232, Gly233, Trp237, Pro297, His299, Tyr301, HEM413

Abbreviations: MYR: myristic acid; PLM: palmitic acid; STE: stearic acid; DCR: icosanoic acid; MYZ: myristoleic acid; PAM: palmitoleic acid; OLE: oleic acid; EIC: linoleic acid; LNL: alpha-linolenic acid; ACD: arachidonic acid; EPA: eicosapentaenoic acid; MGR: malachite green.

3.4. Conclusion

In this study, a 3D model of CYP5619A1 from *S. diclina* was built by homology modeling and assessed for its binding affinity with different predicted substrates based on homolog protein with the same motifs from fungi and with MGR remedy used to treat *S. diclina* infections. The study revealed that eight of the compounds required low energy to bind to the target protein, with binding energies below -6.00 kcal/mol. This suggests that these ligands can act as possible substrates of CYP5619A1. Among all ligands, EIC and MGR showed high binding affinity with the CYP5619A1 model. EIC is a polyunsaturated fatty acid with 18 carbon atoms and two double bonds in its structure. MGR is an organic compound that is widely used in aquaculture to treat *S. diclina* infections. These two compounds appeared to be the compounds with the best affinities to the target protein. In this regard, it is reasonable to believe that EIC-like compounds could be potential substrates for CYP5619A1 and MGR, possibly inhibiting CYP5619A1 in *S. diclina*.

References

- Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology*, *124*(1), pp. 128-145.
- Bernhardt, R., 2013. Cytochromes p450 as versatile biocatalysts for biotechnological application. *Успехи наук о жизни*, (7), pp. 17-17.
- Bowie, J.U., Luthy, R. and Eisenberg, D., 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, *253*(5016), pp. 164-170.
- Brodhun, F., Göbel, C., Hornung, E. and Feussner, I., 2009. Identification of PpoA from Aspergillus nidulans as a fusion protein of a fatty acid heme dioxygenase/peroxidase and a cytochrome P450. *Journal of Biological Chemistry*, *284*(18), pp. 11792-11805.
- Clugston, M. and Flemming, R., 2000. *Advanced chemistry*. Oxford University Press.
- Colovos, C. and Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), pp. 1511-1519.
- Dubey, R.C., 2014. *Advanced biotechnology*. S. Chand Publishing, pp.469-483
- Dunbrack Jr, R.L., Dunker, K. and Godzik, A., 2000. Protein structure prediction in biology and medicine. In *Pacific Symposium on Biocomputing* (Vol. 5, pp. 90-91).
- Fiser, A., 2010. Template-based protein structure modeling. *Computational Biology*, pp.73-94.
- Goodsell, D.S. and Olson, A.J., 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics*, *8*(3), pp. 195-202.
- Graham, S.E. and Peterson, J.A., 1999. How similar are P450s and what can their differences teach us? *Archives of Biochemistry and Biophysics*, *369*(1), pp. 24-29.
- Hardin, C., Pogorelov, T.V. and Luthey-Schulten, Z., 2002. Ab initio protein structure prediction. *Current opinion in structural biology*, *12*(2), pp.176-181.

- Keating, A.E., Malashkevich, V.N., Tidor, B. and Kim, P.S., 2001. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proceedings of the National Academy of Sciences*, *98*(26), pp.14825-14830.
- Kühnel, K., Ke, N., Cryle, M.J., Sligar, S.G., Schuler, M.A. and Schlichting, I., 2008. Crystal structures of substrate-free and retinoic acid-bound cyanobacterial cytochrome P450 CYP120A1. *Biochemistry*, *47*(25), pp. 6552-6559.
- Luthy, R., Bowie, J.U. and Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), p. 83.
- Martí-Renom, M.A., Stuart, A.C., Fiser, A., Sánchez, R., Melo, F. and Šali, A., 2000. Comparative protein structure modeling of genes and genomes. *Annual Review of Bophysics and Biomolecular Structure*, *29*(1), pp. 291-325.
- Metfessel, B.A., Connelly, D.P., Rich, S.S. and Saurugger, P.N., 1993. Cross-validation of protein structural class prediction using statistical clustering and neural networks. *Protein Science*, *2*(7), pp.1171-1182.
- Omura, T. and Sato, R., 1962. A new cytochrome in liver microsomes. *Journal of Biological Chemistry*, *237*(4), pp. PC1375-PC1376.
- Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G. and Aspuru-Guzik, A., 2012. Finding low-energy conformations of lattice protein models by quantum annealing. *Scientific reports*, *2*.
- Sánchez, R. and Šali, A., 2000. Comparative protein structure modeling: introduction and practical examples with modeler. Protein structure prediction: Methods and protocols, pp.97-129.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015.

Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific Reports, 5*, p. 11572.

- Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics, 17*(4), pp. 355-362.
- Sirim, D., Widmann, M., Wagner, F. and Pleiss, J., 2010. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Structural Biology, 10*(1), p. 34.
- Urlacher, V.B. and Eiben, S., 2006. Cytochrome P450 monooxygenases: Perspectives for synthetic application. *Trends in Biotechnology, 24*(7), pp. 324-330.
- Wiederstein, M. and Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*(suppl_2), pp. W407-W410.
- Zhang, Y., 2008. Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology, 18*(3), pp. 342-348.

CHAPTER 4

CLONING AND GENERATION OF RECOMBINANT *ESCHERICHIA COLI* CELLS CONTAINING *SAPROLEGNIA DICLINA*'S *CYP5619A1* IN *pINK-d* EXPRESSION VECTOR

4.1. Introduction

The concept of gene amplification has been recognised over millennia, although limited to phenotypic observation, because of lack of knowledge. By 5000 BC humans had already discovered that they could improve corn crops by sowing the seeds selected from the best plants. Since the establishment of the basic laws of inheritance by the father of genetics, Gregor Mendel, in 1866, the gene has become the cornerstone of most molecular biology technologies. It all evolved into the extraction of the now-known DNA, from the nuclei of leucocytes by Johann Friedrick Miescher in 1869, the establishment of chromosomes as bearers of genetic information by Walter Sutton in 1902, the discovery of the sugar deoxyribose in DNA by Pheobus Levene in 1929 and the finding of the DNA structure by Watson and Crick in 1953. This knowledge has long been used to understand natural cloning or natural reproduction, and to perform artificial cloning processes such as cell cloning, organism cloning and molecular cloning. In molecular cloning the process involves making multiple molecules and is used to amplify DNA fragments containing whole genes, as well as non-coding DNA sequences, randomly fragmented DNA, and DNA sequences such as promoters. These processes evolved with the "cracking" of the genetic code (determining which codon sequences specify each of the 20 amino acids) by Niremberg, Mathaei, and Ochoa in 1966 and the isolation of the enzyme DNA-ligase, the first gene and the first restriction enzyme (RE), in 1967, 1969 and 1970 respectively. In 1972, Paul Berg created the first recombinant DNA. Although naturally occurring genetic processes such as crossing-

overs can technically produce recombinant DNA, the term usually refers to a DNA molecule that results from the combination of two DNAs from different origins, one containing the sequence of the gene of interest, and the other acting as a vehicle or vector that can replicate itself in living host cells. In 1983 Kary B. Mullis developed the polymerase chain reaction (PCR) technique to synthesise DNA rapidly and to be able to transform large amounts into vectors and host cells. The reaction can also be used to confirm the presence of the insert in the cells obtained and thus to guarantee the success of the cloning.

Generally, the cloning of a DNA fragment would involve four steps, namely: (i) the fragmentation of the vector, (ii) the ligation of the gene of interest, (iii) the transfection of the recombinant DNA, or transformation of host cell, and (iv) the screening or selection of successfully transformed host cells. Although these steps are standard in cloning procedures, different cloning strategies can be used, involving a particular strategy for each of the above-mentioned steps. For instance, the cleavage of a vector can be done at a specific recognition site by using REs, which can cut both strands at the same spot, creating blunt ends, or at different spots, creating sticky ends. Furthermore, the transformation of the host cell can be done through different methods: one can allow the entry of the recombinant DNA into the cell by increasing its permeability with chemicals, heat shocks, or an electric field, or by mechanically introducing the vector into the host cell through optical injection or the use of a biolistic particle delivery system.

The choice of expression system is also a crucial part of the cloning process. Some of the expression systems that are most suitable for large-scale production of proteins include the prokaryotic *Escherichia coli* cells, yeast, *baculovirus* infected insect cells, and mammalian cells (the human embryonic kidney – HEK, and the Chinese hamster ovary – CHO). Since each of those expression systems offers advantages and disadvantages, one must take into

account some characteristics such as the cell growth rate, the complexity of the growth medium and its cost, the level of expression, and probable posttranslational modifications in the form of N- or O-linked glycosylation, any phosphorylation, acetylation, acylation or gamma-carboxylation. *E. coli* has proven to be the cheapest, quickest and thus the easiest of these expression systems. It is also the most standardized expression system, with established vectors, strains and protocols commercially available. For these reasons a strain of *E. coli* and its relevant expression system were used in this study to express CYP5619A1.

4.2. Methodology

4.2.1. Target DNA sequence

The sequence of *CYP5619A1* was taken from an article published by the Unit for Drug Discovery Research's laboratory at CUT (Sello et *al.*, 2015).

4.2.2. Novel expression vector

The vector used in this study was initially a kind gift from the late Dr Naheed Kaderbhai, Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, SY23 3DD, United Kingdom. The *pINK*-a vector's multiple cloning site has been modified so as to circumvent the difficulty observed by researchers, in cloning and expressing *Mycobacterium tuberculosis* P450s (Kgosiemang, 2017). A set of steps were carried out to make the vector adequate for cloning and expression. Because the final results are going to be patented and commercial aspects are involved, some features of the expression vector will not be displayed.

4.2.3. Restriction enzymes profiling

The REs were selected from the novel vector's MCS. The cleaving and non-cleaving abilities were assessed using the freely available software pDRAW (http://www.acaclone.com/),

where their recognition sites were explicitly selected on the sequence of the gene of interest. Moreover, the New England Biolabs (NEB) website (https://www.neb.com/products/restriction-endonucleases) was accessed to obtain sequences of the enzymes' recognition sites. Among the three types of restriction endonucleases (Bächi *et al.*, 1979; Kauc & Piekarowicz, 1978; Nathans & Smith, 1975), the use of type II REs was preferred, for they have a predictable pattern and do not require ATP, like types I and III. Moreover, type II REs make cuts directly at the restriction site, unlike type I that makes cuts hundreds of base pairs away from the restriction site (Lautenberger *et al.*, 1978). The purpose of this evaluation was to select two REs belonging to the vector's MCS, which did not have recognition sites on the gene of interest, in order to design the primers.

4.2.4. Primer design

The forward primer was designed by aligning a random, non-coding nucleotide sequence of six base pairs, the RE's sequence, and at least 21 (here 22) of the first base pairs from the 5'- to 3'- sequence of *CYP5619A1*, in that particular order.

The reverse primer consisted of a random, non-coding nucleotide sequence of five base pairs, the RE's sequence, and at least 21 of the first base pairs from the 5'- to 3'- complementary sequence of *CYP5619A1*, in that specific order.

The "Oligo Analyser 3.1" tool in the Integrated DNA Technology (IDT) program (https://eu.idtdna.com/calc/analyzer) was used to analyse the primers and adjust the random nucleotide sequences, in order to have adequate annealing temperatures that match for both forward and reverse primers.

4.2.5. Strains, plasmids, chemicals and kits

E. coli DH5α strain was used in this study. The cells were cultured on Luria-Bertani (LB) broth and LB-agar (LB broth supplemented with 10 g/L agar). For selection of recombinant *E. coli* cells, the LB broth was supplied with antibiotic ampicillin (Amp). LB-Amp plates were prepared by adding Amp at 100 µg/ml final concentration. The Amp stock solution was prepared by dissolving 100 mg of Amp (Catalog No. A6140, Sigma-Aldrich, USA) in 1 ml of DNAse and RNase free water (Catalog No. L3152, Sigma-Aldrich, USA). The Amp stock solution was stored at -20°C. The *pINK-d* expression vector was used for the cloning of *CYP5619A1*. All chemicals used were of high quality and were purchased from Sigma-Aldrich and Merck. The plasmid isolation kit was purchased from Qiagen, USA.

4.2.6. Gene cloning and in-frame analysis

The novel vector's sequence was inserted in the pDRAW software and the DNA property was changed to "circular". The vector was annotated as per the vector's map. The sequence was then truncated at the RE site, and the *CYP5619A1* sequence was inserted and annotated. The Translate tool from ExPASy (http://web.expasy.org/translate/) was used to ensure that the gene of interest remained in frame after cloning.

4.2.7. Gene synthesis

Following the *in silico* cloning of *CYP5619A1* into the vector, the nucleotide sequences were submitted to a gene manufacturing company (GenScript USA Inc, USA), along with the primers' sequences and cloning strategy. The plasmid, *CYP5619A1* and the primers were manufactured, and *CYP5619A1* was cloned into the novel vector, with Res, *KpnI* and *XbaI*.

4.2.8. Transformation

Following the cloning of the gene of interest *CYP5619A1* into the newly-designed vector, the vector obtained with the *CYP5619A1* gene had to be transfected into *E. coli* cells to enable its expression and thus assess the activity.

The host cells (*E. coli* DH5-α) were transformed with the plasmid obtained from the cloning by using the HEPES *E. coli* transformation method described by Inoue and co-workers (1990).

4.2.8.1. Preparation of TB-Buffer

The buffer was prepared by making a 100 ml aqueous solution of 0.26 g of 10 mM HEPES pH 6.7, 0.2206 g of 15 mM $CaCl_2$ and 1.8638 g of 250 mM KCl. The solution was adjusted to pH 6.7 with KOH/HCl. Then 1.0886 g of 55 mM $MnCl_2$ was added and the solution was filtered with a sterile 0.22 µm filter and stored at 4°C.

4.2.8.2. Preparation of SOB medium

The SOB medium was prepared by making a 1500 ml aqueous solution of 10 g of tryptone, 2.5 g of yeast extract, 0.292 g of NaCl, 0.093 g of KCl, 1.017 g of $MgCl_2$ and 1.232 g of $MgSO_4$. The solution was mixed well and divided into 250 ml aliquots in 1 l Erlenmeyer flasks and autoclaved at 121°C for 20 min.

4.2.8.3. Preparation of the SOC medium

The procedure is the same as that for the SOB medium. In addition, 20 ml of sterile 1 M glucose was added after autoclaving.

4.2.8.4. Preparation of competent cells

5 µl of *E. coli* DH5-α glycerol stock was used to inoculate 5 ml LB early in the morning and the inoculum was incubated for 10 hours, while shaking at 37°C. The full 5 ml inoculum was then transferred to 250 ml SOB media in a 1 l flask and grown at 18°C overnight on a shaker, until an OD_{600} of approximately 0.55 had been reached. The flask was then put in an ice-water slurry for 10 min, after which the culture was transferred to a pre-cooled centrifuge bottle and spun at 2500 xg for 10 min at 4°C. The supernatant was decanted and the cells were gently re-suspended in 16 ml of ice-cold TB in the fridge. The suspension was put on ice for 10 min and then centrifuged at 2500 x g for 10 min at 4°C. The cells were gently re-suspended in 4 ml of ice-cold TB. 300 µl of DMSO was added to a final concentration of 7%, and the suspension was put on ice for 10 min. Finally, aliquots of 100 µl per pre-cooled tubes were made, snap-frozen in liquid nitrogen and stored at -80°C.

4.2.8.5. Transformation of *E. coli* DH5-alpha cells

The competent cells were removed from -80°C and thawed on ice. The recombinant plasmid DNA was added to the cells at a concentration of 10% v/v. A flix cube was used to mix it and the mixture was incubated on ice for 20 min. The cells were then immersed in a 42°C water bath for 60 sec, and immediately incubated in ice-water slurry for 2 min. 250 µl of SOC was added and the mixture was incubated at 37°C in a water bath for 1 hour. The pellet cells were then spun at 2500 xg for 5 min and the supernatant was decanted until approximately 100 µl was left. The pellet was re-suspended in the remaining supernatant and plated onto an ampicillin medium. The plates were incubated at 37°C for 20 hours.

4.2.9. Plasmid isolation and purification

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

4.2.10. Restriction enzyme analysis of the plasmids

The isolated plasmids from recombinant *E. coli* cells were subjected to restriction enzyme digestion to confirm the presence of the inserts and the correct size of the cloned *CYP5619A1* cDNA. All restriction enzymes used in this study were purchased from NEB, South Africa. Digested DNA fragments were analysed on a 1% agarose gel. Visualisation of DNA fragments was carried out using SYBR® Sae DNA gel stain (Catalog No. S33102, Thermo Fisher Scientific, USA). The agarose gel was photographed using the Gel Doc™ EZ System (Bio-Rad, South Africa).

4.3. Results and discussion

4.3.1. Re-engineered expression vector and sequence landmarks

The *pINK*-a vector was re-engineered at its MCS. More recognition sites were incorporated initially to be able to clone different P450s (Table 4.1).

The vector was annotated on pDRAW and renamed *pINK-d* (Figure 4.1). *pINK-d* is a circular extra-chromosomal DNA molecule that will be able to replicate independently of the host cell's genome.

Table 4.1. Restriction enzymes incorporated in the multiple cloning site of the expression vector. The restriction enzymes and their recognition sequences are shown in the table.

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

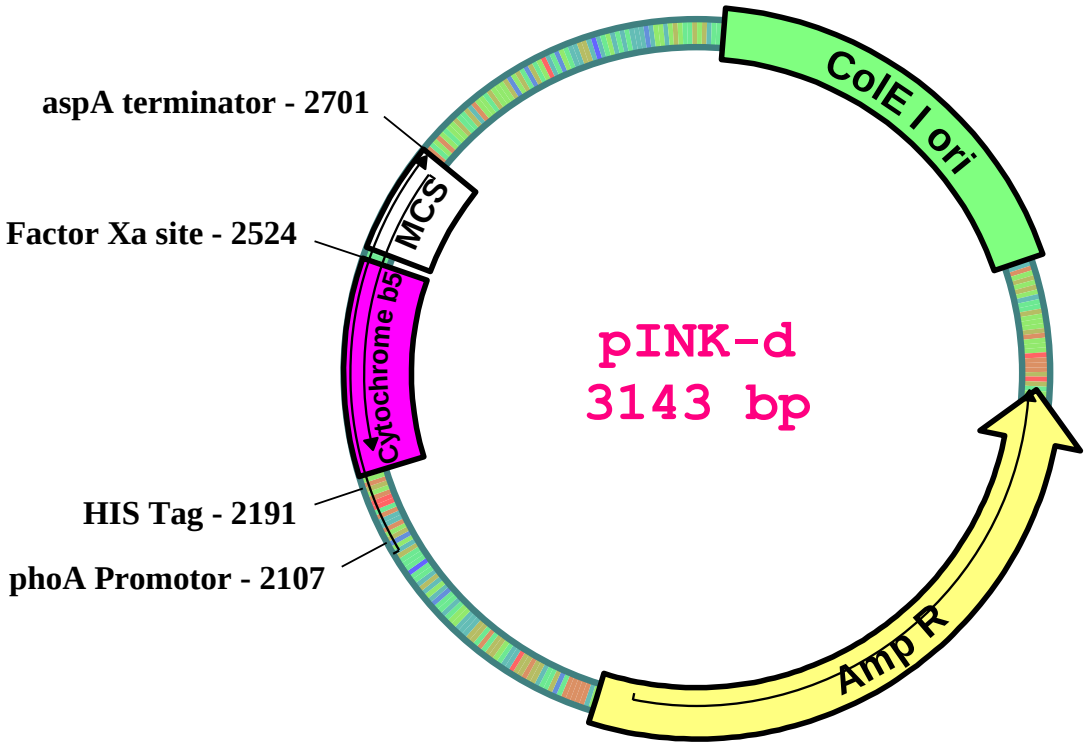


Figure 4.1. Schematic diagram of *pINK-d* expression vector.

4.3.2. Restriction enzymes selection

The selected REs for the cloning of the gene of interest are *KpnI* and *XbaI*. They are part of the vector's MCS, and do not cut *CYP5619A1*.

KpnI recognises the sequence given below:

XbaI recognises the sequence given below:

Both enzymes, *KpnI* and *XbaI*, cause sticky ends restrictions. The cuts are made on the two DNA strands, four base-pairs from each other, creating a four-base 5'-overhang in one molecule and a complementary 5'-overhang in the other. Moreover, their sequences are palindromic, which means that they can be read the same on two strands in the 5'- to 3'- and 3'- to 5'-directions.

4.3.3. *CYP5619A1* primers

Primers are short, single-stranded nucleic acid sequences. When attached to a single-stranded template molecule, they serve as starting point for the synthesis of a complementary DNA strand, directed by a DNA polymerase. The following is the ORF of the gene of interest:

Coding: 5'- ATGGGCAACCTCACCAGCACTGGCGCG […] CCCAACACGATGGAGATGGCCGTGAGCATCTAA -3'

Complementary: 3'- TACCCGTTGGAGTGGTCGTGACCGCGC […] GGGTTGTGCTACCTCTACCGGCACTCGTAGATT -5'

It shows the presence of a start (ATG) and a stop (TAA) codon, as well as the coding (5'- to 3'-) and the complementary (3'- to 5'-) sequences. Restriction site analysis using the software pDRAW showed that the sequence contained neither *KpnI* nor *XbaI* restriction sites.

To design the forward primer, the first 22 nucleotides were selected on the coding sequence of *CYP5619A1*:

5'- ATG GGC AAC CTC ACC AGC ACT G -3'

It is necessary that the 3'- end of a primer be G or C, for they have stronger bonds. In addition, the sequence already begins with a start (ATG) codon, since the gene is being cloned from the very beginning. If cloned from the middle, a start codon would have been added. The *KpnI* site was then added to the 5'- end of the sequence:

5'- **GGT ACC** ATG GGC AAC CTC ACC AGC ACT G

The last step is adding a random, non-coding nucleotide sequence. The reason for this is to have enough support for the enzyme to attach to its restriction site.

5'- NNN NNN **GGT ACC** ATG GGC AAC CTC ACC AGC ACT G

The designing of the reverse primer started in a similar way, by selecting the last 21 bases in the sequence:

5'- TTA GAT GCT CAC GGC CAT CTC C

The sequence that was picked was the complementary sequence of the coding sequence, and the primer was written in the 5'- to 3'-direction for the ordering. The *XbaI* restriction sequence was then added:

5'- **TCT AGA** TTA GAT GCT CAC GGC CAT CTC C

Finally, as on the forward primer, a random, non-coding nucleotide sequence was added at the 5'-end of the restriction site, in order to have enough support for the enzyme to attach.

5'- NNN NNN **TCT AGA** GAT GCT CAC GGC CAT CTC CAT C

The primers obtained were analysed using the "OligoAnalyser 3.1" tool, from the IDT online program (https://eu.idtdna.com/calc/analyzer) and the results (Table 4.2) allowed to observe the melting temperature (Tm) and hence to adjust the annealing temperatures to nearly similar values, by modifying the random sequences at the 5'-end of the primers. Replacing A and T by either G or C would increase the annealing temperatures, because of the stronger bonds of those nucleotides, while replacing G and C by either A or T would decrease the annealing temperatures, because of the weaker bonds of those nucleotides. In addition, the primers were assessed using pDRAW, for the presence of any methylation.

The final designed primers were obtained as follows:

CYP5619A1-FP-KpnI

5'- AAT TAA GGT ACC ATG GGC AAC CTC ACC AGC ACT G -3'

CYP5619A1-RP-XbaI

5'- GC GGC TCT AGA TTA GAT GCT CAC GGC CAT CTC -3'

Table 4.2. Details of the *CYP5619A1* primers. The underlined sections of the sequences are the REs' recognition site.

Primer's code	***CYP5619A1-FP-KpnI***	***CYP5619A1-RP-XbaI***
5'-3' sequence	**AAT TAA <u>GGT ACC</u> ATG GGC AAC CTC ACC AGC ACT G**	**GC GGC <u>TCT AGA</u> TTA GAT GCT CAC GGC CAT CTC**
Oligo concentration	0.25 µM	0.25 µM
Length	34	32
G-C content	50 %	56.2 %
Melting temperature	65.5 °C	65.5 °C
Molecular weight	10404.8 g/mol	9776.4 g/mol

The pDRAW software was used to ensure that the primers did not contain any methylation. Methylation of one or both strands would result in the protection of the enzymes' restriction sites against the binding of those enzymes, thereby inhibiting the enzymes' activity. Although it is not the only possible cause, it is important to verify that factor, to avoid any of the methylation effects (McClelland, 1981). In this case, *KpnI* did not show any methylation, as the enzyme's recognition site is not sensitive to methylation by Dam, Dcm and CpG. On the other hand it was made certain that *XbaI* would not be blocked if Dam overlapped its recognition site.

4.3.4. *In silico* cloning of *CYP5619A1* in *pINK-d*

In silico cloning of *CYP5619A1* into *pINK-d* was carried out using the software pDRAW. A preview of the modified gene was made. From that preview, it was clear that the vector and the gene of interest (*pINK-d* and *CYP5619A1*) could be attached, using both *KpnI* and *XbaI* REs. The plasmid was cleaved by the enzymes, and the insert was attached. The results from the pDRAW software and the preview clearly showed the *pINK-d_CYP5619A1* map (Figure 4.2).

Furthermore, the sequence was assessed to detect any frame shift that could occur after the cloning of *CYP5619A1*. Here, the Translate tool from ExPASy website (http://web.expasy.org/translate/) was used. The sequence of the *pINK-d_CYP5619A1* was entered in the ExPASy Translate tool. The entered sequence goes from the start codon (ATG) following the Shine-Dalgarno sequence, and ends at the stop codon (TGA) of the MCS. The sequence was submitted for translation after requesting the result in a compact output format. The results obtained are displayed in Table 4.3.

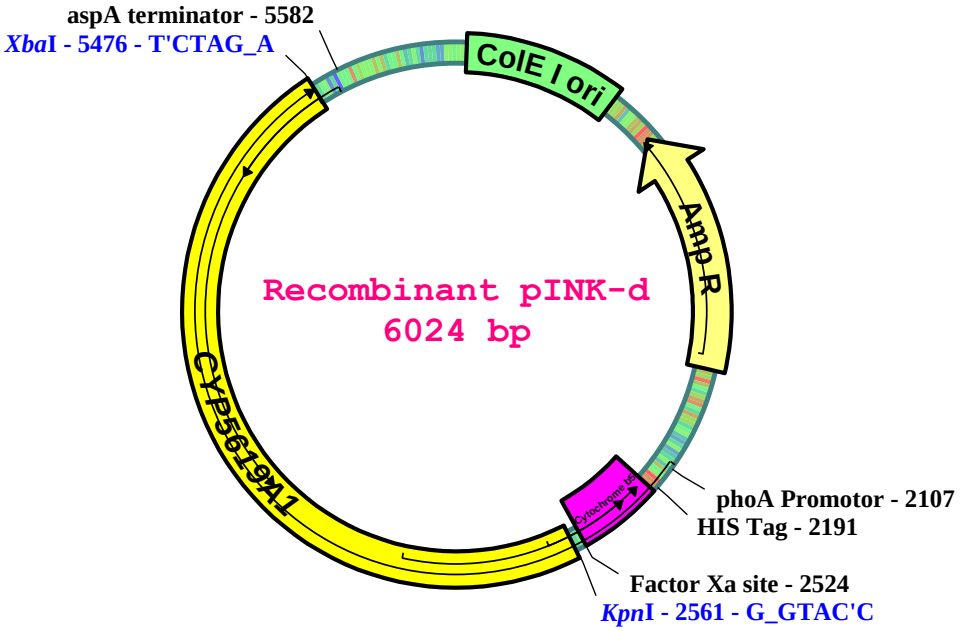


Figure 4.2. ***pINK-d_CYP5619A1*** **recombinant plasmid vector's map.**

Table 4.3. In-frame analysis of *CYP5619A1* in *pINK-d* vector. The highlighted sections of the sequences are the start and the end of the CYP5619A1 amino acids sequence, confirming that no frame shift has occurred during the cloning.

In-frame analysis of *CYP5619A1* in *pINK-d*

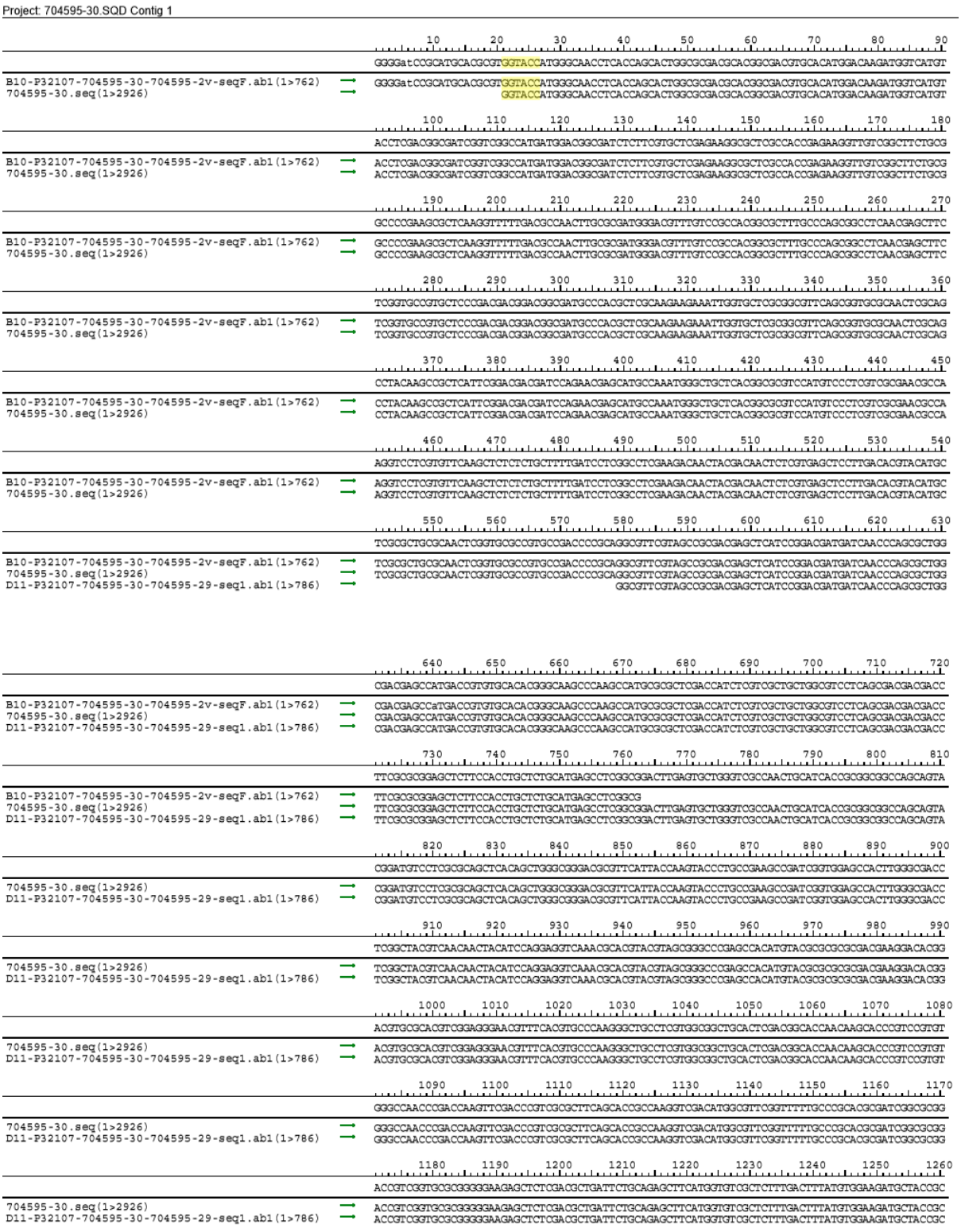
MHHHHHHMAEQSDKDVKYYTLEEIQKHKDSKSTWVILHHKVYDLTKFLEEHPGGEEVLRE
QAGGDATENFEDVGHSTDARELSKTYIIGELHPDDRSKIAKPSETLIMASLQIEGRGGSA
CTRGTMGNLTSTGATHGDVHMDKMVMYLDGDRSAMMDGDLFVLEKALATEKVVGFCGPEA
LKVFDANLRDGTFVRHGALPSGLNELLGAVLPTTDGDAHARKKKLVLAAFSGAQLAAYKP
LIRTTIQNEHAKWAAHGASMSLVANAKVLVFKLSLLLILGLEDNYDNSRELLDTYMLALR
NSVRRADPAGVRSRDELIRTMINPALATSHDRVHTGKPKPCALDHLVAAGVLSDDDLRAE
LFHLLCMSLGGLECWVANCITAAASSTDVLAQLTAGRDAFITKYPAEADRWSHLGDLGYV
NNYIQEVKRTYVAGPSHMYARATKDTDVRTSEGTFHVPKGCLVAAALDGTNKHPSVWANP
TKFDPSRFSTAKVDMAFGFCPHAIGADRRCAGEELSTLILQSFMVSLFDFMWKMLPHQDY
TLDTTLVNPMPKGGLMVVGFHRRTDLSASMVEVAGSEEDWKFLSLPEAKVYRDDKEALHD
MFADERLDLWTHLMLKLLAKKQSMWNKPFANQAITAPKYQKTLPKITLYGLKIQIPTEDE
DWPSDPWNEVATVKFLRDSCPLGDDFEHTWLPGEDMERYVMSKVGSMWPRVNVHWNDRYS
DRALELLVFNGLGQHLVTKLRTAHDDGSYYGICLDFMQALDVRPGYAKYGADAYFNAKGK
VTKIVRLGKTVHPGDEDWEYAKLCFRGSLQTKVTALDHLLGIHITVANGLVTSTREQLPP
THPLRRLLKPFTFRSVIINYNASYALFWPKGMLHRAFSLSVEGMQQTWELGLANFKYETF
PEHKARQNIDTTTLPYHEDGMDFWLIVRGFVGSYIDLYYPCDESLTQDTAVQAFWSYLKT
TLPPNSIRPLSKDNIKDFVAHAIFLVSSMHNHLGTIAEYVSDPAFCPSAWVEGELAGRPG
PCVRGALIMAATGFVQPSIKEDFSHIMLDDAAKAVCRKFTADVCAYAAVVEGRNTKRQHP
YQAFNPNTMEMAVSI-SRPRGS-DRWRAPRRSRTSGGRPAPNCHGSSLQVM

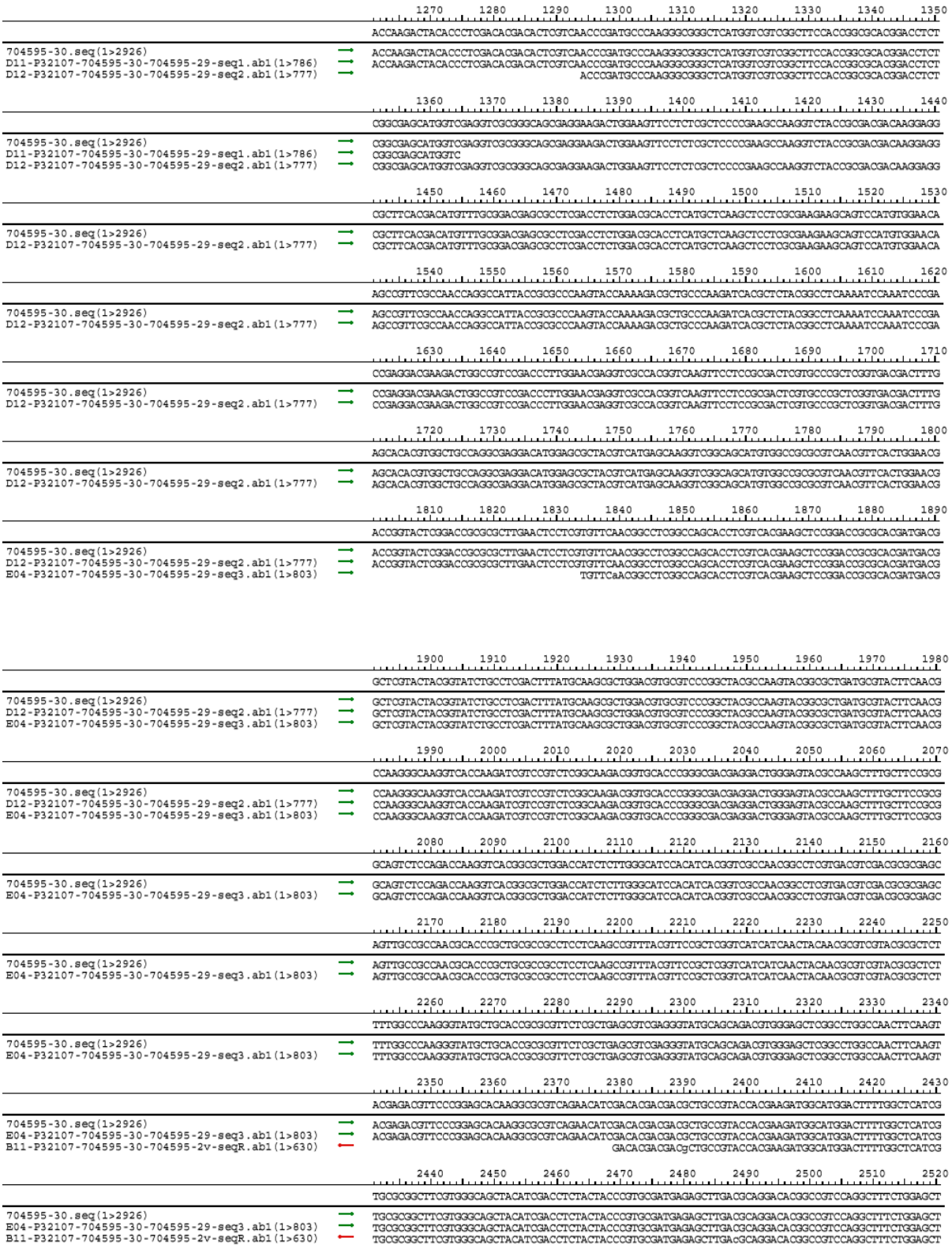
This clearly shows that there has been no frame shift during the *in silico* cloning of *CYP5619A1* and the *CYP5619A1* will be expressed properly.

4.3.5. Synthesis of *pINK-d_CYP5619A1*

The gene sequence was sent to a gene-manufacturing company, GenScript, which is well-known for its good quality products. The vector, *pINK-d*, and the gene, *CYP5619A1*, were synthesised. The latter was cloned into *pINK-d* using *KpnI* and *XbaI* REs (Figure 4.3) and the products were received in a microtiter plate. The constructs and the expression vector were resuspended in 100 µl of Tris-Hcl buffer (pH: 8.0) and 1 µl of this solution was used for transformation.

The restriction sites, highlighted in the figure (Figue 4.3), are linked to the extremities of the gene of interest *CYP5619A1*. This is proof that the gene was cloned perfectly into the vector *pINK-d* and can be expressed properly.





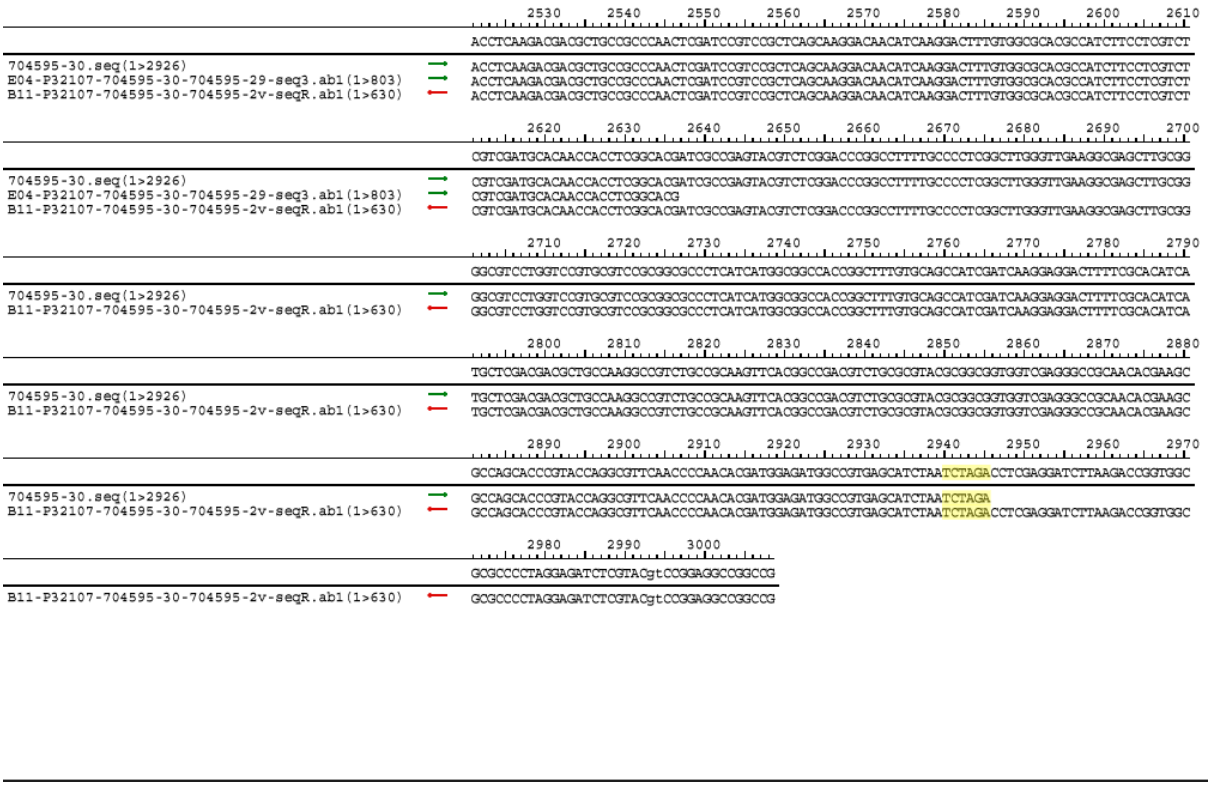
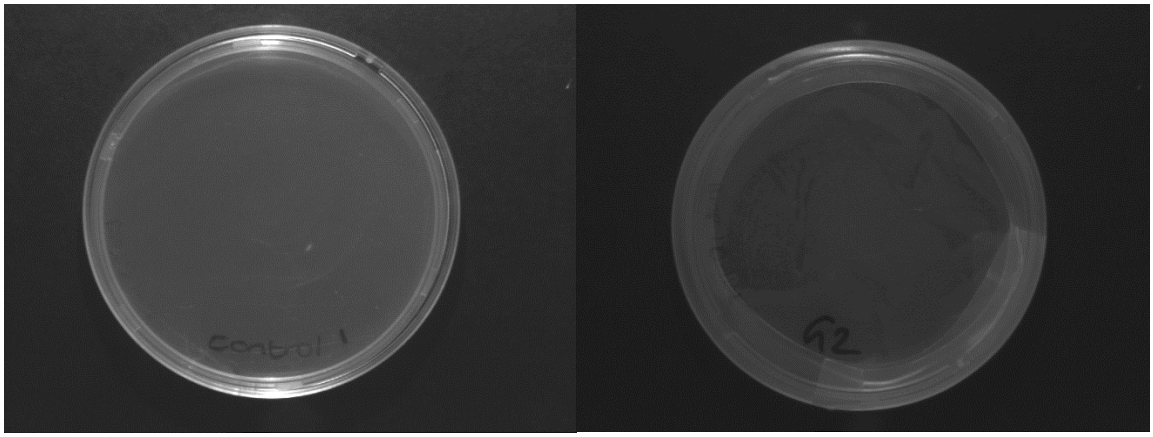


Figure 4.3. Sequencing analysis of *CYP5619A1* in *pINK-d* vector. The figure shows the restriction sites (highlighted) and the sequence of the gene of interest CYP5619A1.

4.3.6. Transformation of the recombinant plasmid into *E. coli* DH5α cells

Competent *E. coli* DH5-α cells were prepared as described in the methodology. These are cells that are able to take up extracellular DNA from their environment. The biochemical mechanisms that underlie that capacity is explained by many authors (Lévy *et al.*, 1990; Dorocicz *et al.*, 1993; MacFadyen *et al.*, 1996; Wang *et al.*, 2002; Bossé *et al.*, 2004; Barabote & Saier, 2005; Bigas *et al.*, 2005; Gioia *et al.*, 2006).

The plasmid *pINK-d_CYP5619A1* was successfully transfected into *E. coli* DH5-α using the *E. coli* transformation procedure described above (Inoue *et al.*, 1990). Some cells took up the plasmids, which contained a gene for Amp resistance. Therefore the cells were screened on an Amp-containing medium (Figure 4.4), which only allowed the growth of the transformed cells. The reason for this growth is that the Amp resistance gene (Amp^R) contained in the plasmid *pINK-d* codes for a β-lactamase: penicillin amido-β-lactamhydrolase, an enzyme that cleaves the β-lactam ring of penicillin and related antibiotics (Sutcliffe, 1978). The amide bond of the β-lactam ring is hydrolysed to produce penicilloic acid, which exhibits no antibiotic activity. The *E. coli* DH5-α strains transformed with *pINK-d_CYP5619A1* produce that enzyme, which is secreted into the periplasmic space of the bacterium, where it performs its activity and detoxifies the Amp in the LB-media.



A B

Figure 4.4. Transformation of *pINK-d_CYP5619A1* into *E. coli* DH5α and screening of the transformed cells on LB medium containing Ampicillin antibiotic. A: Untransformed cells on LB medium with Ampicillin; B: Transformed cells on LB medium with Ampicillin.

4.3.7. Plasmid isolation and confirmation of the presence of *CYP5619A1* cDNA

In order to confirm the presence of the gene of interest in the transformed cells, the latter were isolated and assessed. Firstly, colonies of the transformed *E. coli* DH5α were inoculated into 10 ml of LB broth and incubated at 37°C and 150 rpm. The overnight bacterial culture was used for the isolation of plasmids. The plasmid DNA was extracted using the QIAprep Spin Miniprep Kit and the concentration of the plasmid DNA was measured (Table 4.4) before being subjected to RE (*KpnI* and *XbaI*) digestion analysis.

On lane A, two bands are visible. One represents the plasmids that have a supercoiled conformation. Plasmids with that conformation run faster in agarose gels than they are expected to, since they are not trapped (Akerman, 1998). For that reason, the band appears as that of a 4000 bp DNA sequence, though the plasmid has a 6024-bp long sequence. On the other hand, the nicked, relaxed circular form of the plasmid being the slowest on the gel, it appears between the 6000 and 8000 bands of the KB ladder.

On lane B, two bands are visible. The fastest one appears to be a sequence of nearly 3000 bp, which represents the *CYP5619A1*, with its 2913-bp long sequence. The second, slower one appears to be a sequence of slightly more than 3000-bp long, which is therefore assimilated to the *pINK-d*, with its 3111-bp long sequence. It is important to note that the original vector was 3143-bp long, but lost 32 bp in its MCS from the cleavage by the *KpnI* and *XbaI* REs.

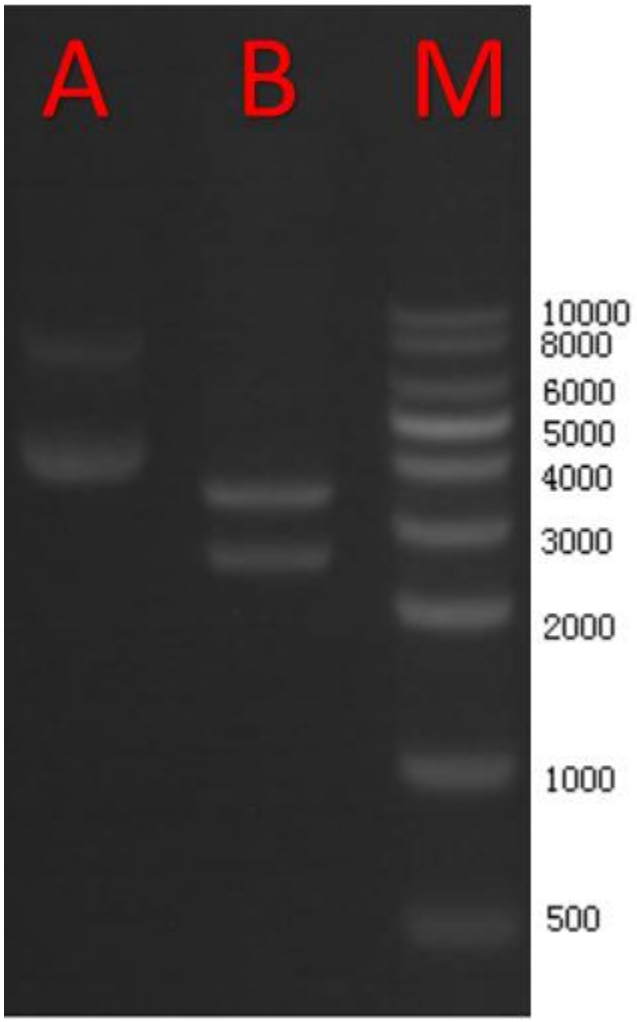


Figure 4.5. Restriction enzyme digestion analysis of *pINK-d_CYP5619A1*. Lane A: undigested plasmid *pINK-d*_CYP5619A1; **Lane B:** digested plasmid (with *KpnI* and *XbaI*); **Lane M:** KB Ladder. Digestion conditions: 300 ng of plasmid digested in water bath, 37°C for 40 min. 1% agarose gel.

4.4. Conclusion

The aim of this chapter was to clone the gene of interest *CYP5619A1* into the newlydesigned vector *pINK-d* using *in silico*, as well as wet-laboratory techniques, in order to be able, at later stages, to express CYP5619A1 and assess the activity of this P450. The gene was cloned into the novel vector using the software pDRAW and the sequence of the designed primers; the vector and gene of interest were sent to GenScript. The results obtained were satisfactory, as the gene was perfectly cloned into the vector, which was verified by running the plasmid through an agarose gel on the one hand, and the plasmid restricted by the *KpnI* and *XbaI* enzymes on the other hand. The gel analysis confirmed the presence of *CYP5619A1*.

References

- Åkerman, B., 1998. Effects of supercoiling in electrophoretic trapping of circular DNA in polyacrylamide gels. *Biophysical Journal, 74*(6), pp. 3140-3151.
- Bächi, B., Reiser, J. and Pirrotta, V., 1979. Methylation and cleavage sequences of the EcoP1 restriction-modification enzyme. *Journal of Molecular Biology, 128*(2), pp. 143-163.
- Barabote, R.D. and Saier, M.H., 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiology and Molecular Biology Reviews, 69*(4), pp. 608-634.
- Bigas, A., Garrido, M.E., de Rozas, A.M.P., Badiola, I., Barbé, J. and Llagostera, M., 2005. Development of a genetic manipulation system for *Haemophilus parasuis*. *Veterinary Microbiology, 105*(3), pp. 223-228.
- Bossé, J.T., Nash, J.H., Kroll, J.S. and Langford, P.R., 2004. Harnessing natural transformation in *Actinobacillus pleuropneumoniae*: A simple method for allelic replacements. *FEMS Microbiology Letters, 233*(2), pp. 277-281.
- Dorocicz, I.R., Williams, P.M. and Redfield, R.J., 1993. The *Haemophilus influenzae* adenylate cyclase gene: Cloning, sequence, and essential role in competence. *Journal of bacteriology, 175*(22), pp. 7142-7149.
- Gioia, J., Qin, X., Jiang, H., Clinkenbeard, K., Lo, R., Liu, Y., Fox, G.E., Yerrapragada, S., McLeod, M.P., McNeill, T.Z. and Hemphill, L., 2006. The genome sequence of *Mannheimia haemolytica* A1: Insights into virulence, natural competence, and *Pasteurellaceae* phylogeny. *Journal of Bacteriology, 188*(20), pp. 7257-7266.
- Inoue, H., Nojima, H. and Okayama, H., 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene, 96*(1), pp.23-28.

- Kauc, L. and Piekarowicz, A., 1978. Purification and properties of a new restriction endonuclease from *Haemophilus influenzae* Rf. *European Journal of Biochemistry, 92*(2), pp. 417-426.
- Kgosiemang, I.K.R., 2017. Construction of world's first *Mycobacterium tuberculosis* cytochrome P450 monooxygenase expression library. Master's thesis, Department of Health Sciences, Central University of Technology, Bloemfontein, Free State, South Africa.
- Lautenberger, J.A., Kan, N.C., Lackey, D., Linn, S., Edgell, M.H. and Hutchison, C.A., 1978. Recognition site of *Escherichia coli* B restriction enzyme on ϕXsB1 and simian virus 40 DNAs: An interrupted sequence. *Proceedings of the National Academy of Sciences, 75*(5), pp. 2271-2275.
- Lévy, S., Zeng, G.Q. and Danchin, A., 1990. Cyclic AMP synthesis in *Escherichia coli* strains bearing known deletions in the pts phosphotransferase operon. *Gene, 86*(1), pp. 27-33.
- Macfadyen, L.P., Dorocicz, I.R., Reizer, J., Saier Jr, M.H. and Redfield, R.J., 1996. Regulation of competence development and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate: Fructose phosphotransferase system. *Molecular Microbiology, 21*(5), pp. 941-952.
- McClelland, M., 1981. The effect of sequence specific DNA methylation on restriction endonuclease cleavage. *Nucleic Acids Research, 9*(22), pp. 5859-5866.
- Nathans, D. and Smith, H.O., 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annual Review of Biochemistry, 44*(1), pp. 273-293.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015.

Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific Reports, 5*, p. 11572.

- Sutcliffe, J.G., 1978. Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. *Proceedings of the National Academy of Sciences, 75*(8), pp. 3737-3741.
- Wang, Y., Goodman, S.D., Redfield, R.J. and Chen, C., 2002. Natural transformation and DNA uptake signal sequences in *Actinobacillus actinomycetemcomitans*. *Journal of Bacteriology, 184*(13), pp. 3442-3449.

CHAPTER 5

CONCLUSION AND FUTURE PERSPECTIVES

In conclusion, in this study, phylogenetic analysis and in silico structural and functional analysis of the novel P450 family CYP5619 from oomycetes have been carried out. This study revealed that the CYP5619 family is not limited to oomycetes, but is also present in other microbes. *In silico* structural and functional analysis revealed that CYP5619A1 is predicted to be capable of oxidising fatty acids, linoleic acid and to tightly bind to malachite green, the agent used to kill oomycetes. Future study could include the expression and functional analysis of CYP5619A1, including assessing the inhibiting effect of malachite green.

RESEARCH OUTPUTS

The research done during the doctoral programme produced conference outputs, as well as articles in peer-reviewed journals.

1. Conference abstract/poster presentations

Results from this study have been presented in the form of posters at national and international conferences detailed below:

1. **Bamal HD,** Mashele SS, Tuszynski JA, Syed K (2017). Phylogenetic and comprehensive *in silico* structural and functional analysis of CYP5619 from *Saprolegnia diclina*. The Annual South African Pharmacology Conference, Faculty of Health Sciences, University of the Free State, Bloemfontein, 01st – 04th October.

2. **Bamal HD,** Mashele SS, Tuszynski JA, Syed K (2017). *In silico structural and functional characterisation of a novel cytochrome P450 CYP5619A1 from Saprolegnia diclina*. 20th International Conference on Cytochrome P450s: Biochemistry, Biophysics and Biotechnology, August 27-31, Dusseldorf, Germany.

3. **Bamal HD,** Mashele S, Syed K (2016) Structural analysis of a novel P450 family CYP5619 from pathogenic oomycete *Saprolegnia diclina*. The 13th International Symposium on Cytochrome P450 Biodiversity and Biotechnology, 22-26 July 2016, Vancouver, BC, Canada. P72-S2.

4. **Bamal HD,** Mashele S, W Chen, Syed K (2015) Understanding the P450 subfamily evolution in biotechnologically valuable and catalytically versatile P450 families CYP63 and CYP5136. International Symposium on Methods for Studying Drug Metabolism and Transport, and African Traditional Medicines (METHODS-2015) from 23-25 November 2015 at St Georges Hotel and Conference Center, Pretoria, South Africa. DMP07.**6.1.**

2. Publications and supervisions

An article detailing the study findings is in preparation for submission to the journal "Marine Drug" (ISSN 1660-3397; IF 3.503) and will be as follows:

Bamal HD, Mashele SS, Tuszynski JA, Syed K (2017) *In silico* structural and functional analysis of novel P450 family CYP5619 from deadliest oomycetes pathogen *Saprolegnia diclina*. *Marine Drugs* (soon to be communicated).

During the course of the program, **four B Tech projects** and **three Masters projects** (three students) were supervised. In view of this contribution, co-authorship was earned in articles in high-impact factor journals such as Scientific Reports and BBA Proteins and Proteomics. Details of the manuscripts are as follows:

1. **Bamal HD** (co-author) (2017) *In silico* analysis of cytochrome P450 monooxygenases in chronic granulomatous infectious fungus *Sporothrix schenckii*: special focus on CYP51. BBA Proteins and Proteomics (BBAPRO-17-174R1).

2. **BAMAL HD** (co-author) (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: special focus on mycobacterial P450s. *Scientific reports*, 6, p.33099.