Candida CYP52: Alkane and fatty acid metabolism.

Price, Claire Louise

Candida CYP52: Alkane and Fatty Acid Metabolism

By

Claire Louise Price

Submitted to the University of Wales in fulfilment of the requirements for the degree of Doctor of Philosophy

Swansea University

2012

Summary

Cytochromes P450 are a superfamily of haem-thiolate proteins found in all kingdoms of life. To date 11294 enzymes have been identified and have been shown to be involved in the metabolism of a wide variety of substrates, including hydrocarbons and xenobiotics. In yeast and fungi the hydroxylation of alkanes is associated with a family of cytochromes P450 enzymes known as CYP52s. These enzymes are involved in the terminal hydroxylation of long-chain alkanes resulting in the production of alcohols, which can be further converted to form fatty acids and diacids. *In vivo* such hydrocarbons can be subjected to β-oxidation for use in growth. Alternatively, the products formed by CYP52 catalysed hydroxylation *in vitro* can be used in biotechnological applications. They can be used as platform chemicals in the production of a number of industrial products, including plastics, fragrances and antibiotics.

The β-oxidation of fatty acids has been less well documented for *Candida albicans* than for other *Candida* species, therefore it was the aim of this study to investigate a) did cytochromes P450 exist in *C. albicans* that could possibly fulfil this function and b) to definitively assign function to a single cytochrome P450. Using a bioinformatic approach, five putative CYP52s were identified in *C. albicans*. Of these CYP52s, Alk1 was shown to have the greatest homology to the archetypal alkane-assimilating CYP52, CYP52A3 from *C. maltosa*. *ALK1* heterologous gene expression in the brewer's yeast *Saccharomyces cerevisiae* allowed growth on hexadecane (C16:0) as the sole carbon source. This showed for the first time that Alk1 is involved in the hydroxylation of long-chain alkanes as normally *S. cerevisiae* is unable to utilise alkanes for growth. This study has also shown that Alk1 is able to interact with sterol substrates suggesting a possible role for CYP52s in sterol metabolism, which was previously unknown.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date ..27/1/12.....

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date ..27/1/12.....

Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date ...27/1/12.....

Contents

Acknowledgements

I would like to thank my supervisors Prof. Steve Kelly and Dr. Diane Kelly for giving me the opportunity to study such a fascinating and complex subject, as well as their continued help and guidance throughout this PhD. I would also like to thank the members of the cytochrome P450 group, particularly Dr. Claire Hull, Dr. Josie Parker and Dr. Andrew Warrilow, for all their help, support and friendship, which has helped me through the bad days and let me celebrate the good. I would like to thank the rest of the fifth floor for making everyday interesting. I have had the great fortune of forging friendships with individuals on this floor that I never thought possible at the start of this journey and for that I thank them.

I would like to thank my parents for their unconditional love and support, which has helped me achieve things I never thought possible. I would also like to thank the rest of my family and friends for being unwitting distractions and reminding me, in times of despair, why I started this degree. Thank you.

List of Figures

List of Tables

Abbreviations

5,8-DiHODE	5,8-dihydroxyoctadecadienoic acid
(8R)-HPODE	(8R)-hydroxyperoxyoctadecadienoic acid
11,12-EET	11,12-epoxy-5,8,14-eicosatrienoic acid
19-HETE	19-hydroxyeicosatetraenoic acid
20-HETE	20-hydroxyeicosatetraenoic acid
ACP	Acyl Carrier Protein
Ala	5-aminolevulinic acid hydrochloride
APS	Ammonium persulphate
BCD	Bietti's corneoretinal crystalline dystrophy
BLAST	Basic Local Alignment Sequence Tool
BSTFA-TCMS	*N*,*O*-bis(trimethylsilyl)trifluoroacetamide-trimethylchloro-silane
CaCPR	NADPH-Cytochrome P450 reductase from *Candida albicans*
CAT	Carnitine Acetyltransferase
CGD	*Candida* Genome Database
CO	Carbon monoxide
CoA	Coenzyme A
CPR	NADPH-Cytochrome P450 reductase
DHA	Docosahexaenoic acid
DLPC	Dilaurylphosphatidylcholine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
EPA	Eicosapentaenoic acid
FAD	Flavin adenine dinucleotide (ferredoxin reductase)
FdR	Ferredoxin reductase
FdX	Iron sulphur redoxin
FMN	Flavin mononucleotide (flavodoxin)
GCMS	Gas Chromatography Mass Spectrometry
hCPR	Human NADPH-Cytochrome P450 reductase
HMMs	Hidden Markov Models
HPLC	High Pressure Liquid Chromatography
IPTG	Isopropyl-β-D-thio-galactoside

K_s	Substrate binding constant
LB	Luria-Bertani media
LCMS	Liquid Chromatography Mass Spectrometry
NGS	Next-Generation Sequencing
ORF	Open Reading Frame
PAH	Polycyclic Aromatic Hydrocarbon
PCR	Polymerase Chain Reaction
PIR	Protein Information Resource
PSI	Precocious Sexual Inducer factor
RDX	Royal Demolition Explosive (hexa-hydro-1,3,5-trinitro-1,3,5-triazine)
ScΔ33:CPR	Truncated *Saccharomyces cerevisiae* NADPH-Cytochrome P450 reductase
SGD	*Saccharomyces* Genome Database
TB	Terrific Broth
TEMED	N,N,N',N',-tetramethylethylene diamine
TLC	Thin Layer Chromatography
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter 1: Introduction

1.1 *Candida albicans*

Candida albicans is a diploid fungus, which can exist in three different morphological forms: yeast, pseudohyphae and hyphae (figure 1.1).

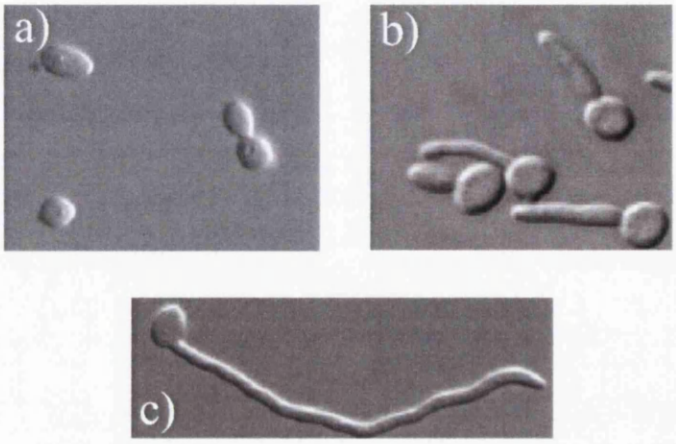


Figure 1.1 – The three morphologies of *C. albicans*: yeast, pseudohyphae and hyphae (taken from Sudbery *et al.*, 2004). a) *Yeast* – In this unicellular form, a new daughter cell buds off the mother cell to form two cells (Sudbery *et al*, 2004; Whiteway & Bachewich, 2007). b) *Pseudohyphae* – Pseudohyphae form when a daughter bud remains fused to the mother cell. Once the septum has formed, the daughter bud elongates to resemble true hyphae (Sudbery *et al.*, 2004). c) *Hyphae (also known as true hyphae)* – Rapid switching between the yeast and hyphal forms of *C. albicans* has been implicated in the pathogenesis of this fungus (Brown & Gow, 1999). It is thought the hyphal form allows *C. albicans* to penetrate and invade the host tissues leading to infection (Gow, 1994).

C. albicans can be located harmlessly in the flora of healthy human gastrointestinal and genitourinary tracts and on the skin (Berman & Sudbery, 2002). However, changes in the host environment (such as changes in pH or glucose concentration) or a weakened immune system (as a result of illnesses, such as cancer or AIDS) can result in the manifestation of superficial or systemic infections (López-Martínez, 2010). Superficial candidosis can manifest as oral thrush (figure 1.2a), a rash/itching in the creases of the body, such as the armpits and in between the fingers (figure 1.2b), vulvovaginal thrush or paronychia, which results in soreness and discoloration of the (finger)nails (figure 1.2c) (López-Martínez, 2010). In patients with a

suppressed/weakened immune system, such as those undergoing chemotherapy, *C. albicans* can enter the bloodstream and cause potentially life threatening infections (Sudbery *et al.*, 2004).

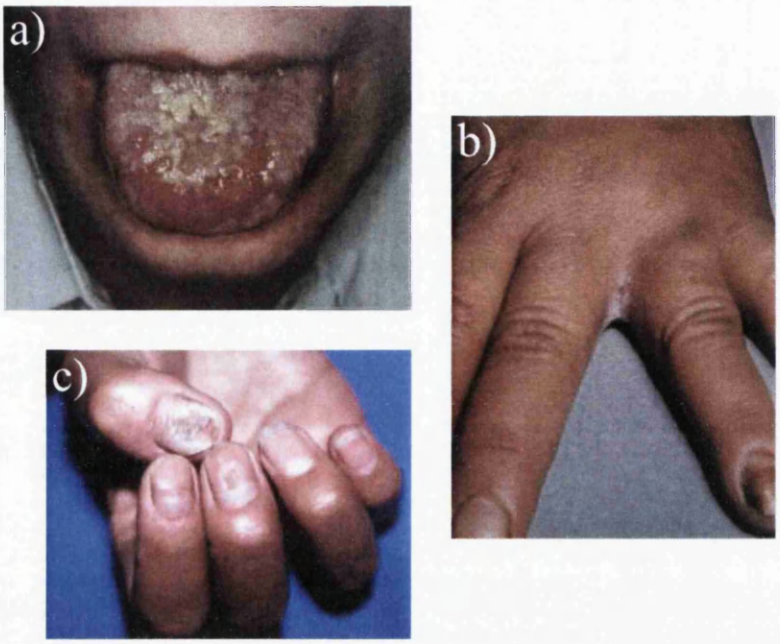


Figure 1.2 – Examples of superficial *Candida* infections (taken from López-Martínez, 2010). a) Oral thrush – most common infection. b) Infection in the crevices between the fingers. c) Paronychia – painful irritation of the fingernails.

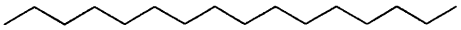
Although *C. albicans* is the major pathogen of fungal infections in humans, other *Candida* spp. are also emerging as causes of these infections due in part to the over-prescription of antifungal agents by doctors (Kothavade *et al.*, 2010; López-Martínez, 2010). These non-*albicans* species include: *C. dubliniensis*, *C. krusei*, *C. lipolytica* and *C. tropicalis*, which is the major non-*albicans* species causing fungal infections in humans (Kothavade *et al.*, 2010; López-Martínez, 2010). Interestingly, *C. tropicalis* has not only been shown to be involved in pathogenesis, but it has also been shown to have potential biotechnological applications too. *C. tropicalis* is able to grow on hydrocarbons, such as alkanes (figure 1.3) and fatty acids (figure 1.4), when they are used as the sole carbon source. The derivatives produced through metabolism by cytochrome P450 enzymes can be used as platform chemicals to form the basis of a number of industrial products, including hot-melting adhesives, plastics, lubricants, fragrances and antibiotics (Eschenfeldt *et al.*, 2003; Liu *et al.*,

2003). Other *Candida* spp., including the soil-dwelling, non-pathogenic fungus *C. maltosa*, have also been shown to have similar capabilities.

Dodecane (C12:0)

Tetradecane (C14:0)

Hexadecane (C16:0)

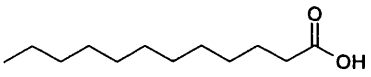


Octadecane (C18:0)

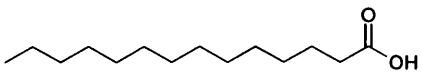
Figure 1.3 – Examples of different alkane compounds. Alkanes are made up of the chemical elements carbon and hydrogen and have the molecular formula C_nH_{2n+2}.

Saturated Fatty Acids

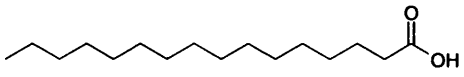
Lauric Acid (C12:0)



Myristic Acid (C14:0)

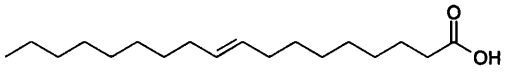


Palmitic Acid (C16:0)



Unsaturated Fatty Acids

Oleic Acid (C18:1)



Linoleic Acid (C18:2)

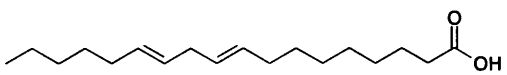


Figure 1.4 – Examples of different saturated and unsaturated fatty acids. Fatty acids contain an alkyl chain joined to a carboxyl (COOH) group and exist in two forms: saturated and unsaturated. Saturated fatty acids do not contain any double bonds, whereas unsaturated fatty acids contain at least one.

1.2 Cytochromes P450

Cytochromes P450 are a superfamily of haem-thiolate proteins, which catalyse the oxidation of a wide variety of structurally diverse, organic compounds, such as fatty acids, steroids and xenobiotics. They were first observed in liver microsomes in 1958 by Garfinkel and Klingenberg (independently) and were described as an unusual carbon monoxide binding pigment with an absorbance maximum at 450nm (Garfinkel, 1958; Klingenberg, 1958). However, it was not until 1964 that Omura and Sato characterised the pigment as a cytochrome protein (Omura & Sato, 1964). Cytochromes P450 get their name from the Soret peak, which is produced at a wavelength of 450nm when carbon monoxide binds to the sodium dithionite reduced haem protein (Omura & Sato, 1964).

Cytochromes P450 have presently been found in all kingdoms of life (animals, lower eukaryotes, plants and bacteria) and currently, there are 11294 enzymes (February 2011) in the cytochrome P450 superfamily (http://drnelson.utshc.edu/CytochromeP450.html). This number, however, is continually increasing and a nomenclature system has been implemented to identify these genes.

Cytochromes P450 can be identified by the CYP prefix, e.g. **CYP**4 (Nebert *et al.*, 1989, Nelson, 2006). This is followed by a number, which represents the family to which the cytochrome P450 belongs. Individual genes become part of this family if they share more than 40% amino acid identity with the other cytochromes P450 already in that family, e.g. CYP**4** (Nelson *et al.*, 1993; Nelson *et al.*, 1996; Werck-Reichhart & Feyereisen, 2000; Kelly *et al.*, 2003; Schuler & Werck-Reichhart, 2003). The cytochromes P450 can then become part of a specific subfamily if they share more than 55% amino acid identity with the other cytochromes P450 and this is represented as a letter in the nomenclature, e.g. CYP4**F** (Werck-Reichhart & Feyereisen, 2000; Kelly *et al.*, 2003; Schuler & Werck-Reichhart, 2003). The final number, e.g. CYP4F**11**, is the unique gene identifier for that specific cytochrome P450 (Werck-Reichhart & Feyereisen, 2000). Figure 1.5 shows an overview of this nomenclature system.

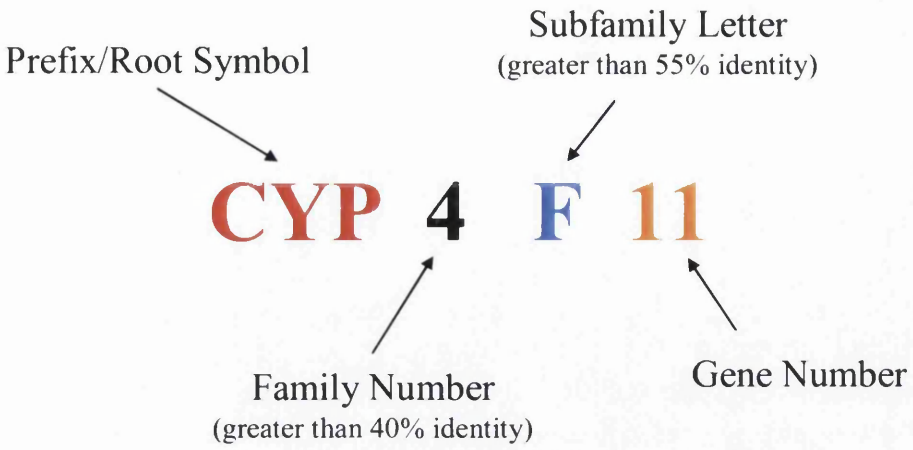


Figure 1.5 - Overview of the cytochrome P450 nomenclature system. The example used is CYP4F11. CYP represents the prefix/root symbol. The first number (4) is the family number (which requires greater than 40% identity amongst members). The subfamily is denoted by a letter (F) and requires members to have greater than 55% identity. The final gene number (11) is a unique gene identifier.

1.2.1 Cytochrome P450 Monooxygenase System

Cytochromes P450 are involved in the catalysis of many different reactions, including carbon hydroxylation (alcohol formation), heteroatom release (also known as dealkylation) (oxidative cleavage at atoms on the molecule that are neither carbon nor hydrogen), heteroatom oxygenation (formation of a heteroatom oxide) and epoxidation (formation of epoxides from the addition of oxygen to alkenes and aromatic compounds) (Guengerich & MacDonald, 1984; Porter & Coon, 1991; Groves, 2005; Guengerich, 2007) (see figure 1.6).

a) Hydroxylation

b) Heteroatom Release

c) Heteroatom Oxygenation

d) Epoxidation

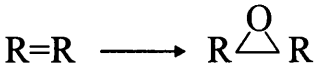


Figure 1.6 – Reactions catalysed by cytochromes P450. a) *Hydroxylation.* Alcohol formation. b) *Heteroatom release.* Oxidative cleavage at atoms on the molecule that are neither carbon nor hydrogen. c) *Heteroatom oxygenation.* Formation of a heteroatom oxide. d) *Epoxidation.* Formation of epoxides from the addition of oxygen to alkenes and aromatic compounds. R represents a hydrocarbon side-chain, O represents oxygen and X represents a halogen molecule.

For each of these described reactions to occur an oxygen atom needs to be incorporated into the substrate to produce the required product. This is achieved by the substrate being subjected to the cytochrome P450 monooxygenase catalytic cycle (figure 1.7). The cycle begins with the haem iron centre (Fe) in its resting state (1). The organic substrate (RH) docks near the haem centre (2), enhancing the cytochrome P450's ability to accept electrons (Porter & Coon, 1991) from an associated electron donor (e.g. NADPH-cytochrome P450 reductase) (see section 1.2.2) (Groves, 2005). This electron transfer causes the haem centre to become reduced (3) allowing molecular oxygen to bind (4) producing a cytochrome P450-dioxygen complex (Groves, 2005). A second electron is transferred from the associated electron donor (5) to cleave the oxygen, incorporating one oxygen atom into the organic substrate and allowing the second to be released as water (Groves, 2005). The monooxygenase cycle is completed when product dissociation occurs (6) and the haem centre is restored back to its original resting state (Porter & Coon, 1991; Groves, 2005).

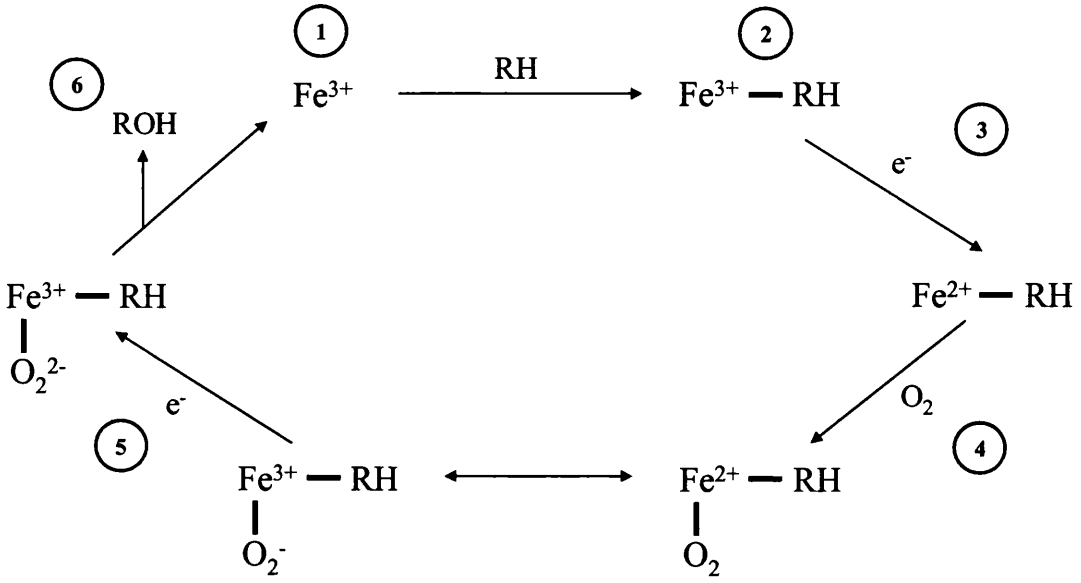


Figure 1.7 – Cytochrome P450 monooxygenase system. Catalytic cycle showing the steps required to incorporate oxygen into the substrate (RH) at the haem centre (Fe). RH = organic substrate, Fe = haem centre, O = oxygen and ROH = oxidised state of the organic substrate.

1.2.2 Classes of Cytochromes P450

For the monooxygenase cycle to progress, electrons are required from an associated electron donor. However, the electron donor required for this process differs depending on the cytochrome P450 enzyme being used. Therefore, cytochromes P450 are separated into ten different classes based on this criterion.

1.2.2.1 Class I

Class I proteins encompass bacterial and mitochondrial cytochromes P450, which require an FAD-containing reductase (ferrodoxin reductase, FdR) as well as an iron sulphur redoxin (iron-sulphur-cluster, Fdx) for the transfer of electrons to occur (Werck-Reichhart and Feyereisen, 2000; Paine *et al.*, 2005). In bacterial class I systems (figure 1.8a), each of these components are soluble, however, in mitochondrial systems (figure 1.8b) the Fdx is soluble (found in the mitochondrial matrix), but the cytochrome P450 and FdR are membrane-bound (bound to the inner mitochondrial membrane) (Hannemann *et al.*, 2007). Electrons are transferred in both classes from NAD(P)H to FdR to Fdx and then finally to the cytochrome P450 for completion of the monooxygenase cycle (Werck-Reichhart & Feyereisen, 2000; Paine *et al.*, 2005). Examples of members of this class are the bacterial cytochrome

P450 CYP101D1 from *Novosphingobium aromaticivorans*, which is involved in the synthesis of camphor to 5-exo-hydroxycamphor (Bell *et al.*, 2010) and the mammalian mitochondrial cytochrome P450 CYP11A1 (P450scc) involved in steroid hormone biosynthesis (Black *et al.*, 1994).

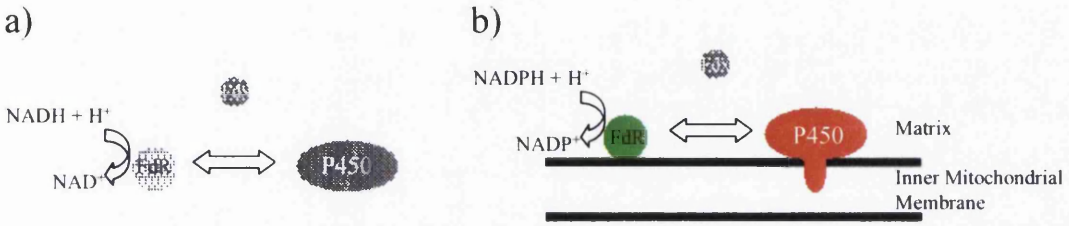


Figure 1.8 – Class I cytochromes P450 (adapted from Hannemann *et al.*, 2007). a) *Bacterial* – all components are soluble. b) *Mitochondrial* – soluble iron sulphur redoxin (Fdx) and membrane-bound cytochrome P450 complex and ferredoxin reductase (FdR).

1.2.2.2 Class II

The cytochromes P450, which make up class II are microsomal proteins that require NADPH-cytochrome P450 reductase (CPR) for electron transfer (figure 1.9) (Werck-Reichhart & Feyereisen, 2000; Paine *et al.*, 2005). CPR is a flavoprotein containing equal amounts of the cofactors FAD (ferredoxin reductase) and FMN (flavodoxin) (He & Chen, 2005; van Bogaert *et al.*, 2007), which are required for electron transfer from NADPH. Initially the hydride ion of NADPH transfers electrons to FAD then a single electron is transferred to FMN from FAD (Gigon *et al.*, 1969; van Bogaert *et al.*, 2007). The electron is transferred to the haem centre of the cytochrome P450. The haem iron then becomes reduced allowing molecular oxygen activation and the eventual formation of the alcohol complex (van Bogaert *et al.*, 2007). A member of this class of proteins is the sterol 14α-demethylase CYP51 (Lamb *et al.*, 2001).

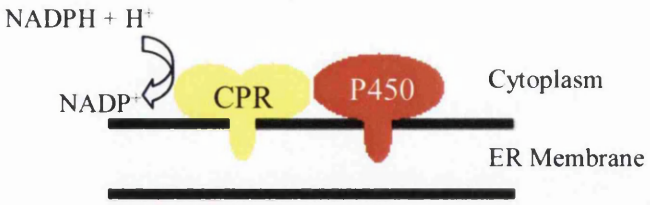


Figure 1.9 – Class II cytochromes P450 (adapted from Hannemann *et al.*, 2007). Membrane-bound NADPH-cytochrome P450 reductase (CPR) and cytochrome P450 complex.

1.2.2.3 Class III

Class III is a bacterial system used by CYP176A1 ($P450_{cin}$ - involved in the biodegradation of cineole 1) from *Citrobacter braakii* (figure 1.10) (Hawkes *et al.*, 2002; Meharenna *et al.*, 2004). It resembles class I as both use a (soluble) three component system to transfer electrons from NAD(P)H. However, unlike the class I system, class III proteins utilise FMN rather than the iron-sulphur-cluster (Fdx) to deliver electrons from FAD to the haem iron (Hannemann *et al.*, 2007). Electron transfer between the redox centres of FAD and FMN also resembles the system utilised by class II proteins, but FAD and FMN are bound to two separate proteins in the class III system whereas they are bound to one protein, CPR, in class II (Hannemann *et al.*, 2007). Although, the electron transfer system utilised by CYP176A1 shares aspects of both the class I and class II systems, it does not fully satisfy the remits of either thus giving rise to a separate group of proteins classified as class III.

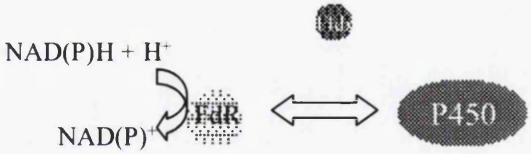


Figure 1.10 – Class III cytochromes P450 (adapted from Hannemann *et al.*, 2007). Soluble, bacterial system utilising FAD and FMN.

1.2.2.4 Class IV

The electron transfer system associated with CYP119 from *Sylfolobus acidocaldarius* belongs to class IV (figure 1.11) (Lim *et al.*, 2010). CYP119 is a thermophilic cytochrome P450, which can remain stable at high temperatures (up to 85°C) (Koo *et al.*, 2000) and unlike the other classes it is not NAD(P)H dependent (Puchkaev & Ortiz de Montellano, 2005). Instead of using NAD(P)H, electrons are transferred from pyruvic acid to a 2-oxoacid-ferredoxin oxidoreductase and then to a thermostable ferredoxin before being delivered to the haem centre of the cytochrome P450 (Puchkaev & Ortiz de Montellano, 2005).

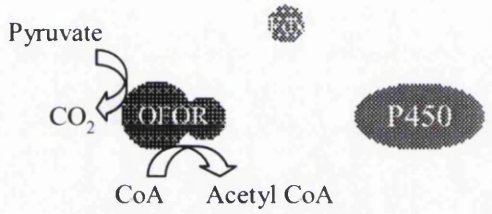


Figure 1.11 – Class IV cytochromes P450 (adapted from Hannemann *et al.*, 2007). Novel thermophilic system.

1.2.2.5 Class V

CYP51 (a 14α-demethylase) from *Methylococcus capsulalus* is the sole member of class V (figure 1.12) (Jackson *et al.*, 2002). Like the class I system, it utilises FdR and Fdx, however, this is a novel system as the Fdx is fused to the C-terminus of the cytochrome P450 complex (Hannemann *et al.*, 2007). Therefore, in this system, electrons are transferred from NADH to the soluble FdR before being passed onto the FdX-cytochrome P450/CYP51 fusion protein (Hannemann *et al.*, 2007).

Figure 1.12 – Class V cytochromes P450 (adapted from Hannemann *et al.*, 2007). Fdx-cytochrome P450 fusion protein.

1.2.2.6 Class VI

The class VI system also utilises a novel fusion cytochrome P450 fusion protein in the passage of electrons from NADPH (figure 1.13) (Hannemann *et al.*, 2007). The sole member of this class is XplA from *Rhodococcus rhodochrous* strain 11Y, which is involved in the biodegradation of RDX (Royal Demolition Explosive – hexa-hydro-1,3,5-trinitro-1,3,5-triazine) (Jackson *et al.*, 2007). It consists of a Fldx redox partner fused at the N-terminus of a cytochrome P450 domain, which accepts electrons from NADPH (Jackson *et al.*, 2007). Electrons are transferred from NADPH to the fusion protein by XplB, a potential adrenodoxin reductase found upstream of XplA (Jackson *et al.*, 2007).

Figure 1.13 – Class VI cytochromes P450 (adapted from Hannemann *et al.*, 2007). N-terminally linked Fdx-cytochrome P450 fusion protein.

1.2.2.7 Class VII

Cytochromes P450 belonging to Class VII are novel fusion proteins that have a phthalate dioxygenase reductase domain linked to the C-terminus of the cytochrome P450 (figure 1.14) (Hannemann *et al.*, 2007). The first member of this class was CYP116B2 (P450RhF) from *Rhodococcus* sp. (Roberts *et al.*, 2002). This enzyme has been show to produce 7-hydroxycoumarin *in vivo* without the need for additional redox partners (Hunter *et al.*, 2005). Interestingly, unlike the other members of this class CYP116B2 has a preference for NADPH as an electron donor over NADH (Roberts *et al.*, 2003).

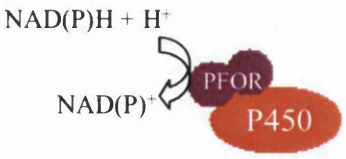


Figure 1.14 – Class VII cytochromes P450 (adapted from Hannemann *et al.*, 2007). Phthlate-family oxygenase reductase (PFOR) fused to a cytochrome P450 complex.

1.2.2.8 Class VIII

Class VIII cytochromes P450, like class II proteins, utilise CPR as a redox partner to transfer electrons from NAD(P)H (figure 1.15). However, the CPR and cytochrome P450 complex exist as a single fusion protein in this system. Two examples are the fusion proteins CYP102A1 ($P450_{BM-3}$) from *Bacillus megaterium* (Miura & Fulco, 1974) and CYP505A1 (P450foxy) from *Fusarium oxysporum* (Nakayama *et al.*, 1996). Both are self-sufficient monooxygenases that are involved in the sub-terminal hydroxylation of fatty acids (Miura & Fulco, 1974; Nakayama *et al.*, 1996; Warman *et al.*, 2005).

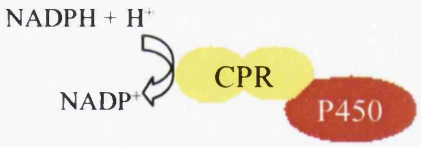


Figure 1.15 – Class VIII cytochromes P450 (adapted from Hannemann *et al.*, 2007). CPR-cytochrome P450 fusion protein.

1.2.2.9 Class IX

The nitic oxide reductase, CYP55 (P450nor) is the sole family belonging to class IX (figure 1.16). It is found in *F. oxysporum* (Nakahara & Shoun, 1996), *Cylindrocarpon tonkinense* (Usuda *et al.*, 1995; Kudo *et al.*, 1996), *Trichosporon cutaneum* (Zhang *et al.*, 2001) and *Aspergillus oryzae* (Kaya *et al.*, 2004). In anaerobic conditions, CYP55 can accept electrons from NADH to catalyse the conversion of nitrogen oxide to nitrous oxide (Kaya *et al.*, 2004; Daiber *et al.*, 2005).

This transfer of electrons is achieved without the use of redox partners (Kaya *et al.*, 2004).

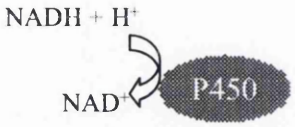


Figure 1.16 – Class IX cytochromes P450 (adapted from Hannemann *et al.*, 2007). Nitric-oxide reductase cytochrome P450 system.

1.2.2.10 Class X

Class X cytochromes P450 are membrane-bound proteins found in the endoplasmic reticulum of plants and mammals (figure 1.17) (Hannemann *et al.*, 2007). These proteins are able to synthesise organic substrates independently of electron transfer (via redox partners) (Hannemann *et al.*, 2007). Examples of cytochromes P450 belonging to this class are: CYP74A (allene oxide synthase) (Song *et al.*, 1993), CYP74B (hydroperoxide lyase) (Matsui *et al.*, 1996), CYP74C (allene oxide synthase) (Stumpe *et al.*, 2006) and CYP74D (divenyl ether synthase) (Stumpe *et al.*, 2001), which are all found in chloroplasts and involved in cell defence through the lipoxygenase pathway (Froehlich *et al.*, 2001). As these enzymes neither require oxygen nor a redox partner, they use the acyl hydroperoxide from the substrate to drive the reaction (Lau *et al.*, 1993; Shibata *et al.*, 1995; Itoh & Howe, 2001).

Figure 1.17 – Class X cytochromes P450 (adapted from Hannemann *et al.*, 2007). Independent cytochrome P450 system.

1.3 Alkane and Fatty Acid Hydroxylating Cytochromes P450, CYP52s, from Fungi and Yeast

CYP52s are a family of alkane-inducible cytochromes P450 found only in yeast and fungi, such as *Candida* spp., that can utilise alkanes for growth. Within this family multiple genes have been identified. For example, approximately 23 cytochromes P450 have been found in *C. tropicalis* of which 18 belong to the CYP52 family (this study). However, in *Saccharomyces cerevisiae*, a yeast strain which cannot use alkanes for growth, only 3 cytochromes P450 have been identified (Kelly *et al.*, 2005). These genes are *CYP51* (also called *ERG11* and 14α-demethylase), *CYP61* (also called *ERG5* and 22-desaturase) and *CYP56* (also called *DIT2*) (Kelly *et al.*, 2005).

Cytochrome P450 enzymes were first implicated in alkane and fatty acid hydroxylation in 1971 (Lebeault *et al.*, 1971). *C. tropicalis* was grown on tetradecane (C20:0) and the harvested cells, containing cytochrome P450, were used to hydroxylate various fatty acids and alkanes (Lebeault *et al.*, 1971). The cytochromes P450 in *C. tropicalis* were shown to preferentially hydroxylate mid- to long-chain fatty acids, especially lauric acid (C12:0) and alkanes (hexadecane (C16:0), decane (C10:0) and octane (C8:0)) (Lebeault *et al.*, 1971). However, the first alkane-inducible cytochrome P450, P450alk, from *C. tropicalis* was isolated and cloned in 1987 (Sanglard *et al.*, 1987). This was later characterised and identified by Sanglard and Loper as the first member of the CYP52 family, CYP52A1, in 1989 (Sanglard & Loper, 1989). To date (February 2011), 52 CYP52s have been identified in *C. tropicalis*, *C. maltosa*, *C. albicans*, *C. apicola*, *C. bombicola*, *Yarrowia lipolytica*, *Aspergillus nidulans*, *A. fumigatus*, *A. oryzae* and *Debaryomyces hansenii* (http://drnelson.uthsc.edu/CytochromeP450.html).

1.3.1 Metabolism of Alkanes and Fatty Acids by CYP52s

An overview of the pathway involved in the metabolism of alkanes and fatty acids by CYP52s is shown in figure 1.18

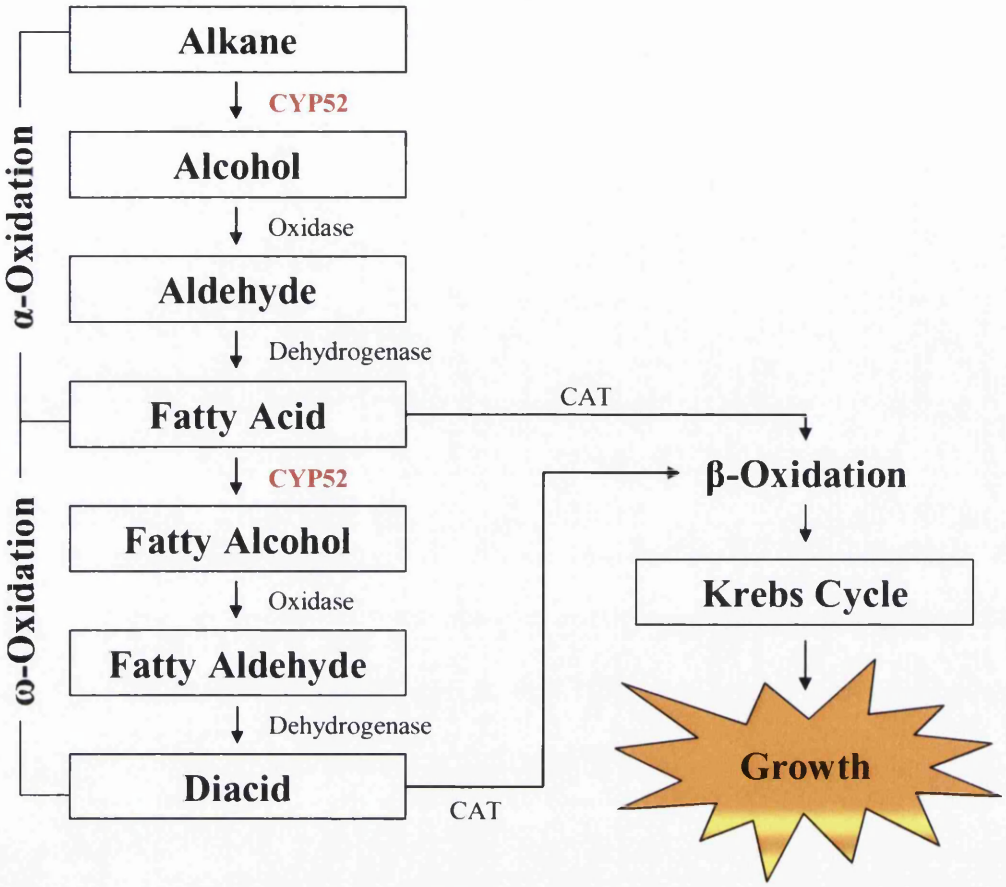


Figure 1.18 – Overview of the metabolic pathway used to convert alkanes and/or fatty acids to diacids. Fatty acids and diacids, produced through α- and ω-oxidation, are transferred to the β-oxidation pathway. This in turns produces acetyl CoA, which enters the Krebs cycle (Porter & Coon, 1991; Jiao *et al.*, 2000; Eschenfeldt *et al.*, 2003). CAT = Carnitine Acetyltransferase

In vivo the production of diacids from long-chain alkanes and fatty acids occurs in the endoplasmic reticulum. Initially, the long-chain alkane undergoes α-oxidation and is hydroxylated to its respective primary alcohol by the NADPH-dependent CYP52s. This is the rate-limiting step of this pathway (Scheller *et al.*, 1998). The primary alcohol can subsequently be converted to its respective aldehyde either by the same CYP52 or by alcohol dehydrogenase. Aldehyde dehydrogenase or CYP52 can then be used to catalyse the insertion of an oxygen atom into the aldehyde producing the respective long-chain fatty acid. This fatty acid can either be transported to the peroxisome for use in β-oxidation and eventual cell growth or can be further hydroxylated to the diacid (Porter & Coon, 1991; Jiao *et al.*, 2000; Eschenfeldt *et al.*, 2003).

The fatty acid is hydroxylated at the ω-carbon (the carbon on the alkyl chain furthest away from the carboxyl side group (see figure 1.19)) during ω-oxidation to produce the fatty alcohol. The fatty alcohol can then be further converted to the fatty aldehyde and the diacid either by CYP52 or by alcohol dehydrogenase and aldehyde dehydrogenase respectively. The diacid can then be transported from the endoplasmic reticulum to the peroxisome to be subjected to the β-oxidation pathway for use in growth (Porter & Coon, 1991; Jiao *et al.*, 2000; Eschenfeldt *et al.*, 2003).

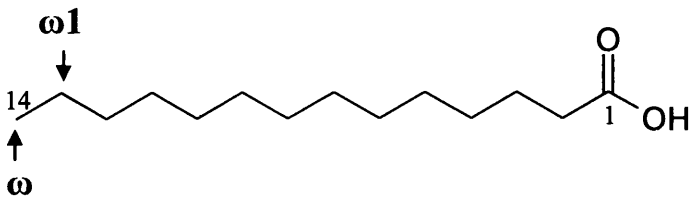


Figure 1.19 – Position of the terminal (ω-) carbon of a fatty acid chain highlighted on myristic acid (C14:0) as an example. Also indicated on this molecule is the sub-terminal (ω1-) carbon, which is found adjacent to the ω-carbon.

Figure 1.20 shows an overview of the structural changes (described above) that occur at the ω-carbon during the production of fatty acids and diacids.

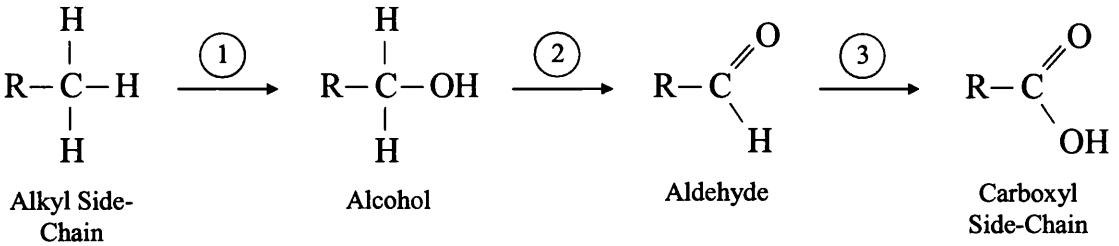


Figure 1.20 – Overview of the structural changes that occur during the metabolism of alkanes and fatty acids. 1) The rate-limiting step of the pathway. CYP52 catalyses the insertion of oxygen into the alkyl side-chain to produce an alcohol. 2) CYP52 or alcohol dehydrogenase catalyse the formation on an aldehyde from the alcohol. 3) CYP52 or aldehyde dehydrogenase catalyse the insertion of oxygen into the aldehyde to produce a carboxyl side-chain. This results in the production of a fatty acid or diacid molecule. R = hydrocarbon side-chain, C = carbon, H = hydrogen and O = oxygen.

As indicated above, CYP52 enzymes could generate diacids alone, which is a benefit from a biotechnological standpoint. However, *in vivo*, this would result in the depletion of NADPH as it would require three times as many electrons to produce aldehydes from alcohols as it would to generate only the alcohol from the alkane and/or fatty acid (Scheller *et al.*, 1998). Therefore, *in vivo*, cytochromes P450 might be used to enhance the overall productivity of the other enzymes (i.e. alcohol dehydrogenase and aldehyde dehydrogenase) in the α- and ω-pathways, but are not required to carry out the further conversions of alcohols and aldehydes by themselves (Scheller *et al.*, 1998; Eschenfeldt *et al.*, 2003).

1.3.2 Functions of CYP52s

As mentioned (section 1.3), approximately 52 CYP52s have been identified in a number of fungal species to date. Many have been shown to be induced preferentially by various alkanes and fatty acids using experimental procedures, such as QC RT-PCR (which is used to determine the transcriptional induction of similar sequences by appropriate substrates) (Craft *et al.*, 2003). For example, CYP52A18 from *C. tropicalis* was shown to be preferentially induced by octadecane (C18:0) and CYP52A17 (from the same species) by oleic acid (C18:1) (Craft *et al.*, 2003). However, very few of the CYP52s discovered thus far have been studied at the protein level.

1.3.2.1 Diacid-Producing CYP52s

Of the CYP52s identified to date only CYP52A3, CYP52A4, CYP52A13 and CYP52A17 have been shown to produce diacids from a mixture of mid- to long-chain alkanes and fatty acids.

CYP52A3 from the soil-dwelling fungi *C. maltosa* is the archetypal alkane-hydroxylating CYP52. It has been shown to produce 1,16-hexadecanedioic acid (C16:0 diacid) (figure 1.21) from hexadecane (C16:0) (Scheller *et al.*, 1998). When incubated with myristic acid (C14:0), palmitic acid (C16:0) and oleic acid (C18:1), CYP52A3 has also been shown to produce the respective ω-hydroxy fatty acid without converting the molecule to the diacid (Zimmer *et al.*, 1996).

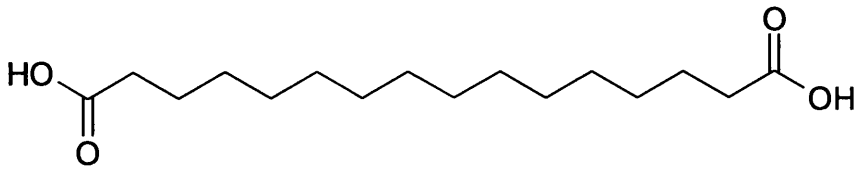


Figure 1.21 – 1,16-Hexadecanedioic acid (C16:0 diacid).

CYP52A4, also from *C. maltosa*, was shown to produce a diacid upon incubation with a hydrocarbon substrate. Unlike CYP52A3, CYP52A4 preferentially hydroxylates fatty acids. It is involved in the conversion of lauric acid (C12:0) to 1,12-dodecanedioc acid (C12:0 diacid) (figure 1.22) (Zimmer *et al.*, 1995). It is also able to hydroxylate dodecane (C12:0), hexadecane (C16:0) and myristic acid (C14:0) at the ω-carbon to the corresponding alcohol (Zimmer *et al.*, 1996).

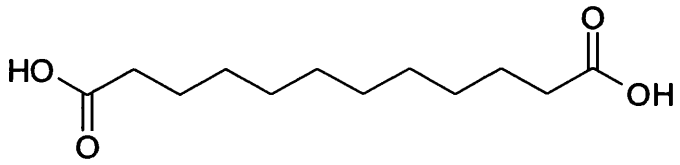


Figure 1.22 – 1,12-Dodecanedioic acid (C12:0 diacid).

CYP52A13 from *C. tropicalis* is able to convert multiple unsaturated fatty acids (oleic acid (C18:1), linoleic acid (C18:2) and arachadonic acid (C20:4)) to their respective diacids (Eschenfeldt *et al.*, 2003). It is also able to produce fatty alcohols by ω-hydroxylating capric acid (C10:0), lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0) and stearic acid (C18:0) (Eschenfeldt *et al.*, 2003).

Another CYP52 from *C. tropicalis*, CYP52A17 is also able to produce diacids. Like CYP52A13, it can convert multiple unsaturated fatty acids (oleic acid (C18:1), linoleic acid (C18:2) and arachadonic acid (C20:4)) to the corresponding diacid, as well as several saturated fatty acids (lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0) and stearic acid (C18:0)) (Eschenfeldt *et al.*, 2003). CYP52A17 is also able to hydroxylate capric acid (C10:0) to its relative alcohol (Eschenfeldt *et al.*, 2003).

1.3.2.2 Other Products of CYP52s

CYP52A21 (Alk8) from *C. albicans* has been shown to hydroxylate lauric acid (C12:0) to the ω-hydroxy fatty acid, as well as, to a lesser extent, the sub-terminal or ω1-hydroxylated version (Kim *et al.*, 2007). The ω1-hydroxy fatty acid produced is thought to be used for growth like the ω-hydroxy form, but the actual mechanism by which this occurs is unclear. However, the following route can be proposed (see figure 1.23). The fatty alcohol undergoes sub-terminal oxidation whereby an oxygen atom is inserted between the ω1- and ω2-carbons. This oxidation results in cleavage and the production of two molecules: an ω-hydroxy fatty acid and acetic acid. The hydroxy fatty acid can then be further converted to the diacid and used in β-oxidation. The acetic acid enters into the Krebs cycle.

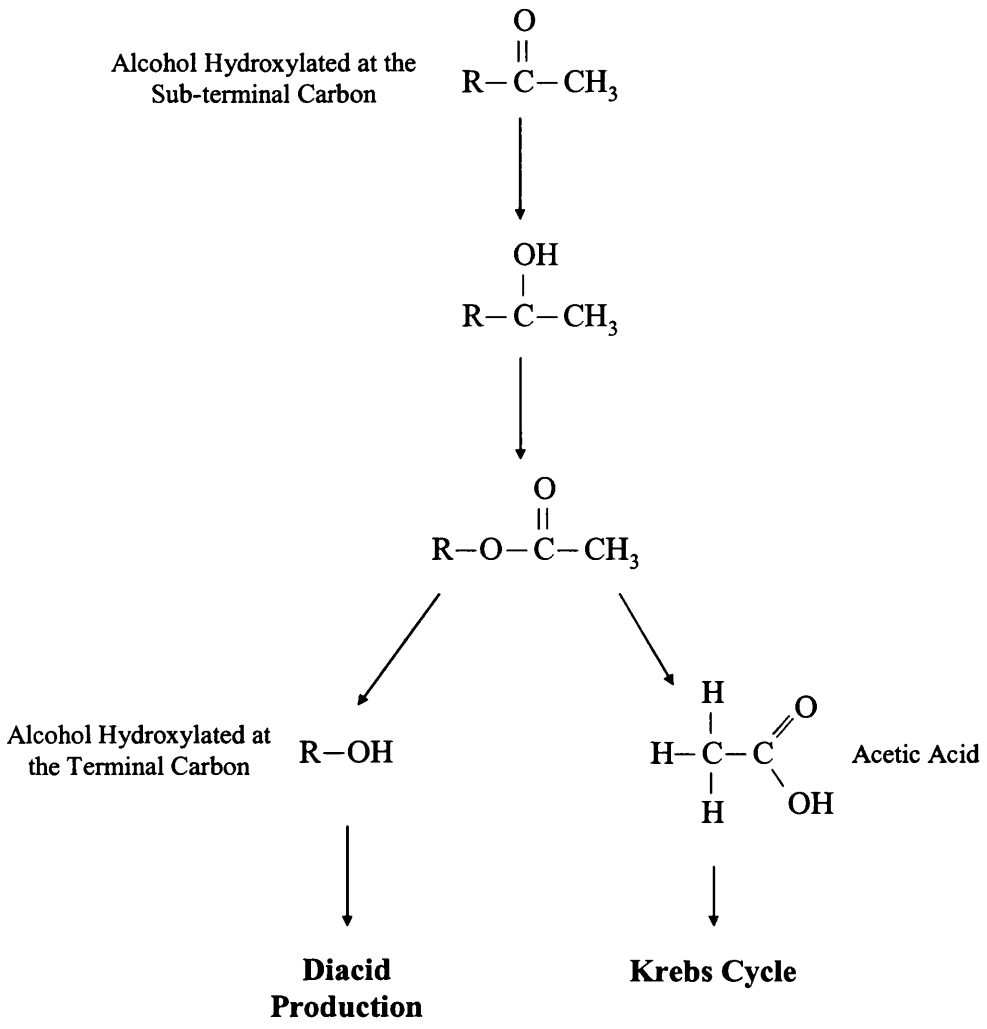


Figure 1.23 – Proposed sub-terminal oxidation of ω1-hydroxy fatty alcohol. Based on the subterminal oxidation of secondary alcohols (Ratledge, 1984; van Bogaert *et al.*, 2011). Oxygen is inserted into the molecule between the ω1- and ω2-carbons. The molecule is then cleaved between the oxygen atom and the ω1-carbon producing an ω-fatty alcohol and acetic acid. R = hydrocarbon side-chain, C = carbon, O = oxygen and H = hydrogen.

In *C. apicola* and *C. bombicola*, CYP52s are involved in the production of sophorolipids (Lottermoser *et al.*, 1996; van Bogaert *et al.*, 2008). Sophorolipids are extracellular glycolipids, which can be used commercially as biosurfactants (Konishi *et al.*, 2008). They are made up of a disaccharide sophorose sugar, which is linked glycosidically to a (saturated or unsaturated) fatty acid of usually 16 to 18 carbons in length (figure 1.24) (Casas & García-Ochoa, 1999; Nuñez *et al.*, 2001; Shah *et al.*, 2005). The glycosidic link occurs at the hydroxyl group found on the ω1-carbon of the fatty acid (Shah *et al.*, 2005), which is thought to arise due to hydroxylation by

CYP52E1 and CYP52E2 from *C. apicola* (Lottermoser *et al.*, 1996) and CYP52E3 from *C. bombicola* (van Bogaert *et al.*, 2008).

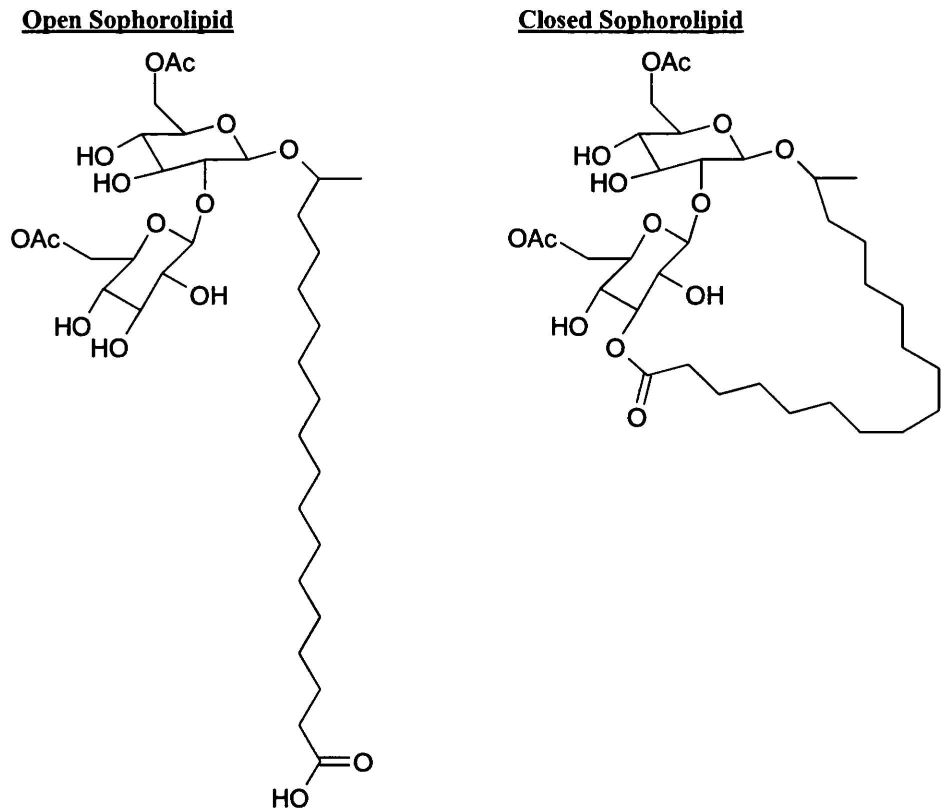


Figure 1.24 – Sophorolipids. Sophorolipids can exist in two conformations: open and closed. The carboxyl group of the fatty acid can bind to the –OH group of the sophorose sugar to produce a closed ring or it can remain unbound to produce an open compound.

1.4 Novel Cytochrome P450 Enzymes Involved in the Hydroxylation of Fatty Acids in Yeast and Fungi

Members of the CYP52 family are not the sole cytochrome P450 enzymes in yeast and fungi to be involved in the catalysis of fatty acids. Two novel cytochrome P450 enzymes, P450foxy and PpoA, have also been identified.

1.4.1 P450foxy (CYP505A1)

P450foxy (CYP505A1) is a cytochrome P450-CPR fusion protein from *F. oxysporum* (Nakayama *et al.*, 1996). It is involved in the sub-terminal and in-chain hydroxylation (ω1- to ω3-carbons) of several fatty acids, including lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0) and stearic acid (C18:0) (Nakayama *et al.*, 1996). However, unlike members of the CYP52 family P450foxy is not involved in the hydroxylation of alkanes (Nakayama *et al.*, 1996).

1.4.2 PpoA

PpoA is a novel fusion protein from *Aspergillus nidulans* that exists as a haem peroxidase (a dioxygenase enzyme) fused to a cytochrome P450 domain and is involved in the production of precocious sexual inducer (psi) factors from polyunsaturated fatty acids, such as linoleic acid (C18:2) (Brodhun *et al.*, 2009). The haem peroxidase converts linoleic acid (C18:2) to (8R)-hydroperoxyoctadecadienoic acid ((8R)-HPODE) by incorporating two oxygen atoms into the structure at carbon 8 (see figure 1.25) (Brodhun *et al.*, 2009). (8R)-HPODE is then converted, by the cytochrome P450 enzyme, to produce 5,8-dihydroxyoctadecadienoic acid (5,8-DiHODE), a psi factor involved in sexual cleistothecia and asexual conidiospore development (see figure 1.25) (Brodhun *et al.*, 2009).

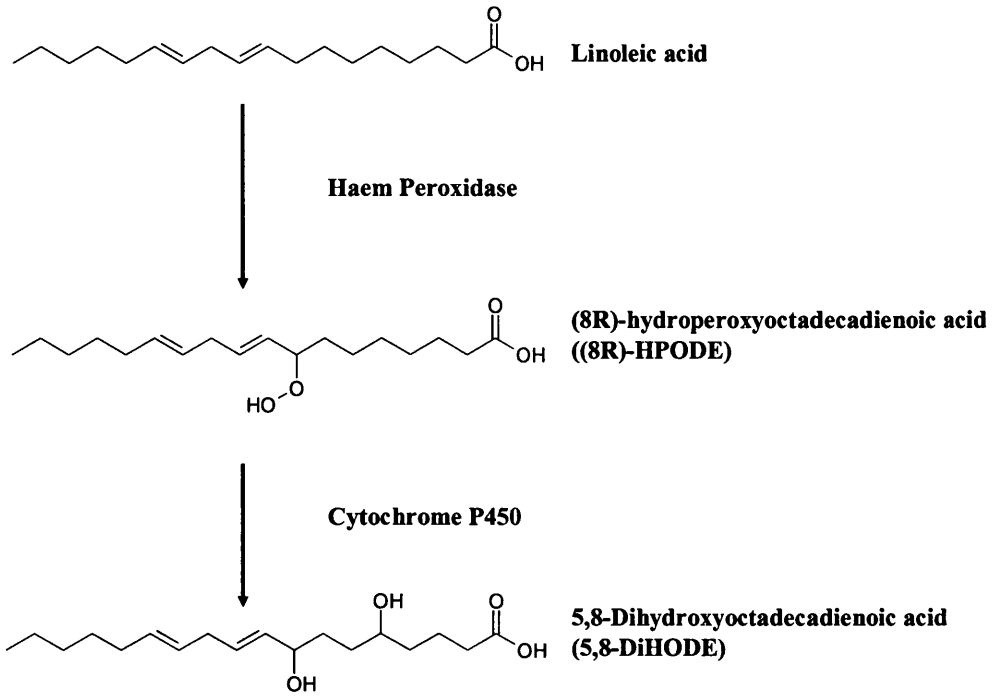


Figure 1.25 – Conversion of linoleic acid (C18:2) to the PSI factor 5,8-dihydroxyoctandecadienoic acid by the novel cytochrome P450 fusion protein, PpoA.

1.5 Alkane and/or Fatty Acid Hydroxylating Cytochromes P450 from Other Organisms

Alkane and/or fatty acid hydroxylating cytochromes P450 are not solely found in yeast and fungi, they can also be found in many other kingdoms of life, including bacteria, plants and animals. Table 1.1 shows an overview of these cytochrome P450 families, which are discussed in further detail below.

Bacteria	Plants	Animals
CYP102	CYP76	CYP1
CYP105	CYP77	CYP2
CYP107	CYP78	CYP4
CYP119	CYP81	
CYP124	CYP86	
CYP152	CYP92	
CYP153	CYP94	
	CYP703	
	CYP704	
	CYP709	
	CYP726	

Table 1.1 – Cytochrome P450 families involved in the hydroxylation of alkanes and/or fatty acids from bacteria, plants and animals.

1.5.1 Alkane and Fatty Acid Hydroxylating Cytochromes P450 from Bacteria

To date 13 proteins from 7 cytochrome P450 families in bacteria have been shown to catalyse the hydroxylation of alkanes and fatty acids *in vitro* (see table 1.2). As with CYP52, the physiological roles of many of these enzymes are undocumented, however, their activities are discussed below.

Cytochrome P450 Name	Species	Activity	Reference
$P450_{BM-3}$ (CYP102A1)	*Bacillus megaterium*	In-chain fatty acid hydroxylation, fatty acid epoxidation	Muira & Fulco, 1975; Gustafsson *et al.*, 2004; Lamb *et al.*, 2010
CYP102A2	*Bacillus subtilis*	In-chain fatty acid hydroxylation	Gustafsson *et al.*, 2004
CYP102A3	*Bacillus subtilis*	In-chain fatty acid hydroxylation	Gustafsson *et al.*, 2004
CYP102A7	*Bacillus licheniformis*	In-chain fatty acid hydroxylation	Dietrich *et al.*, 2008
CYP102B1	*Streptomyces coelicolor* A3(2)	Fatty acid hydroxylation, fatty acid epoxidation	Lamb *et al.*, 2010
CYP105D5	*Streptomyces coelicolor* A3(2)	In-chain fatty acid hydroxylation	Chun *et al.*, 2007
$P450_{BioI}$ (CYP107H1)	*Bacillus subtilis*	Terminal (ω) and in-chain fatty acid hydroxylation, C-C bond cleavage	Cryle *et al.*, 2003
CYP119	*Sulfolobus acidocaldarius*	Terminal (ω) fatty acid hydroxylation	Lim *et al.*, 2010
CYP124	*Mycobacterium tuberculosis*	Terminal (ω) hydroxylation of branched fatty acids	Johnston *et al.*, 2009
$P450_{BS\beta}$ (CYP152A1)	*Bacillus subtilis*	α- and β-hydroxylation of fatty acids	Lee *et al.*, 2003
CYP152A2	*Bacillus subtilis*	α- and β-hydroxylation of fatty acids	Girhard *et al.*, 2007
$P450_{SP\alpha}$ (CYP152B1)	*Sphingomonas paucimobilis*	α-hydroxylation of fatty acids	Lee et al., 2003
CYP153A6	*Mycobacterium* sp. HXN-1500	Terminal (ω) hydroxylation of alkanes	Funhoff *et al.*, 2006

Table 1.2 – Alkane and fatty acid hydroxylating cytochromes P450 from bacteria. Overview of the cytochromes P450 from bacteria, which have been shown to be involved in the hydroxylation of alkanes and fatty acids *in vitro*.

1.5.1.1 CYP102A1, CYP102A2, CYP102A3, CYP102A7 and CYP102B1

$P450_{BM-3}$ (also known as CYP102A1) from *Bacillus megaterium* is a catalytically self-sufficient cytochrome P450-CPR fusion protein (Warman *et al.*, 2005). It can catalyse the hydroxylation of long-chain, saturated fatty acids (lauric acid (C12:0) to stearic acid (C18:0)) at the ω1-, ω2- and ω3-carbons (but not at the terminal ω-carbon) (Miura & Fulco, 1975; Boddupalli *et al.*, 1992; Schneider *et al.*, 1998; Gustafsson *et al.*, 2004; Lamb *et al.*, 2010) as well as the hydroxylation and epoxidation of arachidonic acid (C20:4) (Schneider *et al.*, 1998; Lamb *et al.*, 2010). The rate of hydroxylation of fatty acids by $P450_{BM-3}$ is the highest amongst all cytochrome P450 enzymes (for example, the catalytic rate is greater than $15000min^{-1}$ when arachidonic acid is used as the substrate) (Noble *et al.*, 1999; Girvan *et al.*, 2006) and the structure remains the most studied within the superfamily.

Other members of the CYP102A subfamily have also been expressed and characterised. These include: CYP102A2 and CYP102A3 from *B. subtilis* (Gustafsson *et al.*, 2004) and CYP102A7 from *B. licheniformis* (Dietrich *et al.*, 2008), which, like $P450_{BM-3}$, are all cytochrome P450-reductase fusion proteins. CYP102A2 and CYP102A3 are both involved in the hydroxylation of myristic acid (C14:0) with CYP102A2 preferentially hydroxylating the ω2- and ω3-carbons and CYP102A3 the ω1-carbon (Gustafsson *et al.*, 2004). The CYP102A7 protein can catalyse the hydroxylation of ω1-, ω2- (preferred) and ω3-carbons of a variety of fatty acids, including straight-chained saturated (lauric acid (C12:0) to palmitic acid (C16:0)) and unsaturated (palmitoleic acid (C16:1) and linoleic acid (C18:2)) molecules as well as branched versions of myrisitic acid (C14:0) (12-methyl-myristic acid and 13-methyl-myristic acid) (Dietrich *et al.*, 2008).

Interestingly, members of the CYP102B subfamily can also hydroxylate fatty acids, but unlike the members of CYP102A they are not fusion proteins and, therefore, require the addition of redox partners for activity. CYP102B1 from *Streptomyces coelicolor* A3(2) is involved in catalysing the hydroxylation and epoxidation of arachidonic acid (C20:4) in the presence of spinach ferredoxin and spinach ferredoxin reductase (Lamb *et al.*, 2010). As with $P450_{BM-3}$, CYP102A2 and CYP102A3, the physiological role of this enzyme is unknown as the production of a

transposon mutant of CYP102B1 did not result in any alterations in cell growth and the detectable metabolites produced (Lamb *et al.*, 2010).

1.5.1.2 CYP105D5

CYP105D5 from *S. coelicolor* A3(2) is involved in the hydroxylation of lauric acid (C12:0) at the ω1-, ω2-, ω3- and ω4-positions and oleic acid (C18:1) at the ω1-carbon, however, no ω-hydroxylation has been identified to date (Chun *et al.*, 2007). Despite *in vivo* and *in vitro* experiments showing the role of CYP105D5 as a fatty acid hydroxylase, the physiological role of this enzyme is unknown much like CYP102B from the same species (Chun *et al.*, 2007).

1.5.1.3 CYP107H1

$P450_{BioI}$ (CYP107H1) from *B. subtilis* is involved in fatty acid hydroxylation and C-C bond cleavage (Cryle *et al.*, 2003). Fatty acids, including myrisitic acid (C14:0) and palmitic acid (C16:0) are hydroxylated at either the ω-, ω1-, ω2-, ω3-, ω4- or ω5-carbons and then subjected to chain-shortening through C-C bond cleavage, which is also catalysed by $P450_{BioI}$ (Cryle *et al.*, 2003). This results in the formation of pimelic acid a diacid 7 carbons in length (figure 1.26) (Cryle *et al.*, 2003). Pimelic acid can then be used in the synthesis of biotin, a water-soluble vitamin B complex (Cryle *et al.*, 2003). The specific cleavage of the C-C bond at carbons 7 and 8 is thought to occur with the aid of an acyl carrier protein (ACP) linked to a fatty acid as experiments using ACP from *E. coli* and $P450_{BioI}$ have shown to produce pimelic acid (Cryle & Schlichting, 2008). The fatty acyl-ACP complex binds tightly to the $P450_{BioI}$ causing the alkyl chain of the fatty acid to form a U-shape, presenting the 7^{th} and 8^{th} carbons to the haem (Cryle & Schlichting, 2008). This, in turn, results in the cleavage of the hydrocarbon chain and the eventual formation of pimelic acid (Cryle & Schlichting, 2008).

Pimelic Acid

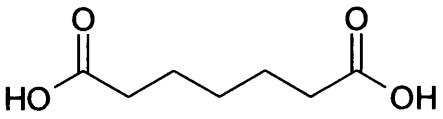


Figure 1.26 – Pimelic acid, a 7-carbon chain diacid, which is the product of fatty acid hydroxylation and C-C bond cleavage by $P450_{BioI}$.

1.5.1.4 CYP119

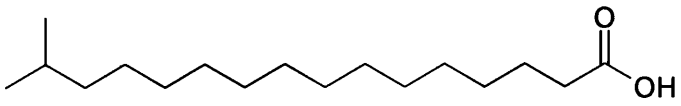
CYP119 is a novel thermophilic cytochrome P450, which is stable at 85°C and found in the hot-spring archaebacteria, *Sulfolobus acidocaldarius* (Koo *et al.*, 2000; Lim *et al.*, 2010). Although, the physiological function of CYP119 is unknown, it has been shown to be capable of the ω-hydroxylation of lauric acid (C12:0) (Lim *et al.*, 2010).

1.5.1.5 CYP124

CYP124 from *Mycobacterium tuberculosis* is active in the terminal hydroxylation of methyl-branched fatty acids, including 15-methyl palmitic acid and phytanic acid (3,7,11,15-methyl palmitic acid) (figure 1.27), to produce the respective ω-hydroxy fatty acids (Johnston *et al.*, 2009). Due to the methyl groups being situated on the sub-terminal carbon in both branched fatty acids mentioned, no hydroxylation occurs at the ω1-position. However, when palmitic acid (C16:0) is used as a substrate, CYP124 is involved in its hydroxylation at the ω-, ω1- and ω2-carbons (Johnston *et al.*, 2009). The *in vivo* uses of ω-hydroxy methyl-branched fatty acids in *M. tuberculosis* are undocumented, but they are thought to be utilised by the β-oxidation pathway (Johnston *et al.*, 2009).

Other methyl-branched hydrocarbons, known as isoprenoids, can also be hydroxylated at the ω-position to form the respective hydroxy version (Johnston *et al.*, 2009). These substrates include: farnesol, geranylgeraniol and farnesyl diphosphate (figure 1.28), all of which can be used by the isoprenoid biosynthetic pathway in *M. tuberculosis* (Johnston *et al.*, 2009). This pathway is essential as it is involved in the production of the cell wall (Brown *et al.*, 2010).

15-Methyl Palmitic Acid



Phytanic Acid

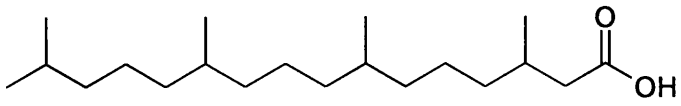
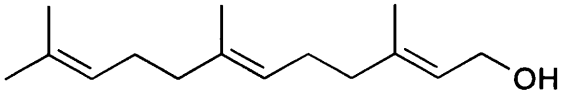
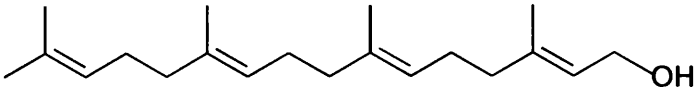


Figure 1.27 – The structures of 15-methyl palmitic acid and phytanic acid (3,7,11,15-methyl palmitic acid).

Farnesol



Geranylgeraniol



Farnesyl Diphosphate

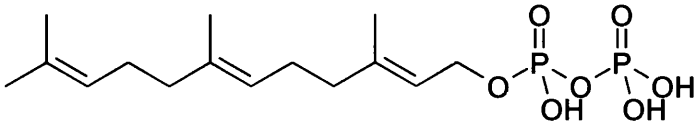


Figure 1.28 – Isoprenoids: farnesol, geranylgeraniol and farnesyl diphosphate.

1.5.1.6 CYP152A1, CYP152A2 and CYP152B1

The CYP152 family of cytochromes P450 are involved in the α- and/or β-hydroxylation of long-chain fatty acids. Hydroxylation occurs at the second (α) and third (β) carbons rather than at the terminal (ω) and sub-terminal (ω_n) carbons (figure 1.29). Members of the CYP152A subfamily are able to catalyse the hydroxylation of both α- and β-carbons to produce 2-hydroxy and 3-hydroxy fatty acids respectively (figure 1.30) (Lee *et al.*, 2003; Girhard *et al.*, 2007). However, $P450_{BS\beta}$ (CYP152A1) from *B. subtilis* is considered a β-hydroxylase as it preferentially hydroxylates at the β-carbon (Lee *et al.*, 2003), whereas CYP152A2 is thought of as an α-hydroxylase (Girhard *et al.*, 2007). $P450_{SP\alpha}$ (CYP152B1) from *Sphingomonas paucimobilis* is also an α-hydroxylase, however, it does not catalyse β-hydroxylation (Lee *et al.*, 2003). The 2-hydroxy and 3-hydroxy fatty acids produced by these enzymes are important in the production of membrane lipids, including sphingolipids. *S. paucimobilis* is a rich source of 2-hydroxy myristic acid-containing sphingolipids, which is the product of myrisitic acid (C14:0) hydroxylation by $P450_{SP\alpha}$ (Matsunaga *et al.*, 1997; Imai *et al.*, 2000).

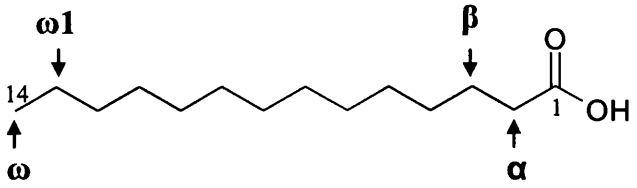
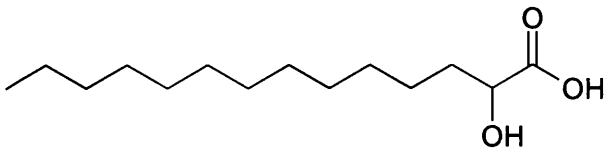


Figure 1.29 – Position of the α- and β- carbons of a fatty acid chain highlighted on myristic acid (C14:0) as an example. The carbon that makes up the carboxyl group is numbered 1. The adjacent carbon (2) is the α-carbon and the third carbon in the chain is the β-carbon. These carbons are hydroxylated by members of the CYP152 family of cytochromes P450.

2-Hydroxy Fatty Acid



3-Hydroxy Fatty Acid

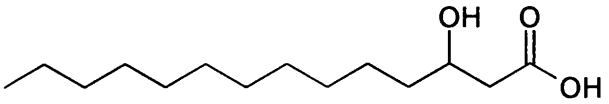


Figure 1.30 – Structural examples of 2-hydroxy and 3-hydroxy fatty acids, the respective products of α- and β-hydroxylation.

CYP152 enzymes are peroxygenases rather than monooxygenases (Imai *et al.*, 2000; Lee *et al.*, 2003; Girhard *et al.*, 2007). Therefore, instead of an oxygen atom from atmospheric oxygen being incorporated into the fatty acid, the oxygen atom comes from hydrogen peroxide (Imai *et al.*, 2000; Lee *et al.*, 2003; Girhard *et al.*, 2007). The use of hydrogen peroxide in this manner fits with the hypothesis that one of the functions of CYP152 *in vivo* is as a rapid hydrogen peroxide scavenging enzyme (Girhard *et al.*, 2007).

1.5.1.7 CYP153A6

CYP153 from Eubacteria are capable of metabolising alkanes of a variety of chain lengths (C6:0 to C16:0) (Funhoff *et al.*, 2007). For instance, CYP153A6 from *Mycobacterium* sp. HXN-1500 has been shown to hydroxylate a range of short-chain alkanes from hexane (C6:0) to undecane (C11:0) at the terminal carbon producing the ω-hydroxy alcohol (Funhoff *et al.*, 2006). This hydrocarbon, in turn, can be further metabolised as alcohol dehydrogenase and aldehyde dehydrogenase convert the alcohol to its respective aldehyde and fatty acid (Wentzel *et al.*, 2007). The fatty acid can then be subjected to β-oxidation and used for growth (Wentzel *et al.*, 2007). Interestingly, unlike the CYP52 family, no CYP153 enzymes to date have been shown to metabolise the alcohol further.

1.5.2 Fatty Acid Hydroxylating Cytochromes P450 from Plants

Unlike the fatty acid hydroxylating cytochromes P450 from yeast/fungi and bacteria, the physiological roles of many of these enzymes in plants have been identified. They have been shown to be implicated in evoking defence mechanisms within the plant, reproduction function and in the synthesis of cutin and suberin (Pinot & Beisson, 2011). Cutin and suberin are lipid polymers, which are found in the hydrophobic regions of the cell walls in plants (Pollard *et al.*, 2008). Cutin is primarily found in the water-repellent protective layer of the epidermis known as the cuticle, whereas suberin is found in the roots and plays a role in ion uptake and water control (Pollard *et al.*, 2008). They mostly consist of hydroxy fatty acids, functionalised fatty acids (e.g. epoxides) and diacids linked by ester bonds (Pollard *et al.*, 2008).

To date 19 fatty acid hydroxylating cytochromes P450 from 11 families have been expressed at the protein level and characterised (see table 1.3). These enzymes are discussed below.

Cytochrome P450 Name	Species	Activity	Physiological Role	Reference
CYP76B9	*Petunia hybrida*	Terminal (ω) fatty acid hydroxylation	Possibly evokes defence mechanisms (?)	Imaishi & Petkova-Andonova, 2007
CYP77A4	*Arabidopsis thaliana*	In-chain fatty acid hydroxylation, fatty acid epoxidation	Evokes defence mechanisms, cutin synthesis	Li-Beisson *et al.*, 2009
CYP77A6	*Arabidopsis thaliana*	In-chain fatty acid hydroxylation	Cutin synthesis	Sauveplane *et al.*, 2009
CYP78A1	*Zea mays*	Terminal (ω) fatty acid hydroxylation	Flower development and reproduction	Imaishi *et al.*, 2000; Imaishi & Ohkawa, 2002
CYP81B1	*Helianthus tuberosus*	In-chain fatty acid hydroxylation	Possibly evokes defence mechanisms and/or cutin synthesis (?)	Cabello-Hurtado *et al.*, 1998
CYP86A1	*Arabidopsis thaliana*	Terminal (ω) fatty acid hydroxylation	Possibly cutin synthesis (?), suberin synthesis	Benveniste *et al.*, 1998; Höfer *et al.*, 2008
CYP86A4	*Arabidopsis thaliana*	Terminal (ω) fatty acid hydroxylation	Possibly cutin synthesis (?)	Li-Beisson et al., 2009
CYP86A8	*Arabidopsis thaliana*	Terminal (ω) fatty acid hydroxylation	Possibly cutin synthesis (?)	Wellesen *et al.*, 2001
CYP86A22	*Petunia hybrida*	Terminal (ω) fatty acid hydroxylation	Estolide synthesis in the stigma	Han *et al.*, 2010
CYP92B1	*Petunia hybrida*	Terminal (ω) fatty acid hydroxylation	Possibly flower development (?)	Petkova-Andonova *et al.*, 2002
CYP94A1	*Vicia sativa*	Terminal (ω) fatty acid hydroxylation	Possibly evokes defence mechanisms, cutin synthesis (?)	Tijet *et al.*, 1998
CYP94A2	*Vicia sativa*	Terminal (ω) and in-chain fatty acid hydroxylation	Possibly evokes defence mechanisms, cutin synthesis (?)	Kahn *et al.*, 2001
CYP94A5	*Nicotiana tabacum*	Terminal (ω) fatty acid hydroxylation, diacid formation	Possibly evokes defence mechanisms, cutin & suberin synthesis (?)	Le Bouquin *et al.*, 2001
CYP94C1	*Arabidopsis thaliana*	Diacid formation	Possibly evokes defence mechanisms, cutin & suberin synthesis (?)	Kandel *et al.*, 2007
CYP703A2	*Arabidopsis thaliana*	In-chain fatty acid hydroxylation	Sporopollenin synthesis	Morant *et al.*, 2007
CYP704B1	*Arabidopsis thaliana*	Terminal (ω) fatty acid hydroxylation	Sporopollenin synthesis	Dobritsa *et al.*, 2009
CYP704B2	*Oryza sativa*	Terminal (ω) fatty acid hydroxylation	Sporopollenin synthesis	Li *et al.*, 2010
CYP709C1	*Triticum aestivum*	In-chain fatty acid hydroxylation	Evokes defence mechanisms	Kandel *et al.*, 2005; Pinot & Beisson, 2011
CYP726A1	*Euphorbia lagascae*	Fatty acid epoxidation	Seed oil synthesis	Cahoon *et al.*, 2002

Table 1.3 - Fatty acid hydroxylating cytochromes P450 from plants. Overview of the cytochromes P450 from plants, which have been shown to be involved in the hydroxylation of fatty acids *in vitro*.

1.5.2.1 CYP76B9

CYP76B9 from *Petunia hybrida* is involved in the ω-hydroxylation of capric acid (C10:0) and lauric acid (C12:0) (Imaishi & Petkova-Andonova, 2007). The physiological function of the resultant ω-hydroxy products are unclear, but they are thought to play a role in growth regulation as treatment of the roots of *Arabidopsis thaliana* with these products results in inhibited growth and elongation (Imaishi & Petkova-Andonova, 2007). However, treatment with lauric acid (C12:0) and capric acid (C10:0) did not have any effect (Imaishi & Petkova-Andonova, 2007). The response of the root to the ω-hydroxy products may suggest the hydroxylation of these fatty acids signals a shift from growth to defence in the plant.

1.5.2.2 CYP77A4 and CYP77A6

CYP77A4 and CYP77A6 from *A. thaliana* are the only members of the CYP77A subfamily to be characterised to date (Li-Beisson *et al.*, 2009; Sauveplane *et al.*, 2009). Both of these enzymes have been shown to be involved in the in-chain hydroxylation of long-chain fatty acids, including lauric acid (C12:0) and palmitic acid (C16:0) (and its hydroxy alternative) (Li-Beisson *et al.*, 2009; Sauveplane *et al.*, 2009). However, only CYP77A4 has been shown to catalyse the epoxidation of the unsaturated C18 fatty acids: oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) (Sauveplane *et al.*, 2009). The products formed by both of these enzymes are thought to be involved in the formation of cutin (Pollard *et al.*, 2008; Li-Beisson *et al.*, 2009; Sauveplane *et al.*, 2009; Pinot & Beisson, 2011).

Upon incubation with oleic acid (C18:1), CYP77A4 has been shown to catalyse the formation of 9,10-epoxystearic acid (see figure 1.31) (Sauveplane *et al.*, 2009). This epoxide is thought to play a role in the defence mechanisms within the plant as the epoxide versions of linoleic acid (C18:3) have been implicated in the resistance of rice plants to rice blast disease (Kato *et al.*, 1993). 9,10-epoxystearic acid can also be further converted to 18-hydroxy-9,10-epoxystearic acid in the presence of an ω-hydroxylase for use in cutin synthesis (see figure 1.31) (Sauveplane *et al.*, 2009). Alternatively, it can be converted to 9,10-dihydroxystearic acid by an epoxide hydrolase, which in turn can be hydroxylated at the terminal carbon to form 9,10,18-trihydroxystearic acid (Sauveplane *et al.*, 2009). This product is thought to evoke the defence mechanism of the plant (see figure 1.31) (Sauveplane *et al.*, 2009).

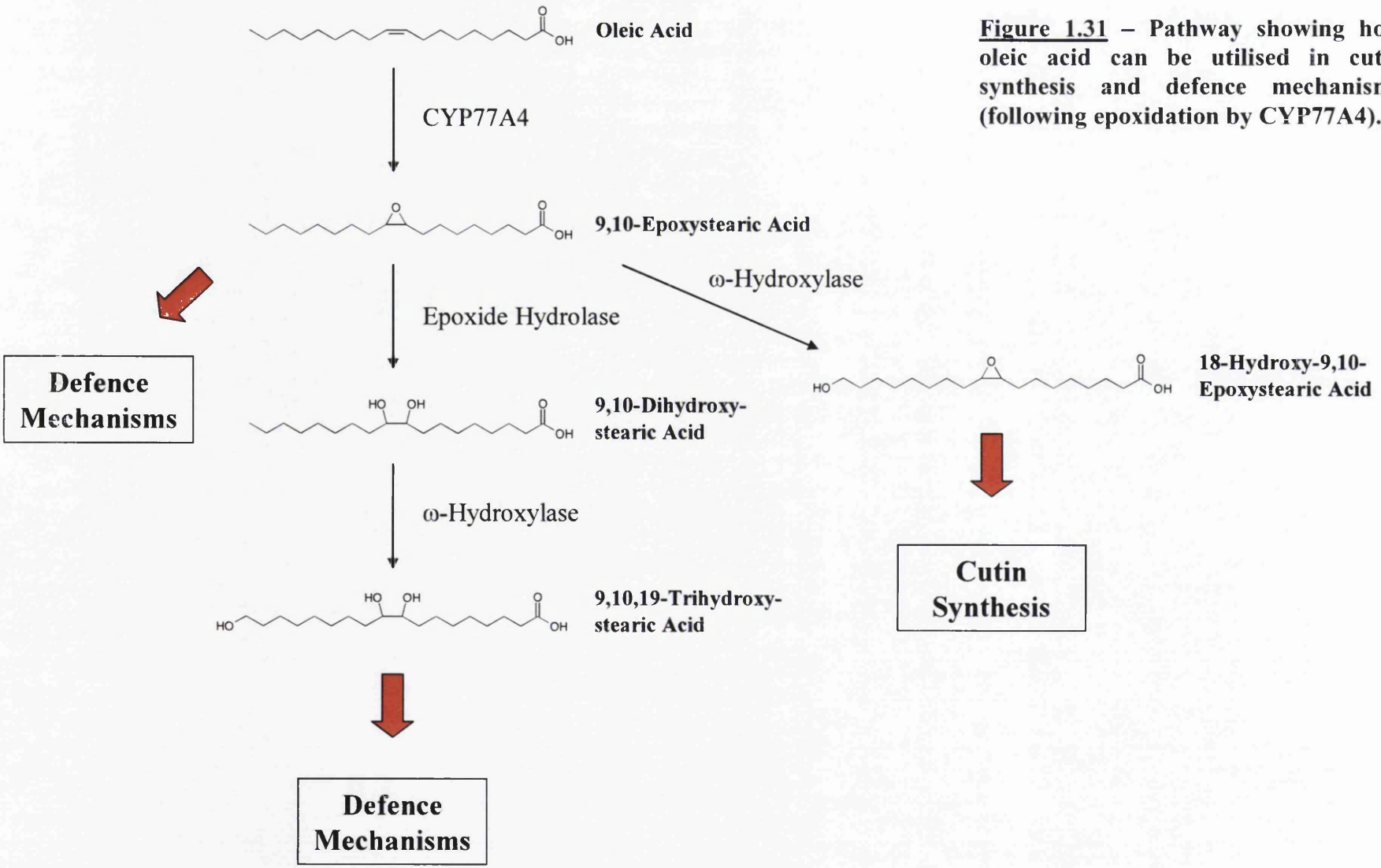


Figure 1.31 – Pathway showing how oleic acid can be utilised in cutin synthesis and defence mechanisms (following epoxidation by CYP77A4).

CYP77A6 plays a role in the pathway converting palmitic acid (C16:0) to dihydroxypalmitic acid, which can then be linked to other molecules via ester bonds to form cutin (see figure 1.32) (Li-Beisson *et al.*, 2009; Pollard *et al.*, 2008). The initial terminal hydroxylation of palmitic acid to 16-hydroxypalmitic acid is catalysed by another cytochrome P450, CYP86A4 (Li-Beisson *et al.*, 2009). The product of this conversion is further hydroxylated at the 10^{th} (ω6-) carbon by CYP77A6 to form 10,16-hydroxypalmitic acid, which can then be used in cutin formation (see figure 1.32) (Li-Beisson *et al.*, 2009).

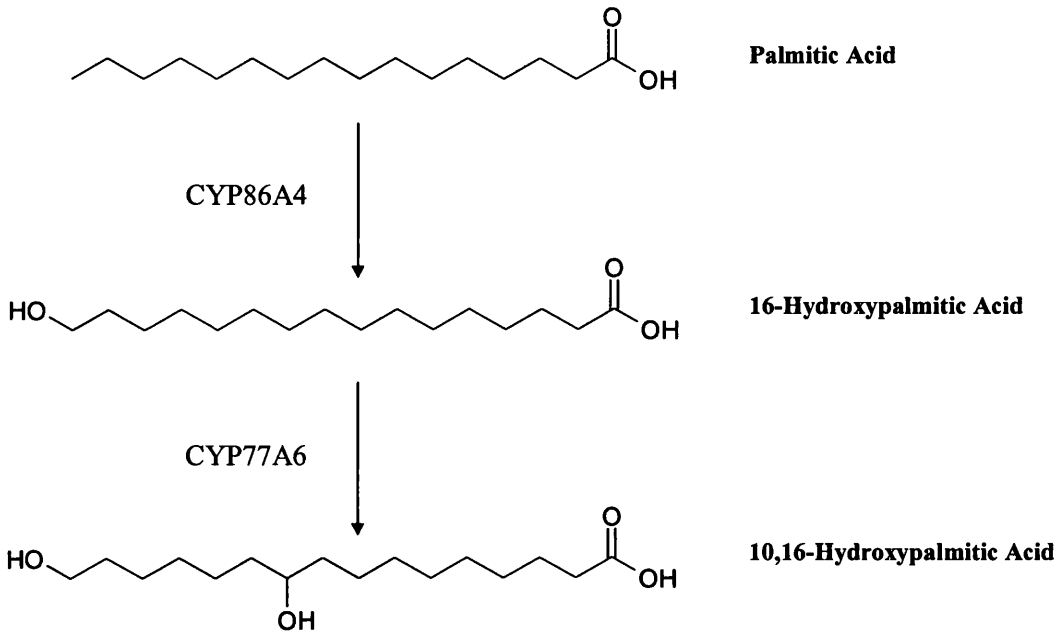


Figure 1.32 – Overview of the enzymes involved in the conversion of palmitic acid (C16:0) to dihydroxypalmitic acid for use in cutin synthesis.

1.5.2.3 CYP78A1

CYP78A1 from *Zea mays* (maize/corn) has been shown to hydroxylate lauric acid (C12:0) at the terminal carbon to produce the ω-hydroxy version (Imaishi *et al.*, 2000; Imaishi & Ohkawa, 2002). The enzyme has been implicated in flower development and reproduction as it has been shown to be expressed *in vivo* during the initial phase of maize tassel formation (Imaishi *et al.*, 2000). The maize tassel is found at the top of the stem in male flowers and upon maturation pollen is released from its anthers and dispersed by the wind to produce more maize crops.

1.5.2.4 CYP81B1

CYP81B1 is a fatty acid hydroxylase found in the tuber (nutrient store) of *Helianthus tuberosus* and is involved in the in-chain (ω1- to ω5-) hydroxylation of capric acid (C10:0), lauric acid (C12:0) and myrisitic acid (C14:0) (Cabello-Hurtado *et al.*, 1998). Although, the physiological function of this enzyme has yet to be elucidated, it may be involved in the initiation of defence mechanisms and/or cutin synthesis like CYP77A4 and CYP77A6 from *A. thaliana*, which have similar functions.

1.5.2.5 CYP86A1, CYP86A4, CYP86A8 and CYP86A22

Four CYP86 enzymes have been characterised and they are CYP86A1, CYP86A4 and CYP86A8 from *A. thaliana* and CYP86A22 from *Petunia hybrida* (Benveniste *et al.*, 1998; Wellesen *et al.*, 2001; Li-Beisson *et al.*, 2009; Han *et al.*, 2010). The enzymes from *A. thaliana* are involved in the terminal (ω) hydroxylation of a range of fatty acids including the saturated fatty acids: lauric acid (C12:0), myrisitic acid (C14:0) and palmitic acid (C16:0) and the unsaturated fatty acids: oleic acid (C18:1) and linoleic acid (C18:2) (Benveniste *et al.*, 1998; Wellesen *et al.*, 2001; Li-Beisson *et al.*, 2009; Han *et al.*, 2010). Originally these three enzymes were thought to have a putative role in cutin synthesis (Benveniste *et al.*, 1998; Wellesen *et al.*, 2001; Li-Beisson *et al.*, 2009; Han *et al.*, 2010). However, CYP86A1 has been shown to be essential in the synthesis of suberin (Höfer *et al.*, 2008). It has been found localised in the roots of *A. thaliana* where suberin accumulates and in CYP86A1 knockout strains the concentration of this biopolymer is reduced by 60% indicating the importance of CYP86A1 in suberin synthesis (Höfer *et al.*, 2008).

Unlike the other CYP86 enzymes mentioned above, CYP86A22 is involved in the ω-hydroxylation of oleic acid (C18:1), C16:0 fatty acyl-CoA, C18:0 fatty acyl-CoA and C18:1 fatty acyl-CoA, but not palmitic acid (C16:0) and stearic acid (C18:0) (Han *et al.*, 2010). Interestingly, CYP86A22 is the only cytochrome P450 so far to be shown to hydroxylate fatty acyl-CoAs (Han *et al.*, 2010). The ω-hydroxy fatty acyl-CoAs formed by CYP86A22 have been identified in the stigma of *P. hybrida*, therefore implicating a role for CYP86A22 in the synthesis of estolide (a lipid polyester) in the stigma of the plant (Han *et al.*, 2010).

1.5.2.6 CYP92B1

CYP92B1 from *Petunia hybrida* is involved in the ω-hydroxylation of lauric acid (C12:0) to form 12-hydroxylauric acid (Petkova-Andonova *et al.*, 2002). Although, the physiological function of CYP92B1 in this plant is unknown, this enzyme does possess a similar function to CYP78A1 from *Zea mays*, which has been implicated in flower development (Imaishi *et al.*, 2000). In metabolism studies, CYP92B1 was also shown to be able to metabolise linoleic acid (C18:2) and linolenic acid (C18:3) (Petkova-Andonova *et al.*, 2002). However, the products of these conversions are unknown as they were not produced at a high enough concentration to be detected on a GCMS trace, although, they were detectable on a TLC (thin layer chromatography) plate (Petkova-Andonova *et al.*, 2002). Comparing the metabolites produced to ω-hydroxylinoleic acid and ω-hydroxylinolenic acid standards showed that CYP92B1 is not responsible for the ω-hydroxylation of linoleic acid (C18:2) and linolenic acid (C18:3) (Petkova-Andonova *et al.*, 2002). This suggests that CYP92B1 may be involved in the epoxidation or in-chain hydroxylation of these hydrocarbons.

1.5.2.7 CYP94A1, CYP94A2, CYP94A5 and CYP94C1

Four members of the CYP94 family have been expressed and characterised to date. They are CYP94A1 and CYP94A2 from *Vicia sativa* (Tijet *et al.*, 1998; Kahn *et al.*, 2001), CYP94A5 from *Nicotiana tabacum* (tobacco) (Le Bouquin *et al.*, 2001) and CYP94C1 from *A. thaliana* (Kandel *et al.*, 2007). Each of these enzymes is thought to be involved in the synthesis of cutin (and suberin) and eliciting defence mechanisms, but the substrate specificity and pattern of regioselectivity for each is different (Tijet *et al.*, 1998; Kahn *et al.*, 2001; Le Bouquin *et al.*, 2001; Kandel *et al.*, 2007).

CYP94A1 is involved in the ω-hydroxylation of a broad variety of fatty acids, including the saturated fatty acids ranging from capric acid (C10:0) to palmitic acid (C16:0) and the unsaturated fatty acids: oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) (Tijet *et al.*, 1998). However, CYP94A2, which shares a 76% amino acid similarity to CYP94A1, only hydroxylates lauric acid (C12:0), myristic acid (C14:0) and palmitic acid (C16:0) at the terminal and/or in-chain carbons (Kahn *et al.*, 2001). Interestingly, the pattern of regioselectivity differs dependent on the chain-length of the fatty acid (Kahn *et al.*, 2001). Myristic acid (C14:0 – the

preferred substrate) is hydroxylated at the ω- and ω1-carbons, whereas lauric acid (C12:0) is only hydroxylated at the ω-carbon and palmitic acid (C16:0) at the ω1- and ω2-carbons (Kahn *et al.*, 2001).

CYP94A5 and CYP94C1, like CYP52A3 and CYP52A4 from *C. maltosa* and CYP52A13 and CYP52A17 from *C. tropicalis*, not only hydroxylate fatty acids to produce the respective alcohol, but can also convert the hydrocarbon further to produce diacids (which are major components of suberin) (Le Bouquin *et al.*, 2001; Kandel *et al.*, 2007; Pollard *et al.*, 2008). CYP94A5 initially converts 9,10-epoxystearic acid to the ω-alcohol, 18-hydroxy-9,10-epoxystearic acid, then further catalyses it to the respective diacid (Le Bouquin *et al.*, 2001). This enzyme can also hydroxylate a variety of other fatty acids at the terminal carbon to produce the ω-alcohol (Le Bouquin *et al.*, 2001). These fatty acids are lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0), oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) (Le Bouquin *et al.*, 2001). CYP94C1 has been shown to convert lauric acid (C12:0) to its respective diacid, dodecanedioic acid (Kandel *et al.*, 2007).

1.5.2.8 CYP703A2

CYP703A2 is an in-chain hydroxylase from *A. thaliana*, which has been shown to monohydroxylate capric acid (C10:0) (at the 6th (ω4-), 7th (ω3-) or 8th (ω2-) carbon), lauric acid (C12:0) (at the 6th (ω6-), 7th (ω5-), 8th (ω4-) or 9th (ω3-) carbons), myristic acid (C14:0) (at the 6th (ω8-), 7th (ω7-) or 8th (ω6-) carbon) and palmitic acid (C16:0) (potentially at the 6th (ω10-), 7th (ω9-) or 8th (ω8-) carbon) (Morant *et al.*, 2007). It preferentially hydroxylates lauric acid (C12:0) at the 7th (ω5-) carbon and its product (7-hydroxylauric acid) has physiological implications for the plant. 7-hydroxylauric acid is essential in the synthesis of sporopollenin, which is found in the exine layer (tough outer walls) of pollen and spores (Edlund *et al.*, 2004; Morant *et al.*, 2007). This layer plays an important role in pollination and protection (from pathogens and dehydration) (Edlund *et al.*, 2004; Morant *et al.*, 2007). A CYP703A2 knockout of *A. thaliana* was shown to have impaired development of pollen, thus further confirming the role of this enzyme in sporollenin synthesis (Morant *et al.*, 2007).

1.5.2.9 CYP704B1 and CYP704B2

As with CYP703A2, CYP704B1 from *A. thaliana* and CYP704B2 from *Oryza sativa* (rice) are involved in sporopollenin synthesis (Pinot & Beisson, 2011). However, unlike CYP703A2, CYP704B1 and CYP704B2 catalyse the terminal (ω-) hydroxylation of fatty acids rather than in-chain hydroxylation (Dobritsa *et al.*, 2009; Li *et al.*, 2010). These enzymes have been shown to produce the ω-hydroxy versions of palmitic acid (C16:0), oleic acid (C18:1) and linoleic acid (C18:2) and linolenic acid (C18:3) (Dobritsa *et al.*, 2009; Li *et al.*, 2010).

1.5.2.10 CYP709C1

CYP709C1 from *Triticum aestivum* (bread wheat) is active in the sub-terminal/in-chain hydroxylation of fatty acids (Kandel *et al.*, 2005). It has been shown to metabolise a variety of fatty acids, including lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0) and stearic acid (C18:0) (Kandel *et al.*, 2005). However, CYP709C1 preferentially hydroxylates 9,10-epoxystearic acid at the ω1- and ω2- carbons (Kandel *et al.*, 2005). The product of this metabolism, 17-hydroxy-9,10-epoxystearic acid, can be further hydrolysed by epoxide hydrolase to produce 9,10,17-trihydroxystearic acid, which is involved in evoking defence mechanisms within the plant (see figure 1.33) (Pinot & Beisson, 2011).

1.5.2.11 CYP726A1

CYP726A1, found in the seeds of *Euphorbia lagascae*, is involved in the epoxidation of linoleic acid (C18:2) and linolenic acid (C18:3) at the double bond between the 12 and 13 carbons (Cahoon *et al.*, 2002). This results in the production of vernolic acid (12,13-epoxyoctadeca-cis-9-enoic acid) (figure 1.34), which is a major component of seed oils (Cahoon *et al.*, 2002).

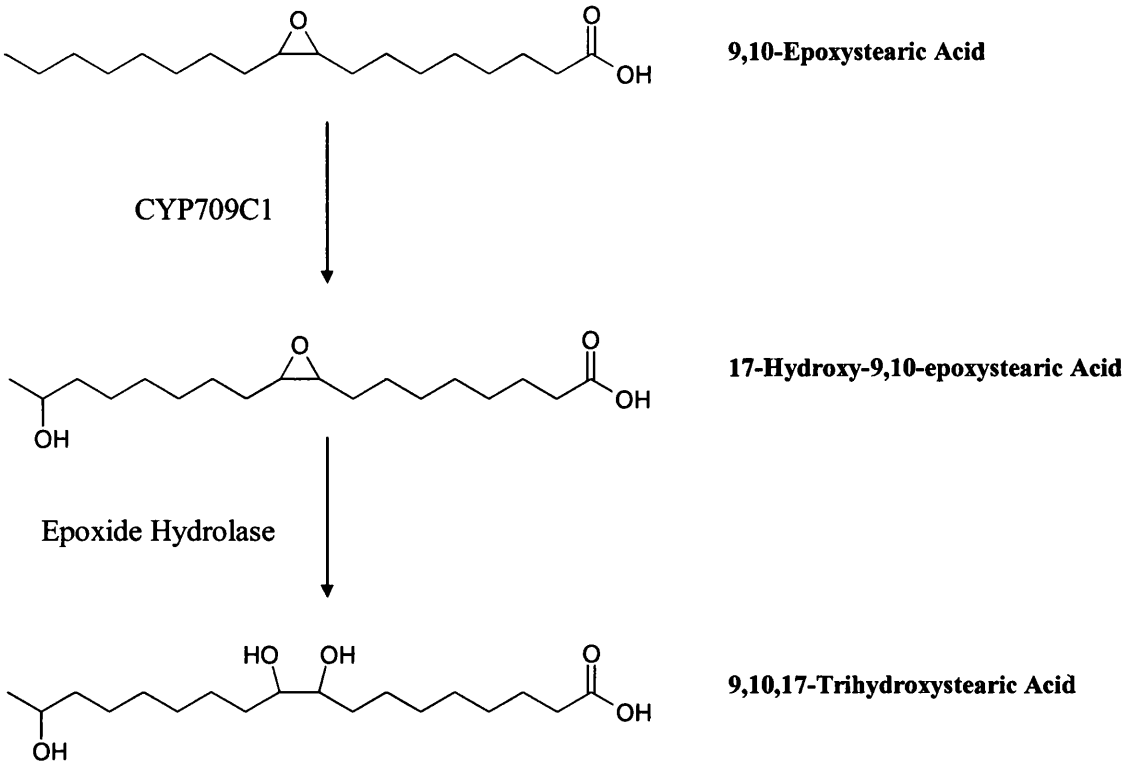


Figure 1.33 - Overview of the enzymes involved in the conversion of 9,10-epoxystearic acid to 9,10,17-trihydroxystearic acid.

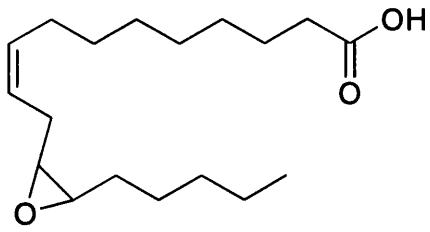


Figure 1.34 – Vernolic acid (12,13-epoxyoctadeca-cis-9-enoic acid), the major component of seed oils.

1.5.3 Fatty Acid Hydroxylating Cytochromes P450 from Animals

Fatty acid hydroxylating cytochromes P450 have been identified in a number of animals including mammals, insects and amphibians (Miura, 1981; Simpson, 1997). However, many mammalian fatty acid hydroxylation cytochromes P450 have been shown to be active at the protein level and the physiological roles of many of these enzymes have been documented. Therefore, mammalian enzymes will be discussed in this section.

A build up of fatty acids in the organs and tissues of mammals with uncontrolled diabetes, chronic alcohol abuse or who have been on a starvation diet can be toxic. Therefore these hydrocarbons need to be converted to diacids for quick elimination through β-oxidation (Dhar *et al.*, 2008; Fer *et al.*, 2008). The initial step in this pathway, as with CYP52 enzymes, is the conversion of the fatty acid to its relative alcohol, which is catalyzed by enzymes belonging to the CYP4 family. Members of this family are not solely involved in the hydroxylation of fatty acids, but can also play a role in the epoxidation of unsaturated fatty acids, such as arachidonic acid (C20:4).

Table 1.4 shows the mammalian fatty acid hydroxylating cytochromes P450 that have been expressed at the protein level and characterised. These enzymes are discussed below.

Cytochrome P450 Name	Species	Location	Activity	Reference
CYP1A1	Human, Rat	Lung, Placenta, Brain	Terminal (ω) or sub-terminal fatty acid hydroxylation (species dependent)	Schwarz *et al.*, 2004
CYP2E1	Rabbit, Human	Liver	Sub-terminal fatty acid hydroxylation	Fukuda *et al.*, 1994; Amet *et al.*, 1995
CYP4A1	Rat	Liver, Kidney	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Chaurasia *et al.*, 1995;Wang *et al.*, 199
CYP4A2	Rat	Kidney	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Wang *et al.*, 1996
CYP4A3	Rat	Kidney	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Wang *et al.*, 1996; Lasker *et al.*, 2000
CYP4A6	Rabbit	Kidney, Liver	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Lasker *et al.*, 2000
CYP4A7	Rabbit	Kidney, Liver, Small Intestine	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Lasker *et al.*, 2000
CYP4A11	Human	Kidney	Terminal (ω) and sub-terminal fatty acid hydroxylation, fatty acid epoxidation	Wang *et al.*, 1996
CYP4B1	Rabbit	Lung, Placenta, Brain	Terminal (ω) fatty acid hydroxylation	Fisher *et al.*, 1998
CYP4F1	Human	Liver, Kidney	Terminal (ω) fatty acid hydroxylation	Lasker *et al.*, 2000
CYP4F3B	Human	Kidney	Terminal (ω) fatty acid hydroxylation	Harmon *et al.*, 2006
CYP4F11	Human	Liver	Terminal (ω) fatty acid hydroxylation	Tang *et al.*, 2010
CYP4V2	Human	Eye	Terminal (ω) fatty acid hydroxylation	Nakano *et al.*, 2009

Table 1.4 – Fatty acid hydroxylating cytochromes P450 from mammals. Overview of the mammalian cytochromes P450, which have been shown to be involved in the hydroxylation of fatty acids *in vitro*.

1.5.3.1 CYP1A1

CYP1A1 has been identified in the lungs, placenta and brain of humans and rats (Schwarz *et al.*, 2004). In both animals, it has been shown to be an arachidonic acid (C20:4) hydroxylase, however, regioselectivity differences have been identified between the two copies (Schwarz *et al.*, 2004). In rats, CYP1A1 preferentially hydroxylases arachidonic acid (C20:4) at the terminal position to produce 20-hydroxyeicosatetraenoic acid (HETE), whereas, in humans, CYP1A1 is involved in the sub-terminal hydroxylation of arachidonic acid (C20:4) to produce 19-HETE (see figure 1.35) (Schwarz *et al.*, 2004). The human CYP1A1 has also been shown to act as an epoxygenase when incubated with eicosapentaenoic acid (C20:5 - EPA) (see figure 1.36) (Schwarz *et al.*, 2004). This reaction results in the formation of 17,18-epoxyeicosatetraenoic acid, which is thought to act as a vasodilator by stimulating calcium-activated potassium channels (Schwarz *et al.*, 2004).

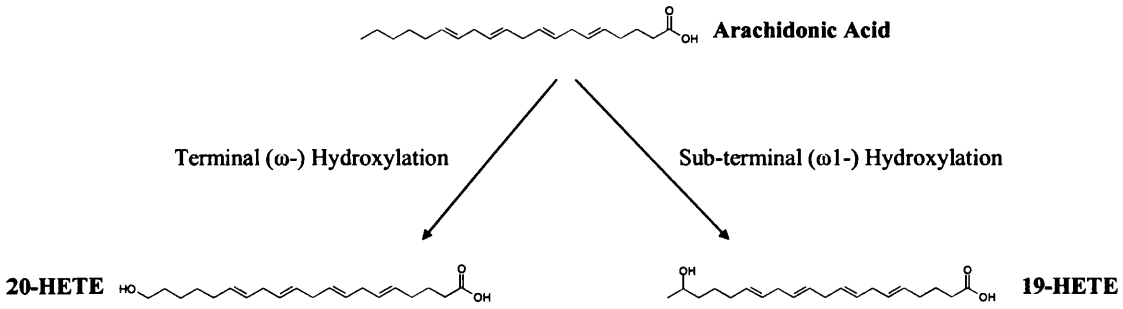


Figure 1.35 – 20-Hydroxyeicosatetraenoic acid (HETE) and 19-HETE, the products of terminal and sub-terminal hydroxylation of arachidonic acid (C20:4).

Eicosapentaenoic Acid (EPA)

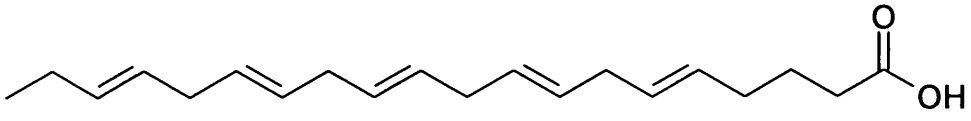


Figure 1.36 – Eicosapentaenoic acid (EPA).

1.5.3.2 CYP2E1

CYP2E1 is a sub-terminal fatty acid hydroxylating cytochrome P450 that has been identified in the liver of both rabbits and humans (Fukuda *et al.*, 1994; Amet *et al.*, 1995). It is induced by ethanol and has been shown to be expressed at high levels in the liver following alcohol ingestion (in both acute and chronic alcoholics) and fasting, as well as in the livers of diabetic patients (Amet *et al.*, 1995). Unlike members of the CYP4 family, CYP2E1 has been shown to be predominantly involved in the sub-terminal (ω1-) hydroxylation of fatty acids, ranging from pelargonic acid (C9:0) to tridecylic acid (C13:0) and in particular lauric acid (C12:0) (Fukuda *et al.*, 1994; Amet *et al.*, 1995). Although, the physiological use of 11-hydroxylauric acid is unknown, CYP2E1 is thought to be the only human liver cytochrome P450 to be involved in the formation of this hydrocarbon and is potentially involved in causing alcoholic liver disease (Morimoto *et al.*, 1993; Amet *et al.*, 1995).

1.5.3.3 CYP4A1-3, CYP4A6-7 and CYP4A11

The CYP4A subfamily of cytochromes P450 have been shown to be involved in the terminal and, to a lesser extent, sub-terminal hydroxylation of fatty acids, including lauric acid (C12:0) and palmitic acid (C16:0), and both the terminal hydroxylation and epoxidation of arachidonic acid (C20:4) (Wang *et al.*, 1996). The terminal hydroxylation of arachidonic acid (C20:4) results in the formation of 20-HETE, which is a potent vasoconstrictor and has been implicated in (genetic) high blood pressure in rats (Lasker *et al.*, 2000). However, the product of arachidonic acid (C20:4) epoxidation, 11,12-epoxy-5,8,14-eicosatrienoic acid (11,12-EET) (figure 1.37) is not a vasoconstrictor, but rather a potent vasodilator and is implicated in reducing blood pressure (Lasker *et al.*, 2000). Although, this product is produced at much lower levels than 20-HETE by the CYP4A subfamily, it does suggest a role for these enzymes in the regulation of blood pressure in mammals.

11,12-Epoxy-5,8,14-eicosatrienoic Acid (11,12-EET)

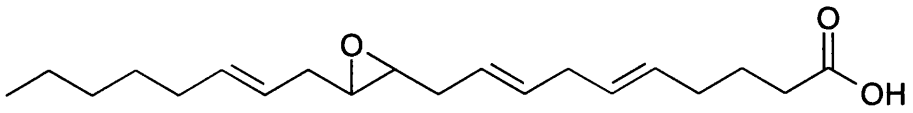


Figure 1.37 – 11,12-Epoxy-5,8,14-eicosatrienoic acid (11,12-EET), formed following the epoxidation of arachidonic acid (C20:4).

Members of the CYP4A subfamily include: CYP4A1 (rat liver and kidney (Chaurasia *et al.*, 1995; Wang *et al.*, 1996)), CYP4A2 (rat kidney (Wang *et al.*, 1996)), CYP4A3 (rat kidney (Wang *et al.*, 1996; Lasker *et al.*, 2000)), CYP4A6 (rabbit kidney and liver (Lasker *et al.*, 2000)), CYP4A7 (rabbit kidney, liver and small intestine) (Lasker *et al.*, 2000)) and CYP4A11 (human kidney (Wang *et al.*, 1996)).

1.5.3.4 CYP4B1

CYP4B1 from rabbit lung was initially considered a xenobiotic metabolising enzyme as it has been shown to be involved in the bioactivation of a number of compounds including valproic acid (an anti-convulsant drug) (figure 1.38) (Fisher *et al.*, 1998; Kikuta *et al.*, 1999). Valproic acid is a branched, 7-carbon chain fatty acid, with the carboxylate moiety situated at the centre of the alkyl chain (Fisher *et al.*, 1998). As this hydrocarbon is considered an isomer of caprylic acid (C8:0), a range of fatty acids were used in experiments with CYP4B1 to elucidate whether this enzyme could also metabolise these substrates. It was shown that CYP4B1 is involved in the ω-hydroxylation of caprylic acid (C8:0 – preferred substrate), pelargonic acid (C9:0), capric acid (C10:0) and enanthic acid (C7:0) (Fisher *et al.*, 1998). It was also shown to be involved in the ω-hydroxylation of alkanes, such as heptane (C7:0), octane (C8:0), nonane (C9:0) and decane (C10:0) (Fisher *et al.*, 1998). The physiological roles of the ω-alcohols produced from fatty acids and alkanes by CYP4B1 are undocumented, however, this enzyme differs from the other known mammalian ω-hydroxylating cytochrome P450 families, CYP4A and CYP4F as it is involved in the metabolism of drugs, alkanes and short-chain fatty acids rather than solely the ω-hydroxylation of medium- to long-chain fatty acids, such as lauric acid (C12:0).

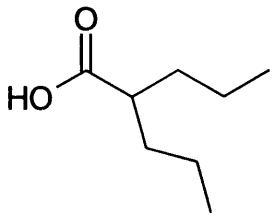


Figure 1.38 – Valproic acid, a branched, 7-carbon chain fatty acid, with the carboxylate moiety situated at the centre of the alkyl chain.

1.5.3.5 CYP4F1, CYP4F3B, and CYP4F11

As with the CYP4A subfamily, the CYP4F subfamily of enzymes is involved in the hydroxylation of fatty acids. However, the enzymes belonging to this subfamily are also involved in the metabolism of xenobiotics and catalyse the hydroxylation of eicosanoids, thus regulating their concentration during inflammation (Kalsotra & Strobel, 2006). Discussed below are those enzymes belonging to the CYP4F family that catalyse the hydroxylation of fatty acids.

CYP4F1 from human liver and kidneys, CYP4F3B from human kidneys and CYP4F11 from human liver have all been shown to be involved in the terminal (ω-) hydroxylation of arachidonic acid (C20:4) to form the vasoconstrictor 20-HETE (Lasker *et al.*, 2000; Harmon *et al.*, 2006; Tang *et al.*, 2010). Interestingly, CYP4F3B is also able to hydroxylate the ω-3 fatty acids, EPA (C20:5) and docosahexaenoic acid (DHA) (C22:6), which are important dietary supplements (Harmon *et al.*, 2006). These hydrocarbons elicit an antihypertensive effect by inhibiting the formation of 20-HETE by CYP4F3B (Harmon *et al.*, 2006). Although, arachidonic acid (C20:4) is the preferred substrate of CYP4F3B at high concentrations EPA (C20:5) and DHA (C22:6) are able to prevent arachidonic acid (C20:4) from binding to the enzyme, thus inhibiting the formation of 20-HETE (Harmon *et al.*, 2006). This in turn leads to the production of 20-hydroxyEPA and 22-hydroxyDHA, which elicit a vasodilatory effect (Harmon *et al.*, 2006).

CYP4F11 has also been shown to be involved in the ω-hydroxylation of a number of other fatty acids, including palmitic acid (C16:0), oleic acid (C18:1), linoleic acid (C18:2), DHA (C22:6) and 3-hydroxy palmitic acid (Tang *et al.*, 2010). At present,

it is unclear whether these hydrocarbons are the main substrates of this enzyme as the effect of CYP4A11 towards other substrates has yet to be documented (Tang *et al.*, 2010).

1.5.3.6 CYP4V2

CYP4V2, found in the human eye, is involved in the terminal hydroxylation of lauric acid (C12:0), myristic acid (C14:0) and palmitic acid (C16:0) (Nakano *et al.*, 2009). Although, the physiological role for the alcohol formed in the eye is undocumented, mutations within the enzyme, which lead to impaired terminal hydroxylation of fatty acids have been implicated in the progressive eye disease, Bietti's corneoretinal crystalline dystrophy (BCD) (Li *et al.*, 2004; Wada *et al.*, 2005; Nakamura *et al.*, 2006; Nakano *et al.*, 2009). BCD is a rare eye disorder, which causes narrowing of the visual field, atrophy of the epithelial cells within the retina and night blindness (Nakano *et al.*, 2009). The implication that CYP4V2 mutations are involved in this disease, suggest that CYP4V2 is required for maintaining healthy eyes.

1.6 Aims of this Thesis

As shown in this chapter, cytochrome P450 enzymes play an important role in the hydroxylation of alkanes and/or fatty acids. However, the specific physiological roles of the products produced are not well documented despite fatty acids playing known roles in signalling, cell growth and the formation of structural lipids, such as phospholipids. The fatty acid derivatives produced have been shown to have roles as platform chemicals in biotechnology (i.e. diacids and surfactants) as chemical synthesis of the same compounds is limited and expensive and further products of value in health care are possible.

CYP52 enzymes from yeast and fungi are of particular interest as of the 52 members of the family identified to date only 8 have been shown to be studied at the protein level. To understand the physiological role this family of enzymes play in yeast and fungi and their uses in biotechnology, the enzymatic activity of each protein needs to be thoroughly understood. To address this issue in *C. albicans* the objectives of this thesis are:

- To determine the ability of *C. albicans* to produce diacids.

- To identify the cytochrome P450 genes in *C. albicans* and the candidate gene with the greatest homology to the archetypal alkane-hydroxylating P450 gene, *CYP52A3* from *C. maltosa* using a bioinformatic approach.
- To express the candidate cytochrome P450 in its native and soluble forms as well as CYP52A3 as a control using *E. coli* and yeast as the host organisms.
- To identify potential substrates and inhibitors of the candidate cytochrome P450 using a spectrophotomer to detect spectral changes.
    - Potential substrates to be used include:
        - Long-chain alkanes
        - Long-chain fatty acids
        - Sterols
    - Potential inhibitors to be used include:
        - Azole antifungal drugs
- To establish the enzymatic activity of the candidate cytochrome P450 (and CYP52A3) *in vitro* and *in vivo* by GCMS.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals and solvents were purchased from Fisher Scientific, unless otherwise stated. Isopropyl-β-D-thio-galactoside (IPTG), 5-aminolevulinic acid hydrochloride (Ala) and growth media were purchased from Formedium.

All restriction enzymes and restriction enzyme buffers were purchased from Promega.

All fatty acids and hexadecane were purchased from Sigma-Aldrich with purities of 98% or above. All sterols were synthesised and purified by Prof. David Nes (Texas Tech University), except cholesterol, which was purchased from Sigma-Aldrich. All azole antifungal agents were purchased from Sigma-Aldrich, except voriconazole, which was purchased from Discovery Fine Chemicals.

2.1.2 *Candida* Strains

Candida albicans SC5314 and *C. maltosa* IAM12247 were purchased from LGC Promochem (in partnership with the American Type Culture Collection (ATCC)).

2.1.3 Media

2.1.3.1 YEPD Media

Broth was made from 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) glucose and was used to culture yeast cells overnight. 2% agar (w/v) was added to make agar plates to grow yeast colonies on.

2.1.3.2 YM Media

Yeast minimal media was made from 0.67% (w/v) yeast nitrogen base (without amino acids) and 1% (w/v) glucose.

2.1.3.3 Luria-Bertani (LB) Media

E. coli cells were grown overnight in broth made from 1% (w/v) bacto tryptone, 0.5% (w/v) yeast extract and 1% (w/v) sodium chloride. Colonies were grown on

plates made of the same media plus 2% (w/v) agar. Selection was made by adding 100μg/ml ampicillin to the media.

In blue/white selection experiments for identifying DNA insertion into plasmids, colonies were grown on LB agar plates containing 100μg/ml ampicillin, 0.5mM IPTG and 80μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal).

2.1.4 Buffers

2.1.4.1 50x TAE Buffer

242g of Tris base (Melford), 57.1ml of glacial acetic acid and 100ml of 0.5M EDTA (Sigma-Aldrich) pH8.0 was made up to 1L with distilled water. Stored at room temperature.

2.1.4.2 1x TAE Buffer

20ml of 50x TAE was made up to 1L with distilled water. Stored at room temperature.

2.1.4.3 DNA Loading Dye (10x Concentrate)

25mg bromophenol blue (Sigma-Aldrich) and 3g glycerol were added to 10ml distilled water and mixed.

2.1.4.4 Resolving Buffer (5x Concentrate)

1.88M Tris-HCl (Melford) pH8.8 dissolved in distilled water. Stored at 4°C.

2.1.4.5 Stacking Buffer (10x Concentrate)

0.6M Tris-HCl (Melford) pH6.8 dissolved in distilled water. Stored at 4°C.

2.1.4.6 Stock Acrylamide Solution

30% (w/v) acrylamide (Amersham) and 0.2% (w/v) N'N'-bis-methylene-acrylamide (Amersham) dissolved in distilled water. Stored at 4°C.

2.1.4.7 Super Cracking Sample Buffer

10% (w/v) SDS (Melford), 5% (v/v) β-mercaptorthanol (Sigma-Aldrich), 120mM Tris-HCl (Melford) pH6.8, 20% (v/v) glycerol and bromophenol blue (a trace) (Sigma-Aldrich) were dissolved in distilled water. Stored at -20°C.

2.1.4.8 Resolving Gel

3.7ml of the stock acrylamide solution, 2ml of the resolving buffer (5x concentrate) and 0.2 ml of 10% (w/v) SDS (Melford) were thoroughly mixed together and made up to a final volume of 10ml with distilled water. 10µl of TEMED (N,N,N',N',-tetramethylethylene diamine) (Sigma-Aldrich) and 100µl of APS (10% (w/v) ammonium persulphate) were then added to the mix.

2.1.4.9 Stacking Gel

1ml of the stock acrylamide solution, 0.6ml of the stacking buffer (10x concentrate) and 0.12ml of 10% SDS (Melford) were added to 4.28ml of distilled water. 10µl of TEMED (Sigma-Aldrich) and 100µl of APS were then added.

2.1.4.10 Protein Marker

Molecular weight 29,000-205,000 marker for SDS-PAGE (Sigma-Aldrich).

2.1.4.11 Electrode Tank Buffer (10x Concentrate)

0.25M Tris Base (Melford) and 1.92M glycine dissolved in distilled water. Stored at room temperature.

2.1.4.12 Upper Tank Buffer

0.1% (w/v) SDS (Melford) was added to 1x electrode tank buffer. Stored at room temperature.

2.1.4.13 Lower Tank Buffer

0.5x electrode tank buffer diluted in distilled water. Stored at room temperature.

2.1.4.14 Coomassie Blue Stain

1.25L methanol, 175ml glacial acetic acid and 6.25g coomassie brilliant blue R250 were made up to a final volume of 2.5L with distilled water. Solution was filtered through Watman No. 1 filter paper. Stored at room temperature.

2.1.4.15 Coomassie Destain

30% (v/v) methanol or 30% (v/v) ethanol dissolved in distilled water. Store at room temperature.

2.2 Methods

2.2.1 Preparation of Competent *E. coli* Cells

Competent *E. coli* cells were made up for use in the laboratory by Ms Nicky Rolley based on the method in Sambrook *et al.*, 2001.

DH5α and XL-1 *E. coli* cells were grown up overnight in 10ml of LB at 37°C and 200rpm. To the XL-1 cells 5mg/ml of tetracycline was also added.

1ml of the overnight culture was used to inoculate 100ml of fresh LB. This culture was grown at 37°C, 200rpm until growth had reached an OD_{650} of 0.5. The culture was then incubated on ice for 10 minutes. The culture was centrifuged at 3000g at 40°C for 10 minutes. The supernatant was discarded and the pellet was resuspended in 20ml of ice cold CM1 (10mM sodium acetate (Sigma-Aldrich), 50mM manganese chloride and 5mM sodium chloride). This suspension was incubated on ice for 20 minutes and then centrifuged at 3000g for 5 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 2ml of CM2 (10mM sodium acetate (Sigma-Aldrich), 70mM calcium chloride, 5mM manganese chloride and 5% (v/v) glycerol). The suspension was aliquoted into 50µl volumes and stored at -80°C.

2.2.2 Isolation of Genomic DNA from Yeast

Genomic DNA was extracted using a method based on the protocol printed in Sambrook *et al.*, 2001.

10ml of overnight YEPD culture inoculated from a single colony *C. albicans* SC5314 or *C. maltosa* IAM12247 colony. 1ml was pipetted into a 1.5ml Eppendorf

tube and spun down for 1 minute at 13,000rpm in a tabletop microcentrifuge. The resultant pellet was washed in 1ml of distilled water and resuspended in 300µl of SED (2M sorbitol (4.5ml), 1M EDTA (1ml) (Sigma-Aldrich), 1M DTT (500µl) and distilled water (4ml)). To this 7µl of lyticase (10units/µl) was added and placed in the 37°C shaker for 120 minutes. The mixture was then spun down for 1 minute at 13000rpm and resuspended in 200µl of Tris-EDTA. To this 30µl of 10% (w/v) SDS (Melford) was added and incubated for 30 minutes at 65°C in a heat block. Following incubation, 100µl of 3M sodium acetate was added and placed on ice for 1 hour. The mixture was spun down for 1 minute at 13,000rpm and the supernatant was added to a fresh 1.5ml Eppendorf tube containing 400µl of ice cold ethanol and then spun for 10 minutes at 13,000rpm. The supernatant was removed and the remaining pellet was washed with 400µl of 70% (v/v) ethanol. The pellet was dried in the SpeedVac until all traces of ethanol were removed and then resuspended in 30µl of distilled water.

2.2.3 Gel Electrophoresis

The 1% (w/v) gel was cast by melting 0.5g of agarose (Melford) in 50ml of 1x TAE buffer until the agarose had entirely dissolved. The solution was poured into a mould and 5µl of SYBR safe (Invitrogen) or ethidium bromide was added before being left to solidify. Once hardened, the gel and the mould were placed into a gel tank, and the tank was filled with 1x TAE buffer, making sure the surface of the gel was covered. DNA, mixed with 10x loading dye, was loaded into the wells. A GeneRuler™ 1Kb DNA ladder (Fermentas) was also loaded into the first lane as a reference marker. The gel tank was attached to a power pack and the gel was run at 80 volts for 50 minutes. Following this, the gel was examined under UV light.

2.2.4 Extaction of DNA from Gels

Extraction of DNA from a 1% (w/v) agarose gel was carried out using a QIAquick Gel Extraction kit (Qiagen) and the supplied protocol was followed. This method was based on that of Vogelstein & Gillespie, 1979.

The DNA fragment was cut from a 1% (w/v) agarose gel, placed into a clean 1.5ml eppendorf tube and weighed. Three times the weight of the gel slice of buffer QG (e.g. to 100mg of gel slice, add 300µl of buffer QG) was added to the tube and

incubated at 50°C in a heat block for 10 minutes. Once the gel slice had completely dissolved, one gel volume of isopropanol was added and mixed. The mixture was pipetted into a QIAquick spin column and centrifuged for 1 minute at 13,000rpm. The flow-through was discarded. 0.5ml of buffer QG was added to the same QIAquick spin column and was centrifuged at 13,000rpm for 1 minute. The flow-through was discarded. The QIAquick spin column was washed by adding 0.75ml of Buffer PE and was spun at 13,000rpm for 1 minute. The flow-through was discarded and the QIAquick spin column was centrifuged for an additional minute at 13,000rpm. The QIAquick spin column was then placed into a fresh 1.5ml eppendorf tube. 50µl of distilled water was added to the QIAquick spin column and centrifuged at 13,000rpm for 1 minute. The QIAquick spin column was then discarded.

2.2.5 Isolation and Purification of Plasmids

Plasmid DNA was extracted and purified using the QIAprep Spin Miniprep kit (Qiagen) and the accompanying protocol was followed. This method was based on that of Birnboim & Doly, 1979.

10ml of LB + ampicillin *E.coli* culture was set up overnight, of which 1ml was pipetted into a new 1.5ml eppendorf tube and spun down for 1 minute at 13,000rpm. The supernatant was discarded and the pellet was resuspended in 250µl of Buffer P1 (Qiagen). To this 250µl of Buffer P2 (Qiagen) was added and the tube was inverted 4-6 times to make sure the solutions were mixed thoroughly. 350µl of Buffer N3 (Qiagen) was then added and again the solutions were mixed thoroughly by inverting the tube 4-6 times. The mixture was spun down for 10 minutes at 13,000rpm. The supernatant was carefully removed and pipetted into the supplied QIAprep spin column, which was then centrifuged for 1 minute at 13,000rpm. The flow-through was discarded. The QIAprep spin column was washed by adding 0.75ml of Buffer PE (Qiagen) and again centrifuged at 13,000rpm for 1 minute. The flow-through was discarded and the QIAprep spin column was spun down for an additional minute to remove residual Buffer PE (Qiagen). The flow-through was placed into a new 1.5ml eppendorf tube. The DNA was eluted by adding 50µl of distilled water to the QIAprep spin column, letting it stand for 1 minute and finally centrifuging for 1 minute at 13,000rpm.

2.2.6 Restriction Enzyme Digest of the Plasmid

Extracted and purified plasmid DNA underwent either a single or a double restriction enzyme digest to isolate the gene of interest.

2.2.6.1 *Not* I Digest – Insert Check

A *Not* I digest was used to screen colonies for the gene of interest as the vector pGEM-T® easy contains a *Not* I restriction site on either side of the cloning region, so a single enzyme can be used to release the gene.

2µl of distilled water, 2µl of the purified plasmid, 0.5µl of Buffer D and 0.5µl of *Not* I were mixed together. The digest mix was incubated at 37°C until completion.

2.2.6.2 Double Digest for Gene Extraction

A *Nde* I/*Hin*d III double digest was used to release the gene of interest from pGEM-T® easy for ligation into pCWori$^+$. For ligation into YEp51 a *Sal* I/*Hin*d III digest was used instead.

28µl of the purified plasmid, 4µl of Buffer D, 4µl of *Nde* I or *Sal* I and 4µl of *Hin*d III were mixed together. The digest mix was incubated at 37°C until completion.

2.2.6.3 Double Digest – Insert Check

A *Nde* I/*Hin*d III double digest was used to check for the presence of the gene of interest in the pCWori$^+$ plasmid. Alternatively, a *Sal* I/*Hin*d III digest was used to check for the gene of interest in YEp51 plasmid.

6µl of the purified plasmid, 1µl of Buffer D, 1µl *Nde* I or *Sal* I and 1µl of *Hin*d III were mixed together. The digest mix was incubated at 37°C until completion.

All digests were visualised on a 1% (w/v) agarose gel.

2.2.7 Transformation of Plasmid into *E. coli*

Based on the method of Hanahan, 1985.

Plasmid DNA was transformed into *E. coli* cells using the heat-shock method. 2µl of DNA was added to an aliquot of competent *E. coli* cells, and incubated on ice for 20 minutes. The cells were heat-shocked for 45 seconds in a waterbath set at 42°C, and then immediately incubated on ice for 2 minutes. 900µl of LB was added to the cells and placed in a shaker for 1 hour at 37°C, 170rpm. The cells were spun down at 13,000rpm for 1 minute and the supernatant was removed by pipetting. Cells were resuspended in 100µl of distilled water and spread onto a LB ampicillin agar plate. The plates were transferred to a 37°C incubator overnight.

2.2.8 Transformation of Plasmid into *Saccharomyces cerevisiae*

One *S. cerevisiae* (AH22 strain) colony was cultured overnight in 10ml of YEPD at 30°C and 150rpm. The *S. cerevisiae* culture was subcultured into 30ml of YEPD and grown until a cell count of $1x10^7$cells/ml was reached. 10ml of this culture was harvested at 3000xg for 3 minutes and the supernatant was discarded. The pellet was washed in 10ml of distilled water and centrifuged for a further 3 minutes. The wash was discarded and the pellet was resuspended in 10ml of 1M sorbitol. The suspension was centrifuged for a further 3 minutes. Again, the supernatant was discarded and the pellet was resuspended in 10ml of 1M sorbitol containing 20mM HEPES (Sigma-Aldrich). The suspension was centrifuged for another 3 minutes. The supernatant was again discarded and the pellet was resuspended in 50µl 1M sorbitol containing 20mM HEPES. The suspension was transferred to an ice-cold 1.5ml Eppendorf tube and placed on ice. To this suspension 5µl of plasmid DNA was added. The cells were transferred to an ice-cold electroporation cuvette. The cells were electroporated using the following conditions: 1.5kV, 25µF and 200Ω. Immediately the cells were resuspended in 1ml of cold YEPD and incubated at room temperature for 1 hour. The cells were then washed twice in 1ml of distilled water and resuspended in 500µl of distilled water. The suspension was spread onto a YM agar plate containing 5µg/ml histidine (Acros Organics).

2.2.9 Plasmid Sequencing

Plasmid DNA was sent for sequencing to Eurofins MWG Operon (London) at a concentration of 50-100ng/µl in a total volume of 15µl. Primers were sent with the plasmid for sequencing (see table 2.1) at a concentration of 2pmol/µl in a minimum volume of 15µl.

Gene Name	Plasmid Name	Primer Name	Primer Sequence
Alk1	pCWori$^+$:Alk1	>Alk1_Internal1_F	5'-ATCAAATTAAACAAA-3'
		>Alk1_Internal1_R	3'-TTTGTTTAATTTGAT-5'
		>Alk1_Internal2_F	5'-GCCAATTTTGGTGTT-3'
		>Alk1_Internal2_R	3'-AACACCAAAATTGGC-5'
Alk1L456S and Truncations	pCWori$^+$:Alk1, pCWori$^+$:Δ33:Alk1, pCWori$^+$:Δ69, pCWori$^+$:modifiedΔ50:Alk1 & pCWori$^+$:Δ44:Alk1	>STNL_Internal1_F	5'-AAGGTTGGAAGCATT-3'
		>STNL_Internal1_R	3'-AATGCTTCCAACCTT-5'
		>STNL_Internal2_F	5'-TCTGCTGAGCTTCGC-3'
		>STNL_Internal2_R	3'-GCGAAGCTCAGCAGA-5'
Δ65:Alk1	pCWori$^+$:Δ65:Alk1	>Alk1_Sol_OS_F	5'-GCGGCCGCCATATGGCGATTGATGCAATT-3'
CYP52A3	pCWori$^+$:CYP52A3	>CYP52A3_Internal1_F	5'-ACTGAATTTTTGTTT-3'
		>CYP52A3_Internal1_R	3'-AAACAAAAATTCAGT-5'
		>CYP52A3_Internal2_F	5'-TTGAATTGGCCAGAA-3'
		>CYP52A3_Internal2_R	3'-TTCTGGCCAATTCAA-5'
CPR	CPR	>CaCPR_F	5'-CGAATCATATGGCATTAGACAAATTA-3'
		>CaCPR_R	3'-ATTATTAAGCTTCCGGAAGCTTTTAGTGATGGTGATGGTGATGCCAAACATCTTCTTG-5'

Table 2.1 – Primers used for the sequencing of plasmid.

2.2.10 SDS-PAGE

3.6ml of the resolving gel was pipetted into the gel forming equipment, immediately followed by 1ml of water saturated with butanol, which was used to remove any bubbles that may have formed in the gel. The resolving gel was left for 90 minutes to set. Once set the butanol saturated water was removed and the gel was washed twice with distilled water. Any remaining traces of the water were removed before the stacking gel was then pipetted on top of the resolving gel and left to set for 30 minutes.

The protein samples were prepared by adding cracking buffer and incubated at 100°C in a heat block for 5 minutes. The cracking buffer was added at a ratio of 1:1 to purified protein and 5:1 with unpurified protein samples. The samples were then centrifuged for 2 minutes at 13,000rpm. 5µl of the sample was then loaded into one of the wells of the gel. 7µl of protein marker was also added. The gel was run at 150V for 70 minutes.

Following electrophoresis, the stacking gel was removed and the remaining part of the gel was stained for 10 minutes in Coomassie Blue stain. The gel was then destained overnight in Coomassie Blue destain.

2.2.11 Reconstitution Assays

The truncated *S. cerevisiae* NADPH-cytochrome P450 reductase (ScΔ33:CPR) protein used in these assays was a gift from Dr. Andrew Warrilow, Swansea University.

2.2.11.1 Micelle Formation

10µl of 2mg/ml of substrate (dissolved in chloroform) was added to 100µl of chloroform and 100µl of 1mg/ml dilaurylphosphatidylcholine (DLPC – dissolved in chloroform). The solution was thoroughly mixed prior to the addition of 200µl of distilled water. This solution was then thoroughly mixed producing a white emulsion. The chloroform was evaporated under a nitrogen stream until a clear solution of aqueous substrate-DLPC micelles was produced.

2.2.11.2 Reconstitution System

Added to 200µl of the aqueous micelles (section 2.2.11.1) was 50µl of 40mM glucose-6-phosphate, 50µl of 10U/ml glucose-6-phosphate dehydrogenase and 100µl 1M Tris-HCl (Melford), pH6.4. Various concentrations of purified cytochrome P450 protein and purified CPR protein were also added to the system (see chapters 5 and 6 for details). The final volume of this solution was made up to 900µl with the addition of distilled water. The solution was mixed thoroughly and pre-incubated at 37°C for 5 minutes. The reaction was initiated by the addition of 100µl of 10mM NADPH dissolved in 0.2% (w/v) sodium bicarbonate. The solution was mixed thoroughly before incubation at 37°C for 1 hour.

After incubation the reaction was either terminated on ice, extracted with 5ml of a chloroform:methanol (2:1, v/v) solution and evaporated to complete dryness in a SpeedVac or the reaction was subjected to saponification (section 2.2.11.3).

2.2.11.3 Saponification

The reaction was terminated by the addition of 1ml of 15% (w/v) potassium hydroxide dissolved in ethanol. The stopped reactions were mixed and then incubated at 85°C for 90 minutes. The samples were then extracted three times with 1ml of hexane. The extracted fractions were evaporated to complete dryness in a SpeedVac.

2.2.12 Silylation of Samples

To the dried sample (in a GC vial), 100µl of pyridine (Sigma-Aldrich) and one vial of *N,O*-bis(trimethylsilyl)trifluoroacetamide-trimethylchlorosilane (BSTFA-TCMS [90:10]) was added. The lid was crimped onto the GC vial and the sample was heated to 70°C on a hotblock for 1 hour.

2.2.13 GCMS

To determine the products of reconstitution assays using hydrocarbons and sterols as substrates, the extracted samples require analysis by GCMS (Gas Chromatography - Mass Spectrometry).

For separation by gas chromatography, the extracted sample is suspended in an organic solvent, such as methanol, which is injected into a mobile phase, in this case the inert gas, helium. The sample is carried to a column (known as the stationary phase) where it is separated and eluted. The products which interact with the column the quickest are eluted first, whereas those that take the longest time to interact, elute the slowest. This separation can be aided by ramping (increasing) the oven temperature (where the column resides) so that the products can be separated by boiling point. Therefore, products with lower boiling points are eluted from the column faster than those with higher ones. The eluted products then enter a detector producing a signal, which can be used to identify the point/time at which the product was eluted from the column. This is known as the retention time. Using a software program, such as MSD Enhanced ChemStation, a chromatograph can be produced plotting retention time against abundance (signal intensity). If the conditions used remain constant, then the retention time can be used to identify the product eluted as this will remain almost constant. However, different products may have similar or very close retention times, which require more analysis to correctly determine the product. This can be done using mass spectrometry.

Following elution from the GC column, the product can enter a mass spectrometer detector where it is fragmented by a cascade of electrons. These charged fragments are separated by electromagnets and are recorded as a graph, known as the mass spectrum. This mass spectrum shows the reproducible fragmentation pattern for a given product/chemical, allowing for accurate identification when used in conjunction with the retention time produced by the GC.

Figure 2.1 shows a schematic overview of the GCMS.

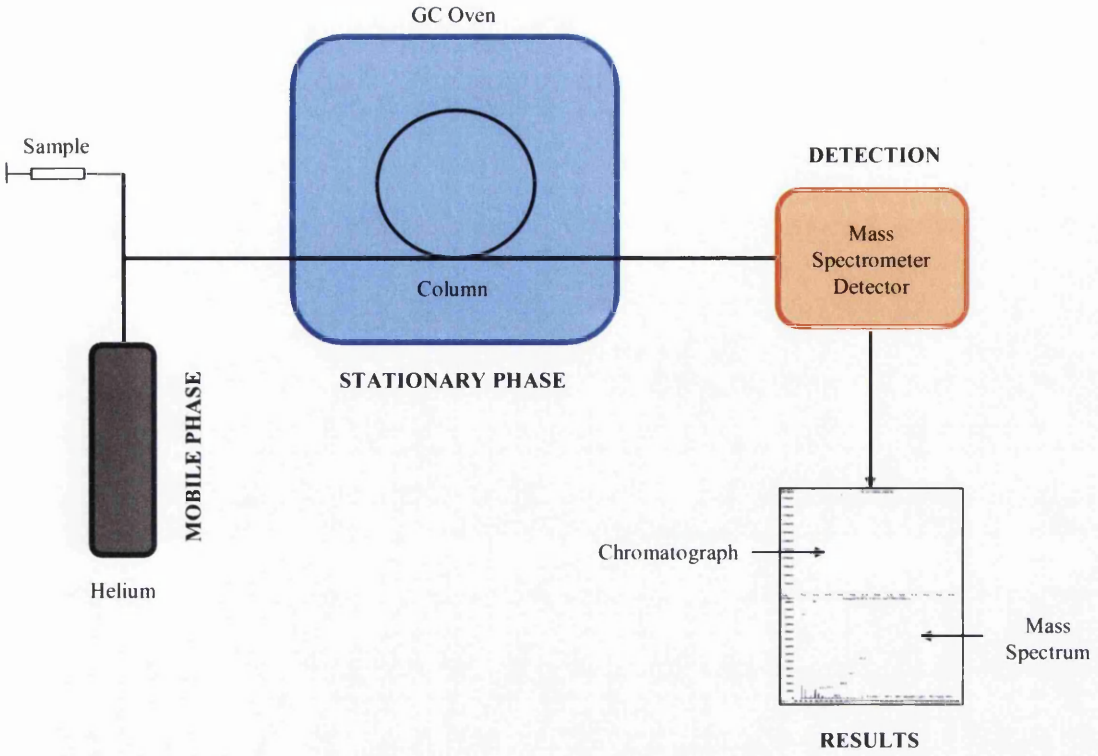


Figure 2.1 – Schematic overview of the GCMS.

Analysis of the products of reconstitution assays using hydrocarbons and sterols as substrates were performed using a 7890A GC system (Agilent Technologies), 5975C mass spectrometer (Agilent Technologies) and MSD Enhanced ChemStation software. A DB-5MS fused silica column with a 30m x 0.25mm x 0.25µm film thickness (J & W Scientific) was used to separate the products on the GC. Samples were injected in 1µl volumes in the splitless mode using helium as the carrier gas.

2.2.13.1 GCMS Method for the Analysis of Alkanes, Fatty Acids and their Derivatives

The GCMS method used was adapted from the methods of He *et al.*, 2005 and Basconcillo & McCarry, 2008.

The oven program used was as follows. The oven temperature was held at 70°C for 4 minutes and then ramped, 20°C/min, to a final temperature of 300°C. The oven was held at this temperature for 5 minutes and a solvent delay of 5 minutes was used.

2.2.13.2 GCMS Method for the Analysis of Sterols

The GC oven program was initially set at 70°C for 4 minutes. The temperature was then ramped, 25°C/min, to a final temperature of 280°C. This temperature was held for 25 minutes.

Chapter 3: Identification of Cytochrome P450 Genes in *Candida albicans* Using a Bioinformatic Approach

3.1 Introduction

Bioinformatics is an interdisciplinary field combining biological sciences with mathematics (including statistics) and computer sciences (Golding, 2003). It allows biological data, such as sequence data, to be collated and accessed, as well as enabling the user to search through vast collections of data, using bioinformatic tools (discussed in detail in section 3.1.5). The user can look for sequence data (nucleotide or protein) that is a homologue, orthologue or paralogue of another gene/protein. A homologue is a DNA sequence or protein molecule that has the same function as another and is related by a common ancestor (Westhead *et al.*, 2002). Paralogues and orthologues are different types of homologues. An orthologue is a homologue, that has arisen after speciation (Westhead *et al.*, 2002), for example CYP51. CYP51 is a member of the cytochrome P450 family and is involved in sterol production. It can be found in a plethora of species including yeast, mycobacteria and humans (Kelly *et al.*, 2001). A paralogue however, is a homologue with a different function, which has arisen after gene duplication, e.g. trypsin and haptoglobin. Trypsin is an enzyme involved in the breakdown and digestion of proteins and shares a common ancestor with haptoglobin. However, haptoglobin has a completely different function to trypsin and is involved in the prevention of unrequired protein aggregation by acting as a chaperone (Lesk, 2005). Enzymes belonging to the CYP52A subfamily of cytochromes P450 are examples of paralogues. Proteins in this subfamily are only found in yeast and share homology, but they preferentially hydroxylase different hydrocarbon molecules. For example, CYP52A3 can catalyse the hydroxylation of both alkanes and fatty acids to their respective alcohols and diacids (Zimmer *et al.*, 1996; Scheller *et al.*, 1998) whereas CYP52A4 can only hydroxylate fatty acids (Zimmer *et al.*, 1995; Zimmer *et al.*, 1996). Bioinformatics is an excellent tool for discovering new genes, including those that make up the cytochrome P450 family, as it can be used to identify putative orthologous genes.

The objective of this chapter is to identify all possible cytochrome P450 genes in *Candida albicans* and to find which of these genes has the greatest homology to the alkane-hydroxylating gene from *C. maltosa*, *CYP52A3* (Ohkuma *et al.*, 1991).

CYP52A3 is of interest as *CYP52* genes and proteins of yeasts and fungi are associated with the terminal hydroxylation of alkanes where the alcohols produced are further converted to fatty acids and then subjected to β-oxidation. Alternatively, fatty acids can be converted to diacids, which can then either be subjected to β-oxidation or be extracted to form the chemical basis of many industrial products, such as perfumes, plastics, lubricants, hot-melting adhesives and antibiotics (Eschenfeldt, *et al.*, 2003; Liu, *et al.*, 2003).

3.1.1 Genome Databases

In 1976, Walter Fiers *et al.* from the University of Ghent, Belgium published the first complete RNA genome sequence from bacteriophage MS2 (Fiers *et al.*, 1976). This was quickly followed by the publication of the first complete DNA genome from phage φ-X174 in 1977 by Fred Sanger and his team (Sanger *et al.*, 1977). Since then there has been an explosion in the number of genome sequences published. These have included *Haemophilus influenza* (the first genome to be published from a living organism, (Fleischmann *et al.*, 1995)), *Escherichia coli* (Blattner *et al.*, 1997), *Genlisea margaretae* (Greilhuber *et al.*, 2006), *Sacchromyces cerevisiae* (http://yeastgenome.org/) and *Homo sapiens* (draft version in 2001 (International Human Genome Sequencing Consortium, 2001); completed in 2003 (Collins *et al.*, 2003)). It is important to note this is not an exhaustive list and many other genome projects have been completed or are currently being undertaken.

Large-scale genomic sequencing, such as those mentioned above, can be undertaken using robot-fed capillary sequencers, which produces sequence data in the form of chromatographs (Golding, 2003). Alternatively, next-generation sequencing (NGS) technologies can be used to produce sequence data. NGS permits an entire genome to be sequenced by breaking it down into fragments, which can be read and assembled to produce a whole genome sequence (Zhang *et al.*, 2011). This technology is rapidly developing and utilisation of high-throughput techniques (such as microarray analysis) has led to advancements in the understanding and identification of genes involved in human disease as multiple human genomes can now be sequenced (Zhang *et al.*, 2011).

Genome sequences that are nearly or completely sequenced can be found in their own specific database as well as in nucleotide databases (which are described in section 3.1.2). Genomic databases contain genetic and molecular biological information for a particular species. They contain DNA sequences, protein sequences (to a lesser extent), descriptions and classifications of biological roles, molecular functions, subcellular localisations, links to literature, links to functional genomic datasets (e.g. expression arrays) and bioinformatic tools for the analysis and comparison of sequences.

An example of such a genome database is the *Candida* Genome Database (CGD, www.candidagenome.org) (see figure 3.1). As the name implies, this archive provides data on the *Candida* genome sequence, particularly *C. albicans*. It provides the user with helpful resources, such as those mentioned above, as well as information on biochemical pathways and comparisons to the *S. cerevisiae* genome. The database also provides researchers with the latest news from the *Candida* research community. The CGD is funded by the National Institute of Dental and Craniofacial Research, US National Institutes of Health and is based on the *Saccharomyces* Genome Database (SGD, http://yeastgenome.org/), which was set-up to provide researchers within the yeast community fast and easy access to information relating to the *S. cerevisiae* genome. This collaboration has enabled CGD to offer improved bioinformatic tools, such as advanced search, batch download, pattern matching and primer design (www.candidagenome.org).

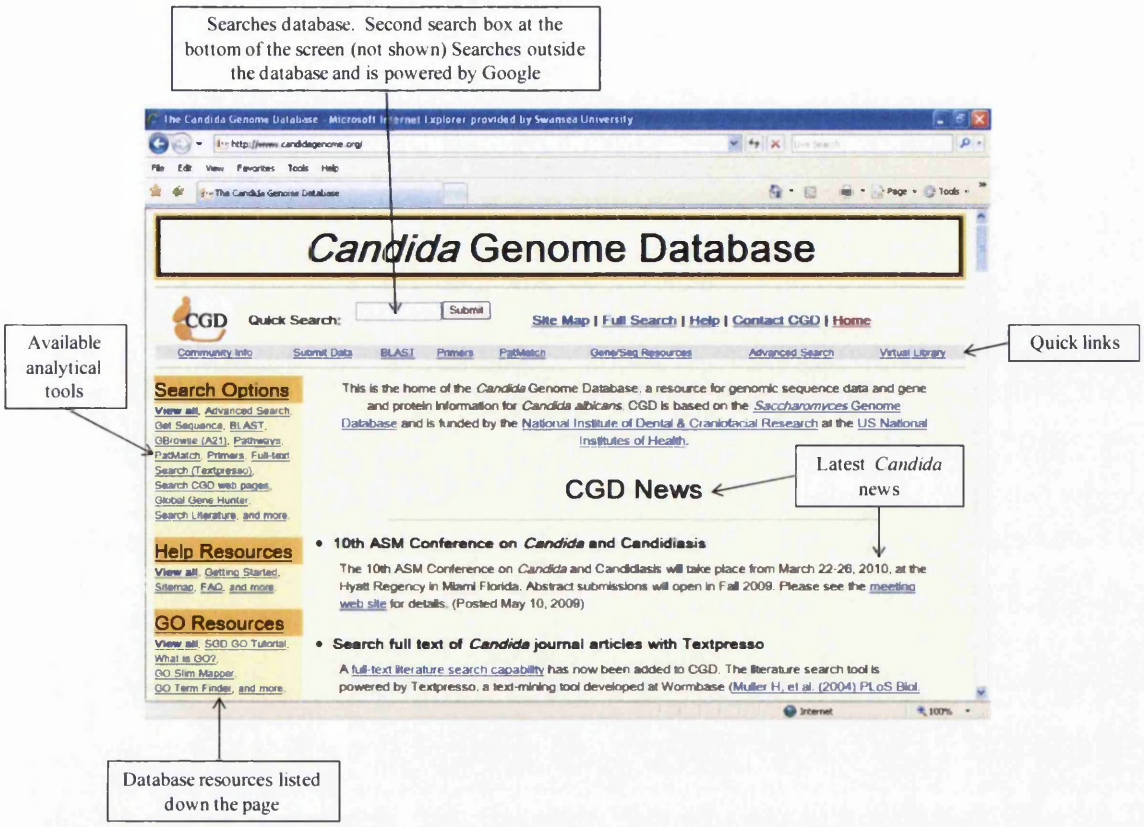


Figure 3.1 – View of the *Candida* Genome Database (CGD) homepage.

3.1.2 Nucleotide Databases

Nucleotide databases are repositories for nucleic acid sequences (and to less extent translated protein sequences). They have become a more popular medium in recent years as fewer and fewer journals are publishing new nucleotide sequence data (Ashburner & Goodman, 1997). There are three primary databases, which store raw nucleic acid sequence data: GenBank (NCBI, www.ncbi.nlm.nih.gov/Genbank), the Nucleotide Sequence Database (EMBL, www.ebi.ac.uk/embl) and the DNA Databank of Japan (DDBJ, www.ddbj.nig.ac.jp). Collectively they are known as "The International Nucleic Acid Sequence Data Library." The data in these archives come from genome projects, scientific publications and patent applications (Lesk, 2005). This information is shared between the three databases on a daily basis, meaning that all three archives contain a lot of the same raw data (Lesk, 2005). However, this information is said to have a lifespan due to the fact most entries are added before they have been fully annotated and checked by library staff (Ashburner & Goodman, 1997). The quality of the information submitted is not the

responsibility of the database but of the researcher/research group who submitted the original sequence (Ashburner and Goodman, 1997). Therefore, once reviewed, many of the sequences may be discarded due to a number of reasons, including sequence errors, mislabelling, contamination and sequence duplication. These are big problems as they cause a needless increase in the size of the archive making the database redundant (Ashburner & Goodman, 1997). Those sequences that remain are extensively annotated and the important properties of each sequence are highlighted, e.g. name and origin, protein attributes, sequence similarities, ontologies and literature reference(s).

Figure 3.2 shows a screen shot from the Nucleotide Sequence Database (EMBL, www.ebi.ac.uk/embl). The database was searched for sequences relating to the alkane-hydroxylating cytochrome P450 from *C. maltosa*, CYP52A3. It was able to identify 2 nucleotide sequences, 1 protein sequence and 5 related articles. This information is sufficient to be used, for instance, to produce primers or to be used in a BLAST search (discussed in more detail in section 3.1.5). Other genes searched will yield more or less information than *CYP52A3* depending on the level of annotation and information available (from the database).

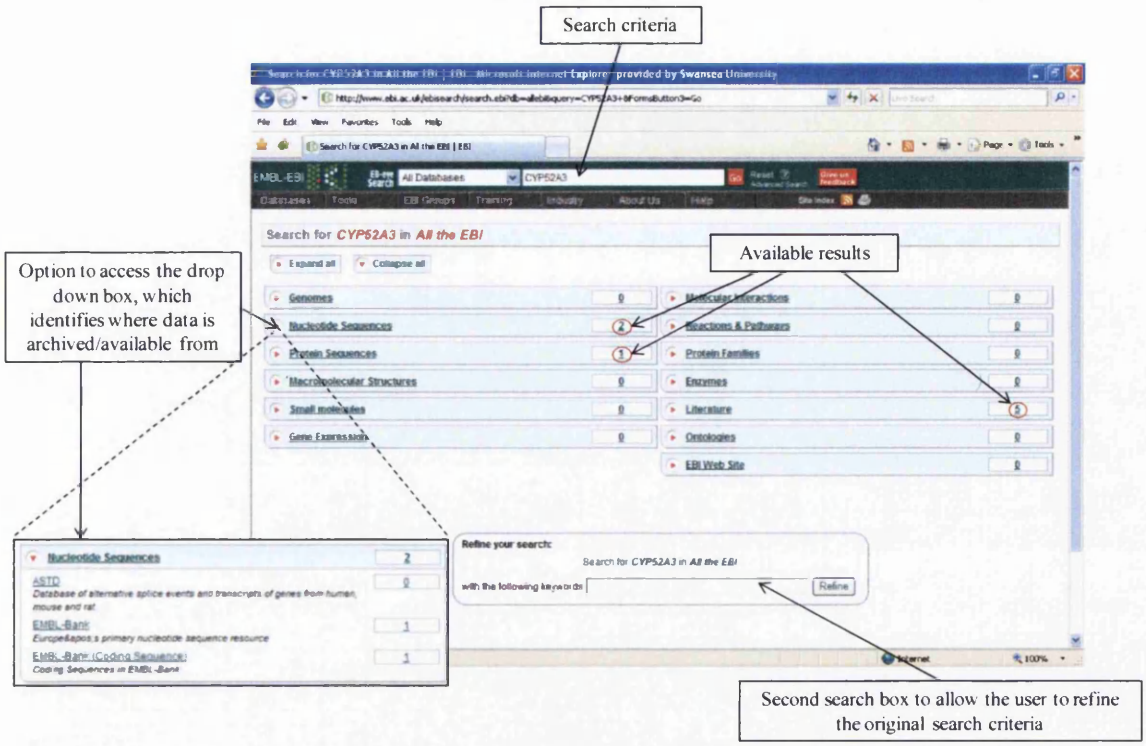


Figure 3.2 – Entry page from the nucleotide sequence database (www.ebi.ac.uk). CYP52A3 was entered into the search box at the top of the homepage (not shown) and Go was pressed generating the page shown in this figure. The user is able to identify the gene, the DNA sequence (as well as the protein-coding sequence), information about the protein and any relevant literature.

3.1.3 Primary Protein Databases

Like nucleotide sequences, raw protein sequences are initially collated into databases, such as SWISS-PROT and TrEMBL. SWISS-PROT is the result of a collaboration between the Swiss Institute of Bioinformatics (http://www.isb-sib.ch/) and the European Bioinformatics Institute (http://www.ebi.ac.uk), and contains protein sequences, from published scientific articles, that are actively monitored and annotated manually by curators (Apweiler *et al.*, 2004; Lesk, 2005). Unlike nucleotide sequence databases, primary protein sequence databases are non-redundant (Apweiler *et al.*, 2004). As mentioned previously, nucloetide databases can contain many of the same or similar sequences (Ashburner & Goodman, 1997), however, SWISS-PROT contains very little repeated data ensuring the database is not unnecessarily large, thus reducing the time the user has to spend searching for a particular sequence. Data in the SWISS-PROT database is extensively annotated and the biological information used includes protein function, post-translational

modifications, the identification of sites and domains, protein similarities and biological pathways involving the protein (Apweiler *et al.,* 2004).

TrEMBL, on the other hand, contains sequences derived from translated protein sequences found in the Nucleotide Sequence Database (EMBL, www.ebi.ac.uk/embl), which are not already integrated into SWISS-PROT (Lesk, 2005). By using translated protein sequences the rate at which new protein sequences are made available is constantly increasing (Ashburner & Goodman, 1997; Apweiler *et al.,* 2004). However, TrEMBL is not just made up of translated protein sequences as it also contains some sequences from published journals and experimental data (Apweiler *et al.,* 2004).

Due to the redundancy associated with primary nucleotide databases, many of the sequences in the TrEMBL database may have to be removed or merged to produce a single entry. The annotations used for many of these sequences are from the nucleotide database (Apweiler *et al.*, 2004). However, the remaining unannotated sequences require annotation. Information for these seqeunces comes from InterPro (an integrated protein database) (Apweiler *et al.*, 2001, www.ebi.ac.uk/interpro/), which is made up of other databases such as PROSITE, PRINTS and Pfam (descibed in more detail in section 3.1.4) (Apweiler *et al.,* 2004). Despite this the sequences in TrEMBL are not as extensively annotated as those in SWISS-PROT (Ashburner & Goodman, 1997). However, these sequences, once completed, are eventually moved to the SWISS-PROT database where they undergo more extensive annotation (Lesk, 2005).

It has been established that SWISS-PROT and TrEMBL are the main repositries of primary protein sequence data, and this recognition led to the formation of a single database in 2003 named UniProt (www.uniprot.org) (Apweiler *et al.,* 2004b, Lesk, 2005). This database not only combines biological data from SWISS-PROT and TrEMBL, but also from a third database, PIR (Protein Information Resource, http://pir.georgetown.edu/) (Apweiler *et al.,* 2004). The PIR database provides UniProt with sequences found in neither the SWISS-PROT nor TrEMBL databases (Apweiler *et al.,* 2004). Therefore, the UniProt database provides the same data as the other primary protein databases, but is split into two parts: those sequences that

are fully annotated (continuing to be called SWISS-PROT) and those sequences waiting to be annotated (referred to as TrEMBL) (Apweiler *et al.,* 2004).

3.1.4 Secondary Protein Databases

In contrast to nucleotide databases, there is a second generation of protein databases available. The secondary protein sequence databases allow an unknown protein to be determined (within the archive) by noting the existence of a particular pattern, motif, signature or fingerprint within the sequence. For instance, cytochromes P450 contain a conserved sequence relating to the haem-binding region, FxxGxxxCxG (Nelson *et al.,* 1993; Nelson *et al.*, 1996). Using this conserved sequence to search against sequences from the database related protein families can be identified (Lesk, 2005). Examples of secondary protein databases are PROSITE (Hofmann *et al.*, 1999, www.expasy.ch/prosite/), PRINTS (Attwood *et al.*, 2000, http://www.bioinf.man.ac.uk/ dbbrowser/PRINTS/) and Pfam (Bateman *et al.*, 2000, http://pfam.sanger.ac.uk/). Each of these databases use a different methodology to determine an unknown protein and these are described below.

PROSITE (Hofmann *et al.*, 1999) is the oldest sequence-motif database (Kriventseva *et al.,* 2001) and it identifies proteins by sequence domains or motifs (see figure 3.3) (Ashburner & Goodman, 1997). The motifs come from published articles and are linked to annotations in the SWISS-PROT database (Liu & Rost, 2003), which means sites and patterns that are biologically significant can be identified so the protein's family can be determined rapidly. However, the identification of new members from diverse families is less reliable because if the information/motif required is not matched in SWISS-PROT then PROSITE is unlikely to be able to match that motif to unidentified proteins (Liu & Rost, 2003).

An example is the FxxGxxxCxG motif associated with cytochrome P450 proteins. This motif contains areas between the conserved residues that can be occupied by any one of the twenty-two amino acids available. PROSITE identifies the defined/highly conserved residues and ignores the amino acids in between by using position-dependent profiles or weight matrices (Bork & Koonin, 1996; Kriventseva *et al.*, 2001). These matrices define the residues of interest so that any insertions or positions that are not part of the motif can be tolerated (Kriventseva *et al.*, 2001).

Figure 3.3 shows an example of a results page from PROSITE. It has been generated using the protein sequence of Alk1 (accession number: Q5A8M1) from *C. albicans* (for more details on this protein see section 3.3). PROSITE identifies this protein as being a member of the cytochrome P450 group and gives a short description of the recognized motif ("Cytochrome P450 cysteine heme-iron ligand signature"). The motif is highlighted on a linear graphic of the protein sequence and there is a legend accompanying these graphic identifying generalised sites. In this case the motif is classified as "other sites". Finally, the region the motif is situated in within the protein sequence is also identified. In this case, it is found between residues 467 and 476 and is identified as FNgGPRICLG.

A second secondary database is PRINTS (Attwood *et al.*, 2000). PRINTS like PROSITE is made up of motifs extracted from searches undertaken in SWISS-PROT (Kriventseva *et al.*, 2001). However, unlike PROSITE, PRINTS uses a technique called "fingerprinting" (Lesk, 2005). Fingerprints are groups of motifs that indicate specific protein families (Attwood *et al.*, 2003). This means rather than searching its archive for one particular sequence motif like PROSITE, PRINTS can search for several by examining multiple alignments (Liu & Rost, 2003). This method is more flexible than the single-motif methods and it has the ability to identify a greater number of relative proteins (Attwood *et al.*, 2003).

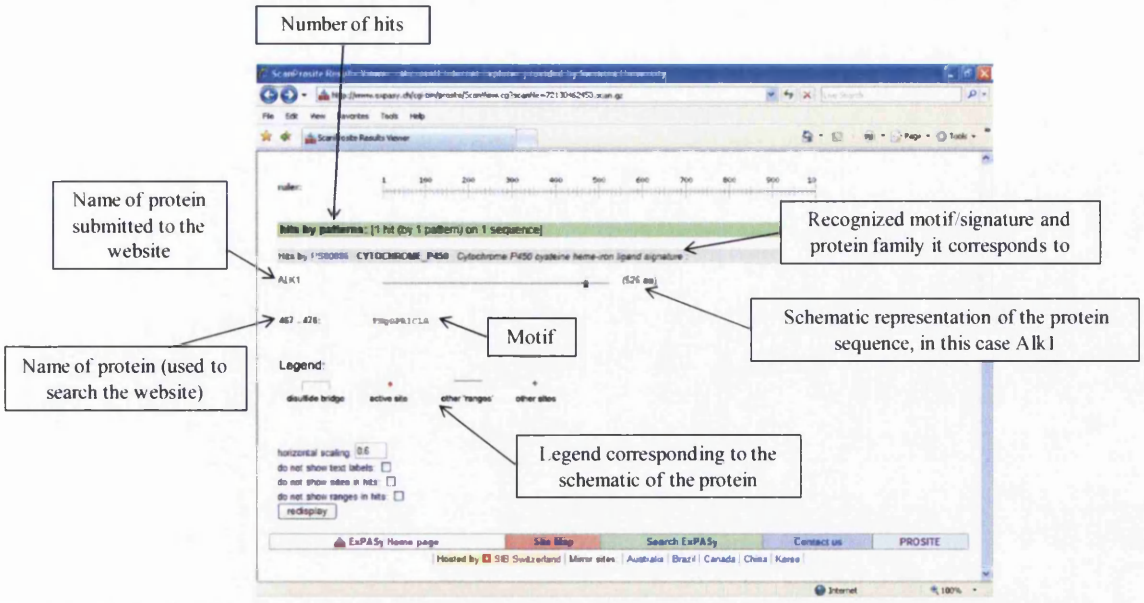


Figure 3.3 – Results page from PROSITE (Hoffman *et al.*, 1999). Results were obtained by searching the database with the protein sequence for Alk1 (in FASTA format). These results show that Alk1 contains a motif associated with cytochromes P450.

Thirdly, Pfam (Bateman *et al.*, 2000) (see figure 3.4) searches for conserved domains in the SWISS-PROT database (Kriventseva *et al.*, 2001) and uses Hidden Markov Models (HMMs) to identify these conserved regions in other protein sequences (Bateman *et al.*, 2002). HMMs are sophisticated systems that take their name from the fact that although the sequences are shown to the user the path through the models (or states) used to identify them are not seen. The HMMs depict the protein domain family by producing sequences from that particular family with a very high probability along with other sequences that produce a less variable probability. This method makes identification of new family members more reliable (Lesk, 2005).

Figure 3.4 shows a results page from Pfam. It was produced using the Alk1 protein sequence. A generated graphic shows the motif identified within the sequence, in this case the motif is cytochrome P450. By clicking on this domain graphic the user is redirected to a second webpage (not shown), which provides information on the protein family the motif belongs to, related references, gene ontology and links to external databases including other secondary protein databases, such as PROSITE and PRINTS. Immediately below there is a table noting the domain boundaries of the identified motif. For Alk1, this is 316 to 525. This is a far greater length of

sequence than was quoted by PROSITE, which identified the motif for cytochrome P450 as being between residues 467 and 476 (see above). A reason for this is that some of the domains identified by Pfam may be obscuring/overlapping other domains giving the entry sequence a larger domain boundary. The alignment of the motif to the protein entry may be observed by pressing the "Show" button (not shown). Finally, the Pfam results also indicate motifs that are deemed less important than the other motifs it deems "significant" (in this case a Nup133 N terminal like motif and a second cytochrome P450 domain). Significance is determined by the bits score and E-value, which are related to sequence homology. The greater the score and the lower the E-value, the higher the homology is (this is described in greater detail in section 3.1.5). Therefore a significant Pfam match is classified as having a very high bits score and a very low E-value.

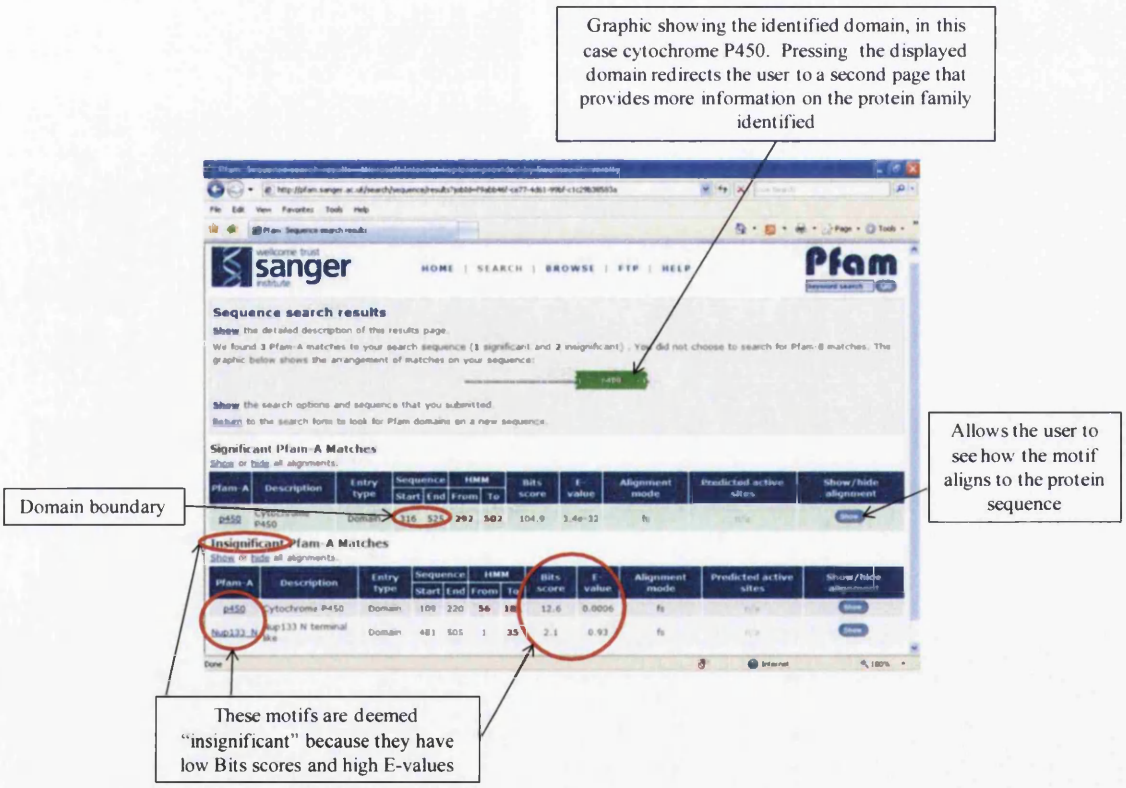


Figure 3.4 - Results page from Pfam (Bateman *et al.*, 2000). Results were obtained by searching the database with the protein sequence for Alk1 (in FASTA format). These results show that Alk1 contains a domain associated with cytochromes P450 and where it is located within the sequence.

Due to these differing approaches, each database has a particular area of expertise. For instance, PROSITE focuses much of its attention on functional sites, whereas Pfam is used primarily to search for divergent domains and PRINTS can identify sub-families. These are just a small subsection of the secondary databases (and methods) available but they convey the notion that there is no one ideal method/database to use. To combat this issue an integrated archive, entitled the InterPro database (Apweiler *et al.*, 2001, www.ebi.ac.uk/interpro/), was formed. InterPro initially allowed quick and easy access to information in PROSITE, PRINTS, Pfam and ProDom (Apweiler *et al.*, 2001, Kriventseva *et al.*, 2001, Lesk. 2005). (ProDom (Corpet *et al.*, 2000, http://www.toulouse.inra.fr/prodom.html) is a sequence cluster database, that unlike the aforementioned databases is produced using clustering algorithms directly from sequence databases (Apweiler *et* al., 2001, Kriventseva *et al.*, 2001)). However, since 2001 other protein databases have been incorporated into the InterPro database. These include: PIRSF (provides information on protein sequence hierarchies) (Nikolskaya *et al.*, 2007, http://pir.georgetown.edu/pirwww/dbinfo/pirsf.shtml), PANTHER (proteins are classified by function) (Mi *et al.*, 2007, http://www.pantherdb.org/), SMART (involved in domain architecture analysis) (Letunic *et al.*, 2006, http://smart.embl-heidelberg.de/), TIGRFAMs (protein family database) (Haft *et al.*, 2003, http://www.jsvi.org/cms/research/projects/tigrfams/overview/), Gene3D (produces theoretical 3D images of protein domains) (Yeats *et al.*, 2008, http://gene3d.biochem.ucl.ac.uk/Gene3D/) and SUPERFAMILY (archives protein functions and structures for completely sequenced genomes) (Wilson *et al.*, 2007, http://supfam.cs.bris.ac.uk/SUPERFAMILY/) (Hunter *et al.*, 2009).

InterPro entries are annotated with information correlating to function, related published articles and links to the original secondary protein database from which they were generated (e.g. PROSITE, PRINTS, Pfam or ProDom) (Apweiler *et al.*, 2001). The future aim of this database is to make it truly integrated by linking information from one database to another enabling the user to identify how motifs from different databases correlate to one another (Hunter *et al.*, 2009).

The different types of databases mentioned above are excellent collections of nucleotide and protein sequences. As described, they provide the user with quick

and easy access to a constantly increasing volume of biological data. However, the data needs to be analysed and interpreted to obtain scientific answers of use. This can be done using a growing number of bioinformatic tools that have become available in recent years, some of which are mentioned below.

3.1.5 Bioinformatics Tools

Bioinformatic tools can be used, for instance, to find comparisons between new sequences and characterised sequences from the aforementioned databases, to determine the degree of homology between sequences and to evaluate evolutionary relationships. One important bioinformatic tool is BLAST (or Basic Local Alignment Sequence Tool) (Altschul *et al.*, 1997). BLAST compares a given (nucleotide or protein) sequence against every entry in a database to find (other) sequences with varying degrees of homology to the original search sequence. There are many different BLAST programs and most use amino acid sequences as the query sequence against databases containing protein or translated nucleotide sequences. Protein sequences are preferred to nucleotide ones as the query sequence because they allow homology (of sequences and characteristics through evolution) to be assessed (Rehm, 2001). Table 3.1 shows a sample of various BLAST programs available and the types of query and database sequences they use. The BLAST program checks each query sequence for short adjoining regions found in each sequence within the database (Lesk, 2005). A substitution scoring matrix (BLOSUM62) is used, but no gaps are permitted (Lesk, 2005). This means that a list of potential sequences can be produced quicker than it would take to generate the results manually. Each sequence in this list is given a bit score and an E-value. The bit score is given relating to the sequence similarity, so the greater the score the higher the homology (Lesk, 2005). A bit score is better than a raw score, which is the same as a bit score, except a bit score also takes the statistical parameters of the scoring system into consideration (Karlin & Altschul, 1990; Altschul, 1991; Altschul & Gish, 1996; Rehm, 2001). The E-value describes the probability of homology between the query sequence and the BLAST results, the greater the homology the lower the E-value (Rehm, 2001; Lesk, 2005). For instance, if the E-value is less than 0.02 then the sequences are probably homologous, but if the E-value is greater than 1 then the probability of the sequence being homologous is very slim (Lesk, 2005). The E-value is calculated as follows:

E-value = P-value x Database size

Where the P-value is the probability of the alignment being better than random (Lesk, 2005). The P-value does not depend on the size of the database, unlike the E-value. P-values lower than 10^{-100} are deemed an exact match, with anything below 10^{-1} being classified as probably insignificant (Lesk, 2005).

This bioinformatic tool can be used to find sequences with homology to, for instance, CYP52A3. The results are ranked according to homology so the sequence with the highest bit score and lowest E-value is at the top of the page. Usually, this is the query sequence itself. The user can identify sequences of interest by choosing sequences in the list depending on the species of interest (e.g. *C. albicans*), the family of interest, the bit score and the E-value. An example of a BLAST results page is shown in figure 3.7, which is in the results section (3.3) of this chapter.

Program Name	Query Sequence	Database Type
PSI-BLAST	Amino Acid	Protein
BLASTN	Nucleic Acid	Nucleic Acid
BLASTX	Translated Nucleotide	Protein
TBLASTN	Amino Acid	Translated Nucleotide

Table 3.1 – Example of some of the different BLAST programs available. Each program uses a different type of query sequence to search the different types of database.

Another bioinformatic tool involves sequence alignment. Using the same basic principle as BLAST (but on a smaller scale), pairwise alignments can be used to match identical amino or nucleic acids (depending on the sequence used) in pairs or in multiple arrangements. Usually, a program such as ClustalX (Thompson *et al.*, 1997) is used. This utility not only allows sequences to be matched, but also places the sequences with the greatest homology at the top of the list during multiple alignments. Alignments from ClustalX can then be viewed as a .aln file in a separate program, such as BioEdit (Hall, 1999), where identical units of the sequences can be shaded the same colour to allow conserved regions of interest to be quickly identified. Following alignment in ClustalX (Thompson *et al.*, 1997) a .dnd file is

also produced, which allows the alignment to be viewed as a phylogenetic tree in a separate program, such as TreeView (Hall, 2001).

A phylogenetic tree shows graphically the ancestral relationships of different sequences. Homology can be inferred using this method because sequences with much similarity will share a common ancestor (Westhead *et al.*, 2002). A phylogenetic tree is made up of branches and nodes (external and internal). External nodes are the tips of the tree (e.g. the name of the sequence/gene/organism), whereas the internal nodes are the points where two (or more) nodes intersect (Hall, 2001). A branch connects two (or more) nodes together (Hall, 2001). The distance between two nodes can be calculated by counting the number of amino or nucleic acid changes that have occurred along that branch (Hall, 2001). This means the greater the homology between the two sequences, the easier it is to convert one of the sequences into the other as fewer mutations are required (Westhead *et al.*, 2002). Phylogenetic trees that relate branch length to distance usually have a legend corresponding to this. An example of this is shown in figure 3.10 (which can be found in the results section 3.3).

Such databases and bioinformatic tools can be used to discover new cytochrome P450 genes. These methods will be used to quantify how many cytochrome P450 genes there are in *C. albicans* and which of these has the greatest homology to *CYP52A3*, a known alkane-hydroxylating P450 gene from *C. maltosa* (Ohkuma *et al.*, 1991).

3.2 Methods

3.2.1 The Cytochromes P450 of *C. albicans*

The ExPASy website (www.expasy.org) was searched for potential cytochromes P450 in *C. albicans* using the query "P450 candida albicans." The fungal cytochrome P450 database (http://p450.riceblast.snu.ac.kr) was also searched.

The nomenclature for cytochromes P450 is based upon amino acid identity. 40% identity and above place a CYP in the same family and more than 55% identity places them in the same subfamily (Nebert, *et al.*, 1987; Nelson, *et al.*, 1996). Families are designated a CYP number based on those reserved for different

taxonomic groups (http://drnelson.utmem.edu/CytochromeP450.html), for lower eukaryotes (including fungi) these are CYP51-CYP69, CYP501-CYP699 and CYP5001-CYP6999. Sequences were collated and scoured for the presence of consensus amino acid sequences associated with cytochromes P450. These conserved regions include the PER(W) domain, the haem binding domain and ExxR of the K-helix. The haem-binding domain is characterised as FxxGxxx**C**xG, where the cysteine (highlighted) acts as the fifth ligand of the haem iron (Werck-Reichhart & Feyereisen, 2000). The ExxR motif is found in the K-helix, which is known to provide core structure stability (Werck-Reichhart & Feyereisen, 2000).

The search data was refined by undertaking a TBLASTN (Gish, 1996-2000) search using the protein sequence of each potential cytochrome P450 in the *Candida* Genome Database (see figure 3.5). Sequences not found within the *C. albicans* genome were easily identified and discarded, along with any duplicated sequences. The remaining duplicates were identified by aligning the remaining sequences in ClustalX (Thompson *et al.*, 1997) and viewed in BioEdit (Hall, 1999).

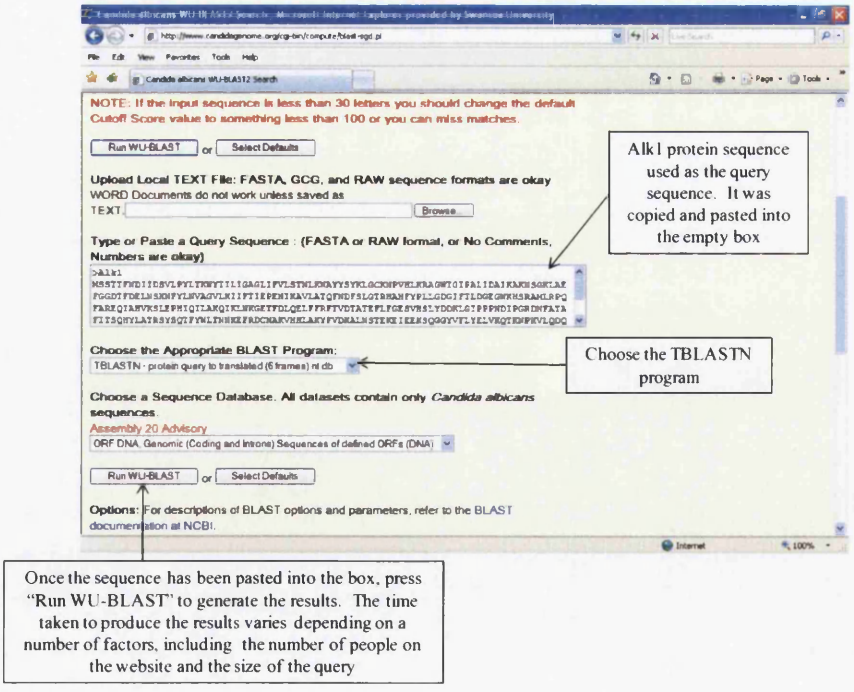


Figure 3.5 – TBLASTN page of the *Candida* Genome Database. The TBLASTN was undertaken using the protein sequence of Alk1 (in FASTA format).

3.2.2 Homologue of CYP52A3

A BLAST search was undertaken using the CYP52A3 protein sequence (accession number: P16496) (See figure 3.6). As a result, a list of homologous CYP52A3 proteins was produced and the *C. albicans* genes with the highest homology were noted. Search results are given a score relating to the degree of sequence similarity, and an E-value pertaining to the probability of the sequences being a match, therefore, the higher the score the greater the degree of similarity, and the lower the E-value the better the match. This means a gene with high/great homology would have as a result a high score and a low E-value.

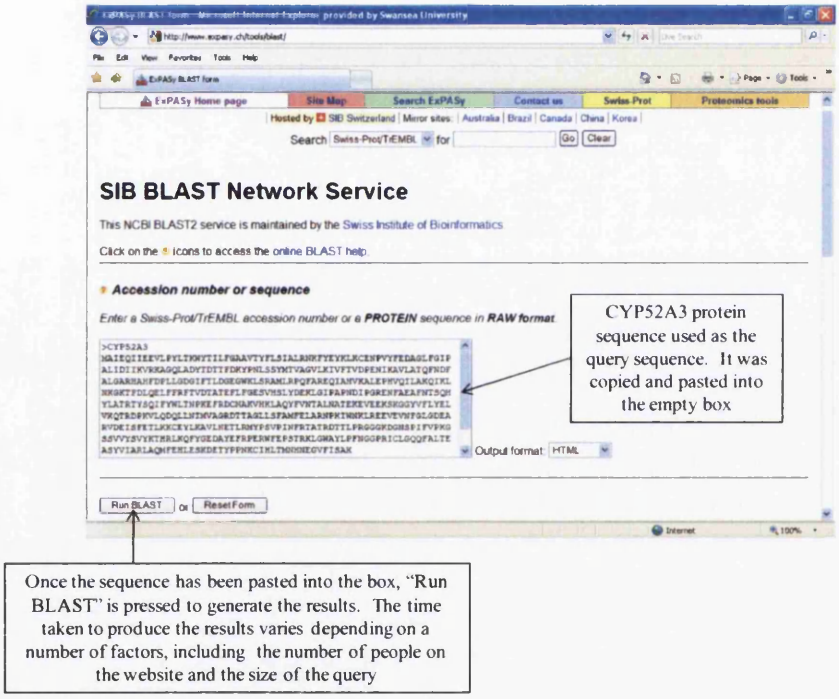


Figure 3.6 – BLAST (Altschul *et al.*, 1997) page of ExPASy (www.expasy.ch). The BLAST search was undertaken using the *C. maltosa* cytochrome P450, CYP52A3 (accession number: P16496).

To ensure the *C. albicans* genes identified were full cytochrome P450 sequences and not partial or duplicated ones, the accession numbers of each gene was compared with those cytochromes P450 known to be in the *C. albicans* genome (see section 3.3.1 for a comprehensive list). Any partial or duplicated sequences were then removed from the list. Finally, the (remaining) sequences were fully aligned with

CYP52A3 using ClustalX (Thompson *et al.*, 1997) and viewed in BioEdit (Hall, 1999). Phylogenetic trees were produced in TreeView (Page, 2001).

3.3 Results

3.3.1 The Cytochromes P450 of *C. albicans*

There are 10 putative cytochrome P450 genes/proteins in *C. albicans*, of which five are potential alkane and/or fatty acid hydroxylating genes or *CYP52*s. The proteins encoded by these genes (*ALK1, ALK2, ALK3, ALK6* and *ALK8*) have been highlighted in table 3.2.

Name	Other Name	Accession No.	ORF No.	Species	K-Helix	PER(W) Domain	Haem-Binding Domain	Protein Length (AA)	Molecular Weight (Da)
CYP52A3		P16496	-	*C. maltosa*	ETLR	PERW	FNGGPRICLG	523	59,840
Alk1		Q5A8M1	ORF19.5728	*C. albicans*	ETLR	PDRW	FNGGPRICLG	526	59,738
Alk2		Q5AAH6	ORF19.7513	*C. albicans*	ECLR	PERW	FNGGPRICLG	522	59,820
Alk3		Q5AAH7	ORF19.7512	*C. albicans*	ESLR	PERW	FNGGPRICLG	538	61,849
Alk6		Q5AGW4	ORF19.6574	*C. albicans*	ESLR	PERW	FCTGPRLCLG	507	58,293
Alk8	CYP52A21	Q59K96	ORF19.10	*C. albicans*	ETLR	PERW	FNGGPRICLG	515	59,332
Dit2		Q5A6K0	ORF19.55	*C. albicans*	ETLR	PERW	FHGRKRACLG	387	44,342
Erg11	CYP51F1	P10613	ORF19.922	*C. albicans*	ESLR	PERW	FGTGPHVCLG	528	60,675
Erg5		O94016	ORF19.5178	*C. albicans*	ETLR	PTRW	FGGGRHRCIG	517	59,652
PAH1	CYP501	Q59PA5	ORF19.1411	*C. albicans*	ETLR	PDRW	FGAGSRMCSG	600	69,012
Tri4		Q5A0Z1	ORF19.3105	*C. albicans*	ENGR	PERW	FGKGVRMCLG	519	60.064

Table 3.2 – The 10 putative cytochromes P450 in *C. albicans*. The conserved regions are highlighted (the PER(W) domain, the haem binding domain and K-helix).

3.3.2 Homologue of CYP52A3

Results from the BLAST query showed that the potential alkane/fatty acid hydroxylating cytochromes P450 highlighted in the search of the *Candida* genome had varying degrees of homology to CYP52A3. Alk1 was shown to have the highest homology to *C. maltosa* CYP52A3 because it had the highest score and the lowest E value of all the *C. albicans* genes meaning the probability of this homology being a coincidence was low (see figure 3.7).

To ensure *ALK1* was in fact the gene with the greatest homology to *CYP52A3*, the protein sequences of CYP52A3, Alk1, Alk2, Alk3, Alk6 and Alk8 were all aligned using ClustalX (Thompson *et al.*, 1997) and viewed in BioEdit (Hall, 1999) (see figure 3.8). Again, Alk1 was shown to have the greatest homology to CYP52A3, which was confirmed visually and by comparing the percentage identities of all the potential CYP52s (see table 3.3).

Final confirmation came in the form of a phylogenetic tree (see figure 3.9). This showed that Alk1 does in fact have the highest homology of all the potential *C. albican* CYP52s to CYP52A3 as Alk1 shares a closer evolutionary history to CYP52A3 than the other potential *C. albican* CYP52s do.

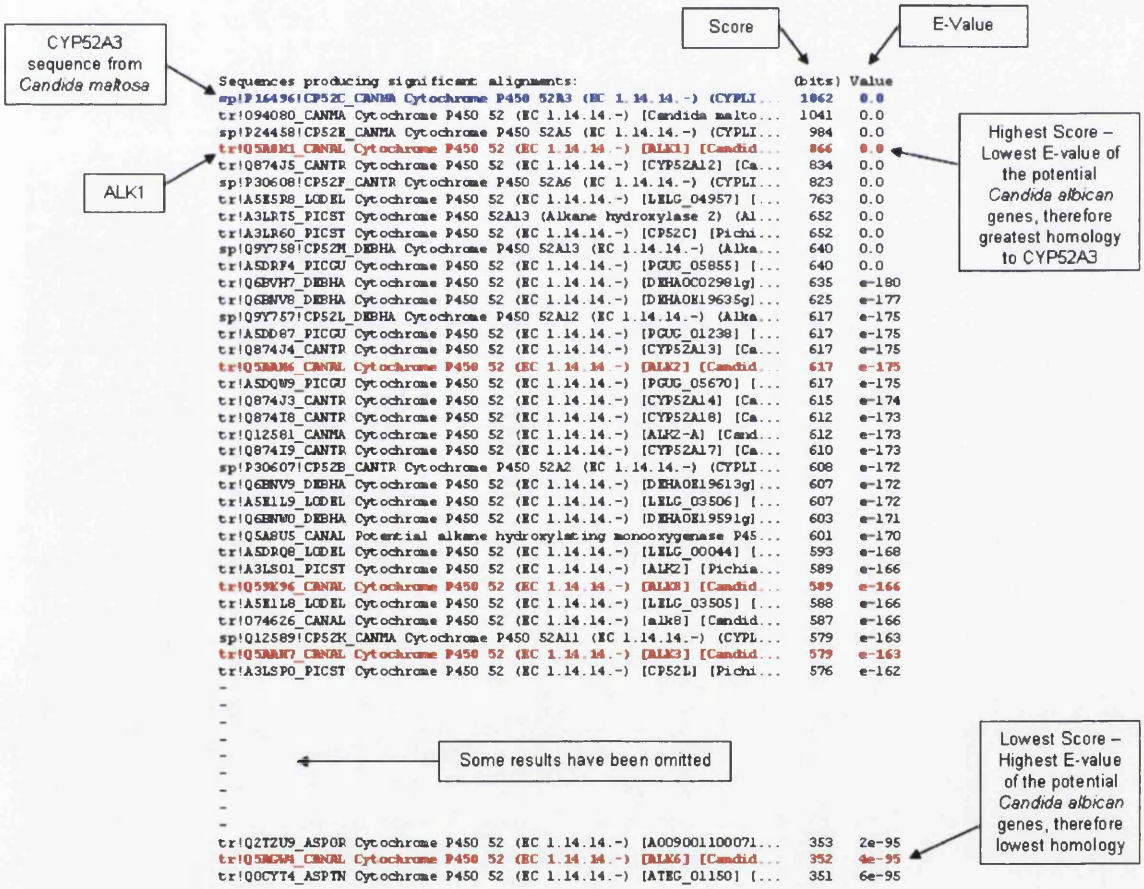


Figure 3.7 - BLAST query results. Results are listed in order of homology to the queried CYP52A3 amino acid sequence, with the sequences with the highest score and lowest E-value at the top, and the lowest score and highest E-value at the bottom. The first result is CYP52A3 itself and is highlighted in blue. The sequences highlighted in red are the potential alkane/fatty acid hydroxylating cytochromes P450 found in the *C. albicans* genome.

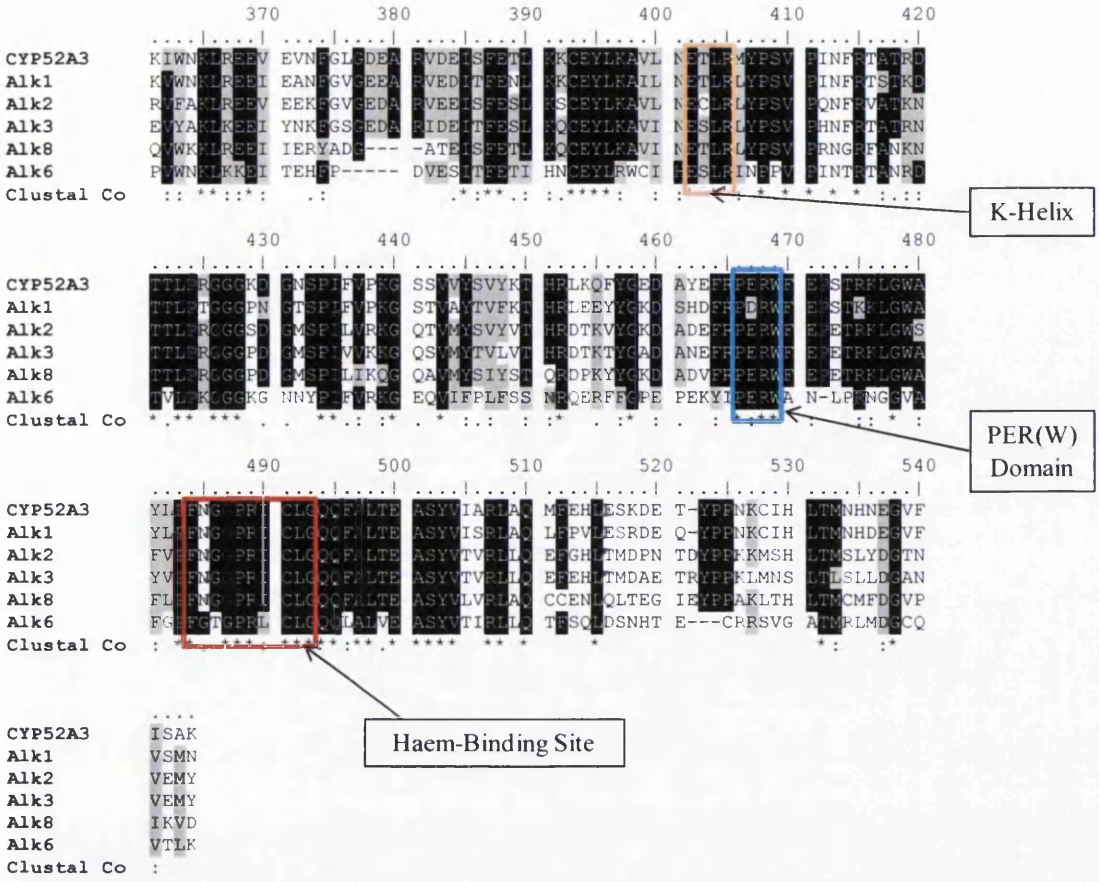


Figure 3.8 – Section of the alignment involving CYP52A3 and the potential *C. albican* alkane hydroxylating cytochromes P450 (from residue 361 to the end). The alignment was undertaken in ClustalX (Thompson *et al.*, 1997) and viewed in BioEdit (Hall, 1999). Threshold for shading was set at 80%. Identical residues are highlighted in black and similar residues are coloured grey. The haem-binding site, K-helix and PER(W) domain are indicated on the alignment.

Protein Name	Accession Name	Identities	
Alk1	Q5A8M1	409/516	79%
Alk2	Q5AAH6	314/521	60%
Alk8	Q59K96	306/507	60%
Alk3	Q5AAH7	302/529	57%
Alk6	Q5AGW4	195/447	43%

Table 3.3 – Potential alkane/fatty acid hydroxylating CYP52s identified in *C. albicans* (http://www.expasy.org/cgi-bin/blast.pl). All proteins compared to CYP52A3. Alk1 is shown to have the highest identity (to CYP52A3) of all the potential CYP52s identified in this table.

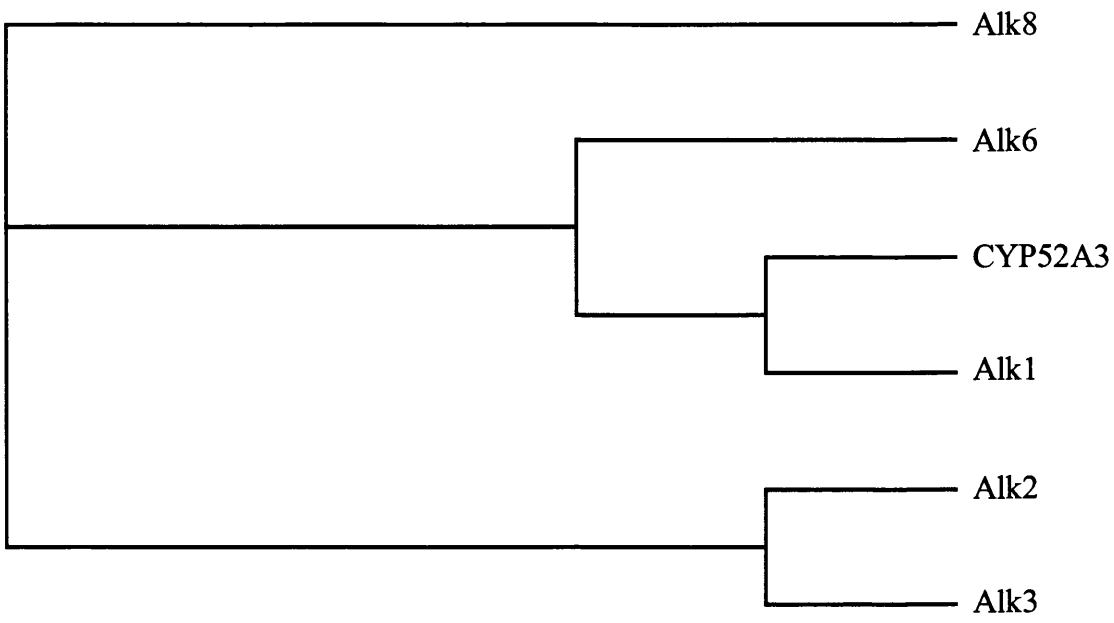


Figure 3.9 - Phylogenetic tree showing the relationship between CYP52A3 from *C. maltosa* and the potential CYP52s from *C. albicans*. This phylogenetic tree shows that Alk1 and CYP52A3 share a close evolutionary history, suggesting that Alk1 has a greater homology to CYP52A3 than the other potential alkane/fatty acid hydroxylating proteins. The distance between Alk1 and CYP52A3 is minimal, which indicates there are few amino acid changes between the two sequences showing relatively high homology.

3.4 Discussion

There is no doubt that since the conception of bioinformatics it has become an integral tool in the biological sciences. However, it is still in its infancy and it is still evolving. More attention needs to be focused on the nucleotide databases as currently they are relatively redundant. A second generation of nucleotide databases need to be developed primarily dealing with the sequence duplication, contamination, mislabelling, sequence errors, etc. that make these databases redundant. This will reduce the size of the databases, making it quicker and easier for the user to access biological data. However, bioinformatics is still an excellent tool and is integral in the discovery of new cytochrome P450 genes.

Using a bioinformatic approach, it has been shown that there are 10 putative cytochrome P450 genes in *C. albicans*. These are *ALK1*, *ALK2*, *ALK3*, *ALK6*, *ALK8*, *TRI4*, *DIT2*, *PAH1*, *ERG5* and *ERG11*. Tri4, according to UniProt, has weak similarity to trichodiene oxygenase, a member of the CYP58 family, which is

involved in the biosynthesis of mycotoxin trichothecene (http://www.uniprot.org/uniprot/Q5A0Z1). However, a BLAST search showed that Tri4 shares homology with several uncharacterised proteins, as well as a putative N-formyltyrosine oxidase (Dit2) from *C. dubliniensis* (B9WF59, bit score = 933, E-value = 0.0, identities = 86%). Dit2 from *C. albicans* shares sequence homology with N-formyltyrosine oxidases (CYP56s) from, amongst others, *C. dubliniensis, S. cerevisiae, Aspergillus clavatus, A. fumigatus, Neosartorya fiocheri* and *Paracoccidioides brasiliensis.* This means Dit2 potentially has a similar function to other CYP56 proteins, which are known to be involved (in conjunction with Dit1) in the biogenesis of the meiotic spore wall of *S. cerevisiae* (Briza *et* al., 1994; Briza *et* al., 1996; Melo *et al.*, 2008). PAH1 is a phenylacetate hydroxylase, which is homologous to putative genes from *C. dubliniensis, Penicillium marneffei, Talaromyces stipitatus, Ajellomyces dermatitidis* and *Neosartorya fischeri* and is identified as CYP501 on the Cytochrome P450 Homepage (http://drnelson.uthsc.edu/CytochromeP450.html). Erg11 (also called CYP51 and lanosterol 14α-demethylase) from *C. albicans* shares homology with CYP51s from *C. glabrata, C. dubliniensis, C. krusei, S. cerevisiae, S. elongisporus, Kluyveromyces polysporus* and *Pichia stipitis.* CYP51s are associated with sterol production and are the main target of azole antifungal drugs (Kelly *et al.*, 2001). Erg5 (also called CYP61 and sterol 22-desaturase), on the other hand, shows homology to putative CYP61 proteins from *P. stipitis, P. pastoris, S. cerevisiae, Symbiotaphrina buchneri, S. kochii, Microsporium canis, A. clavatus, A. fiocherianus, A. fumigatus, A. niger* and *A. terreus.* Like CYP51, CYP61 is involved in sterol production, but is a sterol C-22 desaturase rather than a lanosterol 14α-demethylase (Kelly *et al.*, 1997). The remaining five putative cytochromes P450 in *C. albicans*; Alk1, Alk2, Alk3, Alk6 and Alk8 are potentially involved in alkane and/or fatty acid hydroxylation. Of these Alk1 has been shown to have the greatest homology to a known alkane hydroxylating cytochrome P450 from *C. maltosa*, CYP52A3 and is, therefore, of interest for the studies that follow.

Chapter 4 : Cloning and Protein Expression of Alk1, its Truncations and NADPH-Cytochrome P450 Reductase (CPR) from *Candida albicans* and CYP52A3 from *C. maltosa*

4.1 Introduction

In chapter 3 it was shown that Alk1 from *C. albicans* shares homology with CYP52A3, the archetypal alkane-hydroxylating cytochrome P450 in *C. maltosa*, suggesting Alk1 has the ability to hydroxylate alkanes. However, homology does not necessarily result in the same protein function. This has been shown with the CYP52A subfamily, of which CYP52A3 is a member. CYP52A3 and CYP52A4 from *C. maltosa* have been shown to hydroxylate both mid-to-long chain alkanes and fatty acids (Zimmer *et al.*, 1995; Zimmer *et al.*, 1996; Scheller *et al.*, 1998), whereas CYP52A13 and CYP52A17 from *C. tropicalis* have only been shown to hydroxylate fatty acids (Eschenfeldt *et al.*, 2003). All these genes share greater than 55% identity as they are all part of the same subfamily, but they cannot all hydroxylate alkanes, thus showing differences in activity between the proteins does exist. These differences are not exclusive to the CYP52A subfamily as similar occurrences have been shown to arise between other cytochromes P450. CYP77A4 and CYP77A6, from *Arabidopsis thaliana*, are both able to catalyse the in-chain hydroxylation of lauric acid (C12:0) (Li-Beisson *et al.*, 2009; Sauveplane *et al.*, 2009). However, only CYP77A4 has been shown to be involved in the epoxidation of unsaturated fatty acids, such as oleic acid (C18:1) (Sauveplane *et al.*, 2009). This again shows that exact function cannot be conferred from homology. Therefore, to study the substrates of Alk1 and the reactions it is involved in the protein needs to be expressed for *in vitro* studies.

Cytochromes P450 can be expressed in a number of host organisms, most commonly *Escherichia coli* and yeast cells.

4.1.1 Expression in *E. coli*

E. coli is a preferred host for protein expression as the method is relatively quick and easy (Waterman *et al.*, 1995). The initial transformation of DNA into *E. coli* only requires a small amount of DNA to be successful and is thought to be relatively easy when compared to other host organisms, such as mammalian and insect cells (Verma

et al., 1998). Plasmid vectors used for the cloning and expression of individual genes in *E. coli* contain an origin of replication (*ori*) allowing for efficient replication in the host and multiple restriction cleavage sites for insertion of the gene fragment into the vector. The vector usually contains an antibiotic resistance gene, such as Amp^r (which conveys resistance to ampicillin) allowing *E. coli* colonies containing the plasmid to grow in media containing the antibiotic. Those cells that do not contain this plasmid will not grow in this media, meaning the antibiotic can be used as a selectable marker. Cloning vectors, such as pUC and pGEM-T easy, can also be used for blue/white selection of colonies when complemented with an *E. coli* strain that contains a partial *lacZ* gene deletion (*lacZΔM15*), such as the XL-1 blue strain. The *lacZ* gene, which is also present in these vectors, makes up part of the *lac* operon. The *lac* operon is involved in the transport and metabolism of lactose in *E. coli* and requires the enzyme, β-galactosidase to digest lactose to glucose (Müller-Hill, 1996). *LacZ* encodes for β-galactoside, which is induced in the presence of isopropyl-β-D-thio-galactoside (IPTG). IPTG is a thio-galactoside, which is a compound that acts as a β-galactosidase inducer, but does not itself become hydrolysed by the enzyme (Müller-Hill, 1996). Unlike 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal), which is converted to galactose by β-galactosidase (Müller-Hill, 1996). The resultant galactose is then broken down further to produce an insoluble blue-indigo dye (Müller-Hill, 1996). This dye is produced if the *lacZ* gene of the vector is undisrupted during cloning thus producing colonies, which appear blue in the presence of X-Gal and IPTG. However, if the *lacZ* gene is disrupted by the insertion of another gene, e.g. the gene fragment, the β-galactosidase is no longer induced by IPTG so X-Gal is not broken down and this blue-dye is not produced. This results in the production of white colonies, thus giving rise to blue/white selection.

For protein expression, the gene of interest requires ligation into an expression vector. An expression vector is a vector that has been designed so that the gene fragment can be inserted at a particular site (within the vector sequence) enabling it to be transcribed and translated into protein (Alberts *et al.*, 2002). Expression vectors usually contain a strong promoter, which can be regulated by the addition of a regulatory chemical, such as IPTG, to initiate expression. Examples include: the *lac* promoter (Miller *et al.*, 1968), the *trp* promoter (Zubay *et al.*, 1972) and the *tac*

promoter (Amann *et al.*, 1983; de Boer *et al.*, 1983). The *lac* promoter is found upstream of the cloning site and can be induced by the addition of IPTG. The *trp* promoter is involved in tryptophan synthesis and, unlike the *lac* promoter, can be both induced and repressed. The addition of 3-β-indoleacrylic acid can induce the promoter and result in the transcription of the gene of interest. Alternatively, the addition of tryptophan represses the promoter so transcription cannot take place. The *tac* promoter is a *trp/lac* promoter hybrid making it a stronger promoter than the other two. However, like the *lac* promoter, it is induced by IPTG, leading to the expression of the down-stream gene of interest. The use of an expression vector with a strong promoter, along with the fast rate of growth of *E. coli* means that protein can be expressed in high yields and in a much shorter time than in other systems, such as mammalian cells (Verma *et al.*, 1998). However, the use of IPTG, particularly in large-scale protein expression, can increase the costs of this technique as the chemical is relatively expensive to purchase. Furthermore, the cost to express cytochrome P450 protein is increased by the need to add 5-aminolevulinic acid (Ala) to the media as it is a haem precursor required for cytochrome P450 function (Richardson *et al.*, 1995; Gillam *et al.*, 1995).

The gene sequence of the hydrophobic N-terminus of an individual membrane-bound cytochrome P450 can be altered to enhance the amount of protein produced by expression in *E. coli* (Barnes *et al.*, 1991). These alterations were first shown to aid protein expression of the bovine 17α-hydroxylase, CYP17A1 (Barnes *et al.*, 1991). In the gene sequence of this cytochrome P450, the second codon was altered to GCT (which encodes for alanine) as it is the preferred second codon for *lacZ* gene expression in *E. coli* (Looman *et al.*, 1987; Barnes *et al.*, 1991). Furthermore, codons three to eight were made AT-rich to prevent the formation of GC hairpins (Barnes *et al.*, 1991). Collectively, these alterations resulted in an increase in the concentration of cytochrome P450 produced. The yield of CYP17A1 protein (expressed in pCW vector and the *E. coli* strain JM109) increased to 750nmoles/L (using this method) from 0nmoles/L, which was produced using the native sequence and same growth/expression conditions (Barnes *et al.*, 1991; Waterman *et al.*, 1995). Therefore, modifying the sequence is a useful tool for enhancing protein expression in *E. coli*.

However, the overexpression of protein in *E. coli* can lead to the formation of inclusion bodies, which occur if expression exceeds the cells capability for production (Barnes, 1996). These inclusion bodies are the result of incorrectly folded/unfolded protein and lead to the formation of inactive cytochrome P450 protein (Williams *et al.*, 1982; Barnes, 1996).

4.1.2 Expression in Yeasts

Yeasts are also a common choice of host organism for protein expression. *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Pichia pastoris* have all been used to express cytochrome P450 protein, with *S. cerevisiae* AH22 being the most commonly used strain (Guengerich *et al.*, 1991).

Yeasts, unlike *E. coli*, contain native cytochrome P450 genes. For example, *S. cerevisiae* contains three cytochromes P450: CYP51, CYP61 and CYP56. Although, these are not spectrally detectable, allowing the organism to be used for protein expression experiments. However, this is an advantage as the presence of cytochromes P450 in the organism means NADPH-cytochrome P450 reductase (CPR) is also present as it is required for electron transport during cytochrome P450 activity (see section 1.2.1). This means that any microsomal fractions produced as a result of protein expression in this organism could be used for metabolism studies as CPR would be present in the fraction (Pompon *et al.*, 1996). However, in some cases (e.g. CYP3A4) additional CPR would be required for functional metabolism (Pompon *et al.*, 1996).

There are two types of vectors used for cloning and expression in *S. cerevisiae*: integrative and autonomous. Integrative vectors require integration into the chromosome of the yeast for replication, whereas autonomous vectors are able to replicate independently of the chromosome as they contain an autonomously replicating sequence (ARS), which contains a yeast origin of DNA replication. Autonomous vectors are the most commonly used and many are known as "shuttle vectors" as they also contain an *E. coli* origin of replication. This means these shuttle vectors are able to replicate in *E. coli*, as well as yeast, allowing manipulation and high-copy numbers of the plasmid to be undertaken quickly. These vectors also contain an antibiotic resistance gene (e.g. Amp^r or Tet^r) to allow selection on that

antibiotic (e.g. ampicillin or tetracycline). Yeast vectors usually contain a selectable gene marker, which allows colonies containing the plasmid to grow in the absence of certain amino acids that would normally be added to the media. These gene markers include: *LEU2* (Beggs, 1978), *TRP1* (Tschumper & Carbon, 1980), *HIS3* and *URA3* (Struhl *et al.*, 1979). Presence of one of these genes on the vector (e.g. *LEU2*) means that the amino acid (e.g. leucine) does not need to be added to the media as the amino acid can be synthesised independently of the cell. Therefore, using this plasmid with a complementary strain (i.e. a strain that is unable to synthesis that particular amino acid) means only cells containing this plasmid can grow in the absence of that amino acid allowing for the selection and growth of cells containing the gene of interest.

Yeast expression vectors require promoters for transcription. The most commonly used are the *GAL* promoters *GAL1* and *GAL10*, which are induced by the addition of galactose to the media (Broach *et al.*, 1983). Yeast preferentially grows on media containing sugars, most notably glucose, which can be used directly by glycolysis (conversion of sugar to pyridine producing ATP and NADH as by-products) for growth (Romanos *et al.*, 1992). However, if cells containing plasmids carrying such a promoter (such as YEp51 containing a *GAL10* promoter) are grown in glucose-rich media expression cannot be induced until all the glucose has been exhausted. This is because glucose represses the promoter involved in transcription thus preventing protein expression until all the glucose has been consumed (Romanos *et al.*, 1992). Using the *GAL10* promoter also means expensive chemicals, such as IPTG are not required to be added to the media as IPTG is not needed to induce transcription. Furthermore, the yeast cells are able to synthesise their own haem without the need for the addition of Ala, which *E. coli* cells are unable to do for high level expression. However, ferric chloride ($FeCl_3$) can be added to the media to enhance haem production.

There are many advantages and disadvantages to using *E. coli* and yeasts as host organisms for protein expression, but they remain the most commonly used expression systems to date. Table 4.1 outlines these advantages and disadvantages. Other organisms have also been used, to a lesser extent, for the heterologous expression of cytochromes P450. These include mammalian and insect cells.

Host Organism	Advantages	Disadvantages
E. coli	• Relatively quick • Vector manipulation is relatively easy • The amount of membrane-bound protein produced can be increased by altering the N-terminus (e.g. making it AT-rich)	• Requires Ala as a haem precursor • Overexpression can lead to the formation of inclusion bodies • Lack of post-translational modifications
Yeast	• Does not require Ala as a haem precursor (ferric chloride ($FeCl_3$) can be added to enhance haem production instead) • Contains native cytochromes P450 and CPR • Post-translation modifications	• Slower growth than *E. coli* • Cell wall is difficult to break • Vector manipulation is more difficult than with *E. coli*

Table 4.1 – Overview of the advantages and disadvantages of using *E. coli* and yeast as host organisms for protein expression.

4.1.3 Expression in Mammalian Cells

The use of the mammalian expression system is uncommon as it is relatively expensive and expresses protein at low levels compared to the *E. coli* and yeast systems (Guengerich *et al.*, 1991). However, there are two types of this expression system available (COS cells and the vaccinia virus) and they have both been used to express cytochromes P450.

COS cells are the most commonly used mammalian expression system and are produced by integrating SV40 viral DNA into kidney cells (CV-1) from the African Green Monkey (Gluzman, 1981). This cell line has been used to express a number of human cytochromes P450 including CYP1A1 (McManus *et al.*, 1990), CYP3A4 (Ball *et al.*, 1992) and CYP4A11 (Imaoka *et al.*, 1993). However, because of the low expression levels associated with this expression system, it is difficult to properly

measure the cytochrome P450 protein content so this system is primarily used for the screening of cytochrome P450 activity (Gonzalez & Korzekwa, 1995).

The vaccinia virus has also been used to express a number of cytochrome P450 proteins, including the human genes CYP4B1 (Nhamburo *et al.*, 1989) and CYP2A6 (Yamano *et al.*, 1990). However, this expression system is not widely used as the production of the recombinant virus is time-consuming and its availability is quite low (Gonzalez & Korzekwa, 1995). There is also a safety concern with this system as the vaccinia virus is derived from the cow pox virus and there are fears of human infection despite the virus being attenuated (Gonzalez & Korzekwa, 1995).

4.1.4 Expression in Insect Cells

Insect cells have been used as a host for the expression of insect cytochrome P450 proteins, such as CYP6B33 (Mao *et al.*, 2008). However, at present only a limited number of cytochromes P450 from other organisms have been expressed in this system. These include CYP4A11 (Imaoka *et al.*, 1993), CYP3A4 (Buters *et al.*, 1994), CYP19A1 (Sigle *et al.*, 1994), CYP2A13 (Zhou *et al.*, 2010) and CYP2A25 (Zhou *et al.*, 2010). However, higher levels of protein have been produced using an insect cell system when compared with expression in mammalian cells (Verma *et al.*, 1998). The most common type of expression system in insect cells is the baculovirus system as there are over 500 species available, which unlike the vaccinia virus, are not potentially harmful to humans as they are restricted to use in insect cells (Groner, 1986; Gonzalez *et al.*, 1991; Verma *et al.*, 1998). As with the vaccinia virus, the process to produce recombinant bacluovirus is time-consuming (Gonzalez & Korzekwa, 1995) thus putting the systems at a disadvantage when compared to other expression systems, such as *E. coli* and yeast.

Of the cytochrome P450 protein expression systems described above, *E. coli* and yeast cells are the most commonly used host organisms. However, *E. coli* is the preferred host organism for protein expression as it is readily available and relatively quick and easy to use (Waterman *et al.*, 1995). Therefore, in this chapter *ALK1* will be cloned and expressed in *E. coli* as the gene of interest, as in chapter 3 it was shown to have the greatest homology to the archetypal alkane-assimilating cytochrome P450, *CYP52A3* from *C. maltosa*. Expressing the protein will allow

further studies to be undertaken to identify potential substrates through substrate binding and reconstitution assays thus potentially deorphanising Alk1. Truncations of this gene will also be designed to aid the heterologous expression of Alk1 and to be potentially used for crysatillisation studies. The production of a crystal of this protein would increase the understanding of its structure. As Alk1 is a membrane-bound protein it would first need to be truncated as the hydrophobic N-terminus would interfere in crystal production. *CYP52A3* will also be cloned and expressed as will *CPR* from *C. albicans* (*CaCPR*). CYP52A3 will be used as a positive control in reconstitution assays and CaCPR will be used in these assays to transfer electrons from NADPH during the monooxygenase cycle (i.e. insertion of an oxygen atom into the substrate). Therefore, the objectives of this chapter are to clone and express the proteins of *ALK1* and *CaCPR* from *C. albicans* and *CYP52A3* from *C. maltosa*. Also attempts to design an active truncated form of *ALK1* and to clone and express its protein are described.

4.2 Materials and Methods

4.2.1 Media

4.2.1.1 Terrific Broth (TB) for Protein Expression in *E. coli*

Based on the recipe of Tartoff & Hobbs, 1987.

12g of tryptone, 24g of yeast extract and 8ml of 50% (v/v) glycerol were added to 900ml of distilled water and autoclaved.

4.2.1.2 Minimal Media for Protein Expression in *S. cerevisiae*

4.2.1.2.1 Minimal Media Containing Glucose as the Initial Carbon Source

0.67% (w/v) yeast nitrogen base (without amino acids) (Difco) and 2% (w/v) glucose were added to 1 litre of distilled water and autoclaved.

4.2.1.2.2 Minimal Media Containing Raffinose as the Initial Carbon Source

Based on the recipe of Zimmer *et al.*, 1995.

0.67% (w/v) yeast nitrogen base (without amino acids), 2% (w/v) raffinose and 1mg/L $FeCl_3$ (Sigma) were dissolved in 1 litre of distilled water and autoclaved.

4.2.2 Buffers

4.2.2.1 Lysozyme Buffer for Protein Preparation (*E.coli*)

0.25M sucrose, 50mM Tris HCl (Melford) pH7.4, 0.5mM EDTA (Sigma-Aldrich) and 1mg/ml lysozyme (Sigma-Aldrich). Stored at 4°C.

4.2.2.2 Sonication Buffer for Protein Preparation (*E.coli*)

50mM potassium phosphate pH7.4, 20% (v/v) glycerol, 0.1mM DTT, 0.1mM EDTA, 0.5M sodium acetate (Sigma-Aldrich) and 2% (w/v) sodium cholate (Acros Organics). Stored at 4°C.

4.2.2.3 Equilibration Buffer for Ni^{2+}-NTA Agarose Purification

0.1M potassium phosphate pH7.4 (19.8ml of 1M KH_2PO_4 and 80.2ml of 1M K_2HPO_4) and 20% (v/v) glycerol. Stored at 4°C.

4.2.2.4 Wash Buffer for Ni^{2+}-NTA Agarose Purification

0.5M sodium chloride, dissolved in equilibration buffer. Stored at 4°C.

4.2.2.5 Elution Buffer for Ni^{2+}-NTA Agarose Purification

60mM L-Histidine (Acros Organics), dissolved in wash buffer. Stored at 4°C.

4.2.3 Methods

4.2.3.1 Design of Alk1 Truncations

Alk1 truncations were designed using the following web-based software programs: TMHMM (http://www.cbs.dtu.dk/services/TMHMM) (figure 4.1), HMMTOP (http://www.enzim.hu/hmmtop/) (figure 4.2) and SignalP (http://www.cbs.dtu.dk/services/SignalP/) (figure 4.3). TMHMM and HMMTOP use HMM-driven algorithms based on compiled datasets to predict the membrane anchor region of a protein, whereas SignalP predicts signal peptide cleavage sites rather than the membrane anchor region. Together the results from each of these programs can be used to indicate residues, which may be omitted from the Alk1 protein to produce a truncated version lacking the hydrophobic N-terminus.

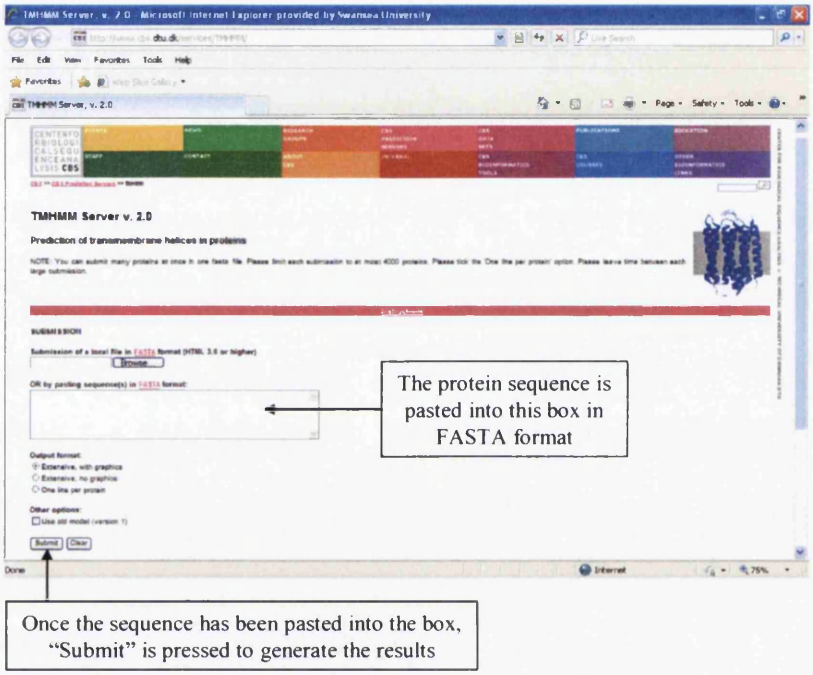


Figure 4.1 – TMHMM (http://www.cbs.dtu.dk/services/TMHMM). TMHMM is a web-based program, which can be used to predict the membrane anchor region of a protein.

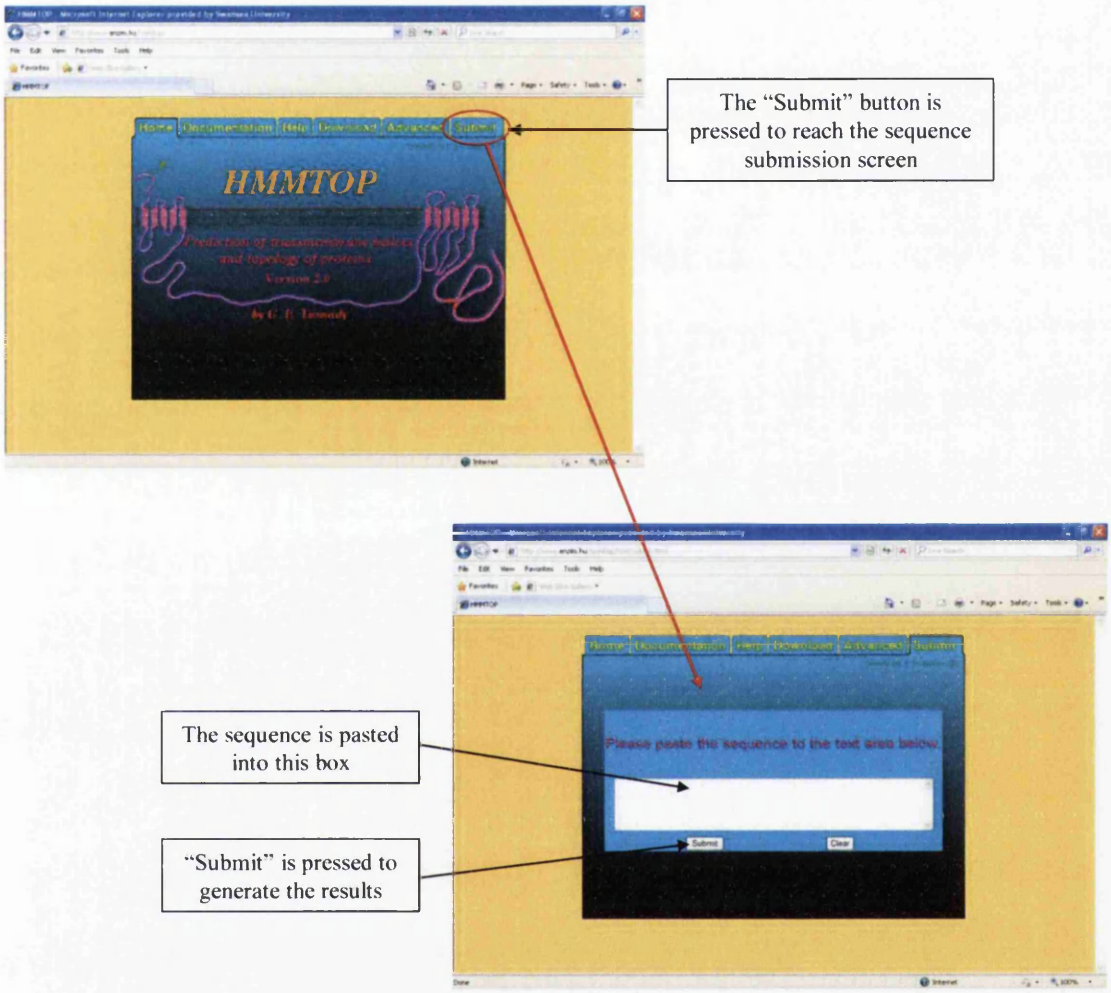


Figure 4.2 – HMMTOP (http://www.enzim.hu/hmmtop/). HMMTOP can be used to predict the potential topology and membrane anchor regions of the protein.

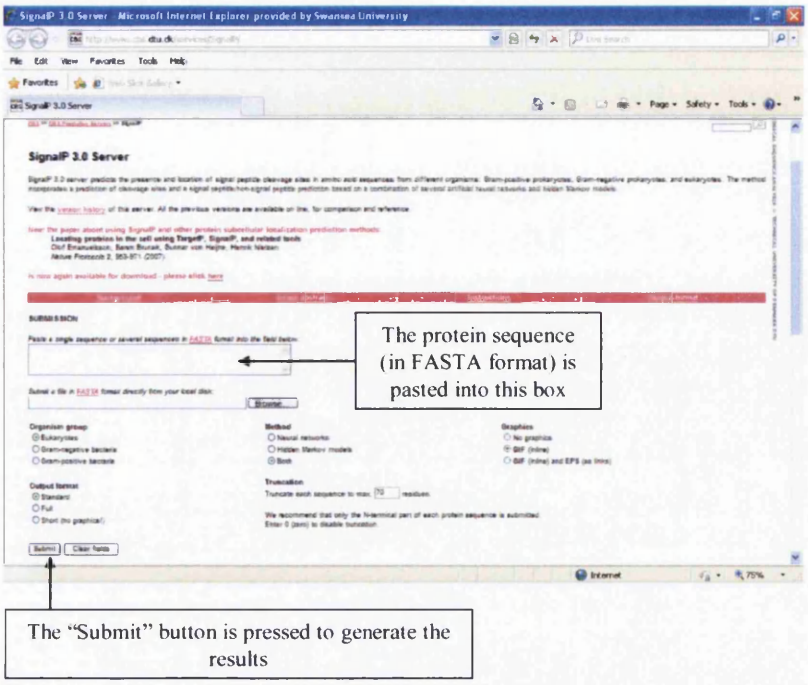


Figure 4.3 – SignalP (http://www.cbs.dtu.dk/services/SignalP). SignalP can be used to predict the signal peptide cleavage sites in a protein sequence.

4.2.3.2 PCR

The complete open reading frames (ORF) of *ALK1*, its truncations, *CaCPR* and *CYP52A3* were amplified from extracted genomic or plasmid DNA using primers that flank the 5' and 3' end of the coding sequence (see table 4.2). Each primer contains a region coding for a restriction enzyme that enables the gene to be cut and ligated into the expression vector of choice (i.e. pCWori$^+$). For this study, the coding sequence of each gene was flanked by *Nde* I and *Hin*d III restriction sites, which allows the gene to be ligated into the vector pCWori$^+$. These flanking regions of the primer do not bind to the template DNA, but during PCR they are replicated producing PCR products that have incorporated terminal restriction sites (Brown, 1995). Both the forward and reverse primers for each gene also contain a GC clamp to aid with PCR (Abrams *et al.*, 1990; Traystman *et al.*, 1990). The reverse primers have a four codon histidine tag to support protein purification by affinity chromatography (using Ni^{2+}-NTA) followed by a stop codon (TAA). Both are incorporated into the terminal ends of the PCR product in the same way as the restriction sites (described above). The *CaCPR* primers used were produced based on the design of Park *et al.*, 2010 (table 4.2).

Primer Name	Primer Sequence	Expected Size
>Alk1_F	5'-CGCCATATGTCCTCAACTACATTC-3'	1581bp
>Alk1_R	3'-GCTAAGCTTTTACTGATGGTGATGGTTCATTGAAACAAAGAC-5'	
>STNL_F	5'-CGCCATATGGCGAGCACCAACCTGAAA-3'	1479bp
>KAKN_R	3'-GCTAAGCTTTTAGTGATGGTGATGATTCATGCTCACGAACAC-5'	
>KAKN_F	5'-CGCCATATGGCTAAAGCTAAAAACAGC-3'	1371bp
>KAKN_R	3'-GCTAAGCTTTTAGTGATGGTGATGATTCATGCTCACGAACAC-5'	
>tAlk1_F1	5'-CATCGCCATATGGCTAAGAAAACGAGCTCTAAAGGGAAGCTCCCGGTCCATCTGAAACGTGCG-3'	1461bp
>KAKN_R	3'-GCTAAGCTTTTAGTGATGGTGATGATTCATGCTCACGAACAC-5'	
>tAlk1_F2	5'-CATCGCCATATGGCTAAGAAAACGAGCTCTAAAGGGAAACTGGGCTGCAAAAATCCG-3'	1473bp
>KAKN_R	3'-GCTAAGCTTTTAGTGATGGTGATGATTCATGCTCACGAACAC-5'	
>CYP52A3_F	5'-CGCCATATGGCTATAGAACAAATTATT-3'	1572bp
>CYP52A3_R	3'-GCTAAGCTTTTAGTGATGCTGATGTTTAGCAGAAATAAA-5'	
>CaCPR_F	5'-CGAATCATATGGCATTAGACAAATTA-3'	2043bp
>CaCPR_R	3'-ATTATTAAGCTTCCGGAAGCTTTTAGTGATGGTGATGGTGATGCCAAACATCTTCTTG-5'	

Table 4.2 – Forward and reverse primers used for PCR of *ALK1*, its truncations, *CYP52A3* and *CaCPR*. The gene sequence is underlined. The restriction sites are highlighted in blue (*Nde* I on the forward primer and *Hind* III on the reverse). The GC clamps are shown in red and the histidine tag is in orange. The stop codon is coloured purple. In bold is the modified N-terminus used for the *modifiedΔ50:ALK1* and *modifiedΔ44:ALK1* truncations.

ALK1 was amplified by PCR from extracted *C. albicans* DNA using the proof-reading enzyme, Platinum® *Pfx* DNA polymerase (Invitrogen). The reaction mixture was made up to a final volume of 50µl by mixing together 27µl of distilled water, 5µl of 10x *Pfx* amplification buffer (Invitrogen), 1.5µl of 10mM dNTP mix (Promega), 3µl of 50mM $MgSO_4$ (Invitrogen), 1.5µl of the forward primer, 1.5µl of the reverse primer, 0.5µl of Platinum® *Pfx* DNA polymerase (1 Unit) (Invitrogen), 5µl of *C. albicans* genomic DNA and 5µl of 10x PCR_x Enhancer solution (Invitrogen). The PCR program was set-up with an initial step at 94°C for 2 minutes. This was followed by 25 cycles of denaturation set at 94°C for 15 seconds, the annealing temperature set at 45°C for 30 seconds and elongation temperature set at 68°C for 2 minutes. Lastly, a 4°C ad infinitum step was added.

To optimise PCR product formation, *ALK1* was titrated against different concentrations of magnesium and genomic DNA (see table 4.3b). The thermal cycling conditions were also altered to try and optimise the reaction (see table 4.3a). Initially, the number of cycles was increased from 25 to 35 cycles. The elongation time was then increased from 2 minutes to 4 minutes with an extra cycle of 2 minutes added on at the end. The duration of the denaturation stages was also increased. The initial cycle was increased from 2 minutes to 5 minutes and the denaturation stage as part of the 35 cycles was doubled from 15 seconds to 30 seconds. The amount of each primer was also increased from 1.5µl to 3µl. The optimisation steps described above are shown in table 4.3b.

The final optimised PCR parameters for the amplification of *ALK1* are as follows. The reaction mixture was made up to a final volume of 50µl by mixing together 22µl of distilled water, 5µl of 10x *Pfx* amplification buffer, 1.5µl of 10mM dNTP mix, 3µl of $MgSO_4$, 3µl of the forward primer, 3µl of the reverse primer, 0.5µl of Platinum® *Pfx* DNA polymerase, 7µl of *C. albicans* genomic DNA and 5µl of 10x PCR_x Enhancer solution. The PCR program was set-up with an initial step at 94°C for 5 minutes. This was followed by 35 cycles of denaturation set at 94°C for 30 seconds, the annealing temperature set at 45°C for 30 seconds and the elongation temperature set at 68°C for 4 minutes. Lastly, a single elongation step at 68°C for 2 minutes was added with a final, ad infinitum, step at 4°C.

ALK1 truncations; *Δ33:ALK1*, *Δ69:ALK1*, *modifiedΔ50:ALK1* and *modifiedΔ44:ALK1* were amplified from pCWori^{+}:Alk1L456S plasmid DNA using the primers detailed in table 4.2 and Expand High Fidelity polymerase (Roche). The reaction mixture was made up to a final volume of 50µl by mixing together 23.25µl of distilled water, 1µl of 10mM dNTP mix (Promega), 3µl of the forward primer, 3µl of the reverse primer, 5µl of 10x Expand High Fidelity buffer (Roche), 7µl of 25mM magnesium (Roche), 7µl of pCWori^{+}:Alk1L456S plasmid DNA and 0.75µl of Expand High Fidelity polymerase (Roche). The thermal cycling conditions were set-up with an initial denaturation step at 94°C for 5 minutes. This was followed by 35 cycles of denaturation at 94°C for 15 seconds, the annealing temperature at 55°C for 30 seconds and the elongation temperature at 68°C for 2 minutes. An additional 1 cycle elongation step of 68°C for 7 minutes was added before a final 4°C ad infinitum step.

CYP52A3 was amplified from *C. maltosa* genomic DNA using the same method as described for the Alk1 truncations above.

CaCPR was amplified from *C. albicans* genomic DNA using Expand High Fidelity DNA polymerase. The reaction mixture was made up as follows: 2µM of 10mM dNTP mix, 3µl of the forward primer, 3µl of the reverse primer, 5µl of 10x Expand High Fidelity buffer, 3µl of 25mM magnesium and 7µl of *C. albicans* genomic DNA. The mixture was made to a final volume of 50µl with distilled water. The thermal cycling conditions were set up as follows. An initial denaturation step of 94°C for 5 minutes was followed by 35 cycles of denaturation at 94°C for 15 seconds, the annealing temperature at 45°C for 30 seconds and elongation at 68°C for 2 minutes. A final elongation step of 68°C for 7 minutes was added before a final 4°C ad infinitum step.

In all cases the PCR product was visualised on a 1% (w/v) agarose gel.

a)

	Step	Temperature	Duration	No. Of Cycles
#1	Denaturation	94°C	2 mins	x1
	Denaturation	94°C	15 secs	x25
	Annealing	45°C	30 secs	
	Elongation	68°C	2 mins	
	Cooling	4°C	∞	x1
#2	Denaturation	94°C	2 mins	x1
	Denaturation	94°C	15 secs	x35
	Annealing	45°C	30 secs	
	Elongation	68°C	2 mins	
	Cooling	4°C	∞	x1
#3	Denaturation	94°C	5 mins	x1
	Denaturation	94°C	30 secs	x35
	Annealing	45°C	30 secs	
	Elongation	68°C	4 mins	
	Elongation	68°C	2 mins	x1
	Cooling	4°C	∞	x1

b)

	1					2	3	
Reaction Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Water	27µl	24µl	25µl	22µl	22µl	25µl	22µl	8µl
10x Pfx Amplification Buffer	5µl	5µl	5µl	5µl	5µl	5µl	5µl	5µl
50mM $MgSO_4$	3µl	6µl	3µl	6µl	3µl	3µl	3µl	3µl
10mM dNTP Mixture	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl
Forward Primer	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	3µl	10µl
Reverse Primer	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	1.5µl	3µl	10µl
Platinum® *Pfx* DNA Polymerase	0.5µl	0.5µl	0.5µl	0.5µl	0.5µl	0.5µl	0.5µl	0.5µl
DNA	5µl	5µl	7µl	7µl	10µl	7µl	7µl	7µl
10x PCR_X Enhancer Solution	5µl	5µl	5µl	5µl	5µl	5µl	5µl	5µl
Thermal Cycling Conditions	#1					#2	#3	

Table 4.3 – *ALK1* PCR methods. a) Shows the thermal cycling conditions used for each PCR reaction; b) The PCR reaction mixtures set-up.

4.2.3.3 A-Tailing and Ligation into pGEM®-T Easy Vector (Promega)

Blunt-ended *ALK1* PCR product was A-tailed as follows: 1µl of 15mM magnesium chloride (Promega), 1µl of 2mM dATP (Promega), 1µl of Taq polymerase (Promega) and 1µl of magnesium free *Taq* buffer (Promega) was added to 6µl of PCR product. This was placed in the thermal cycler for 30 minutes at 37°C. The mixture was then immediately incubated on ice in preparation for ligation into the

pGEM®-T easy vector. However, *CYP52A3*, *CaCPR*, *Δ33:ALK1*, *Δ69:ALK1*, *modifiedΔ50:ALK1* and *modifiedΔ44:ALK1* PCR products did not require A-tailing as the polymerase used for PCR (High Fidelity) contains *Taq*, which adds adenosine to the ends at the gene during PCR.

For ligation into the pGEM®-T easy vector system (Promega) reactions were made up containing 3µl of the (A-tailed) DNA product, 1µl of the pGEM®-T easy vector (Promegag), 5µl of the 2x rapid ligation buffer (Promega) and 1µl of T4 DNA ligase (Promega). A control was also set up with 1µl of the control insert DNA (Promega), 1µl of pGEM®-T easy, 2µl of water, 5µl of 2x rapid ligation buffer and 1µl of T4 DNA enzyme. All reactions were incubated overnight at 4°C.

The ligations were then transformed into competent XL-1 *E. coli* cells and grown on LB ampicillin plates containing X-Gal and IPTG for blue/white selection.

4.2.3.4 Ligation into pCWori$^+$ and YEp51

Isolated vector DNA for ligation into pCWori$^+$ was double digested with *Nde* I and *Hin*d III (section 2.2.6.2), visualised on a 1% (w/v) agarose gel (section 2.2.3) and the DNA was extracted (section 2.2.4). For ligation of DNA into YEp51 the DNA was double digested with *Sal* I and *Hind* III (section 2.2.6.2).

For ligation into the pCWori$^+$ or YEp51 vectors reactions were made up containing 1µl of T4 DNA ligase (Roche), 1µl of 10x T4 ligation buffer (Roche) and various ratios of vector to insert to make a final volume of 10µl. All reactions were incubated overnight at 4°C.

The ligations were then transformed into DH5α *E. coli* cells and spread onto LB agar plates containing ampicillin, which were incubated overnight at 37°C.

4.2.3.5 Cloning Strategies

Overview of the cloning strategies used to express the protein of Alk1, its truncations, CaCPR and CYP52A3 using *E. coli* and *S. cerevisiae* as the host organisms.

4.2.3.5.1 Summary of Cloning Strategy Used for the Expression of Protein in *E. coli*

Figure 4.4 shows the cloning strategy used to express *ALK1*, its truncations and *CaCPR* from *C. albicans* and *CYP52A3* from *C. maltosa* into *E. coli* in preparation for protein expression.

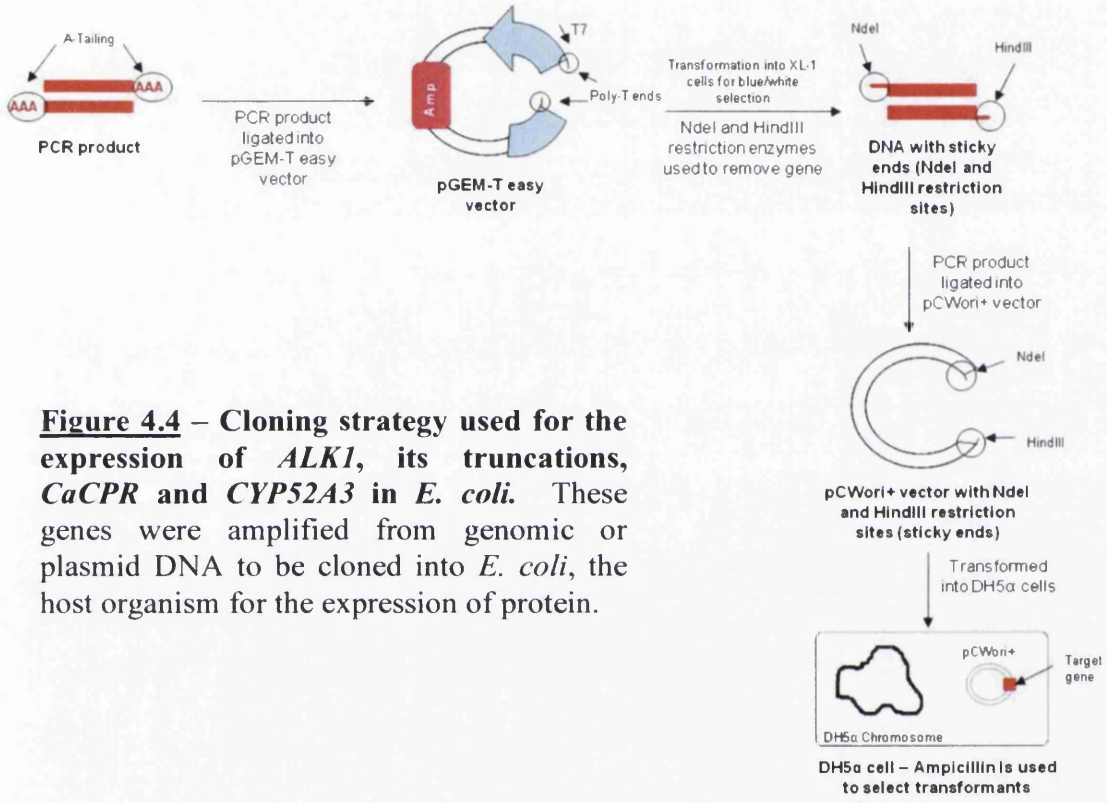


Figure 4.4 – Cloning strategy used for the expression of *ALK1*, its truncations, *CaCPR* and *CYP52A3* in *E. coli*. These genes were amplified from genomic or plasmid DNA to be cloned into *E. coli*, the host organism for the expression of protein.

Genomic DNA was extracted from *C. albicans* and *C. maltosa* (section 2.2.2) and used as the template for PCR of *ALK1*, *CaCPR* and *CYP52A3* (section 4.2.3.2). *Δ33:ALK1*, *Δ69:ALK1*, *modifiedΔ50:ALK1* and *modifiedΔ44:ALK1* were amplified by PCR from pCWori^{+}:Alk1L456S plasmid DNA. Following PCR, the PCR product was visualised on a 1% (w/v) agarose gel (section 2.2.3). The expected size of each gene (shown in table 4.2) was compared with the DNA marker. Correct size amplicons were isolated from the gel and the DNA was extracted (section 2.2.4). The extracted DNA was A-tailed (depending on the PCR polymerase used) and ligated into the vector, pGEM-T easy (section 4.2.3.3). The plasmids were then transformed into XL-1 *E. coli* cells and spread onto LB plates containing ampicillin, IPTG and X-Gal for blue/white selection (section 2.2.7). Individual white colonies

were picked and used to inoculate 10ml of LB broth containing ampicillin for overnight culturing. Plasmid DNA was isolated from 1ml of the culture (section 2.2.5) and digested with the restriction enzyme, *Not* I to check for the gene of interest (section 2.2.6.1) and visualised on a 1% (w/v) agarose gel. Plasmids containing inserts of the correct size were then subjected to a *Nde* I/*Hin*d III double digest (section 2.2.6.2) to release the gene from the pGEM-T® easy vector and to prepare the gene with sticky ends for ligation into pCWori$^+$. A 1% (w/v) agarose gel was used to visualise the band of interest and the correct sized amplicon was isolated from the gel and the DNA was extracted. The DNA sequence of the gene of interest in the pGEM-T easy plasmid was verified by sequencing (section 2.2.9).

The extracted DNA was ligated into pCWori$^+$ (section 4.2.3.4) and transformed into DH5α cells and spread onto LB plates containing ampicillin. Individual colonies were picked and used to inoculate 10ml of LB broth containing ampicillin for overnight culturing. Plasmid DNA was isolated from 1ml of the culture and digested with the restriction enzymes, *Nde* I and *Hin*d III to check for the presence of the insert (section 2.2.6.3). This was visualised on a 1% (w/v) agarose gel. The gene of interest was verified by sequencing (section 2.2.9).

4.2.3.5.2 Summary of Cloning Strategy Used for the Expression of Protein in *S. cerevisiae*

YEp51:CYP52A3_yeast plasmid was a kind gift from Dr. Wolf-Hägen Schunck, Max Delbrück Centre for Molecular Medicine, Berlin, Germany. The plasmid was transformed into *S. cerevisiae* (AH22 strain) cells and spread onto YM plates containing histidine (section 2.2.8).

4.2.3.6 Synthetic Genes – *ALK1L456S, ALK1L456S_yeast* and *Δ65:ALK1*

The synthetic genes, *ALK1L456S* and *Δ65:ALK1* were produced by GeneCust. and optimised for expression in *E. coli*. *ALK1_yeast* was optimised for expression in *S. cerevisiae*. All genes were designed to contain a codon encoding for serine at residue 456.

The *ALK1L456S* and *Δ65:ALK1* were designed as follows (see figure 4.5). The serine in the second position of both the protein sequences was altered to alanine and

the first six codons were made AT rich as these changes are thought to enhance protein expression in *E. coli* (Barnes *et al.*, 1991). The sequence for each gene was also designed to be flanked by two sets of restriction sites. At the 5' end, the protein sequence was flanked by restriction sites encoding for *Not* I and *Nde* I. These were complemented at the 3' end by restriction sited for *Not* I and *Hin*d III. *Nde* I and *Hin*d III were incorporated at either end of the sequence to allow ligation of the genes into the expression vector, pCWori^{+}, which had already been prepared with those restriction enzymes. However, the *ALK1L456_yeast* gene was designed to be flanked by a *Sal* I restriction site rather than *Nde* I to allow ligation into the yeast expression, vector, YEp51 (see figure 4.6).

Incorporated into both of these sequences at the 3' end was the stop codon TAG and a six codon histidine tag to aid with Ni^{2+}-NTA purification (see figures 4.5 and figure 4.6).

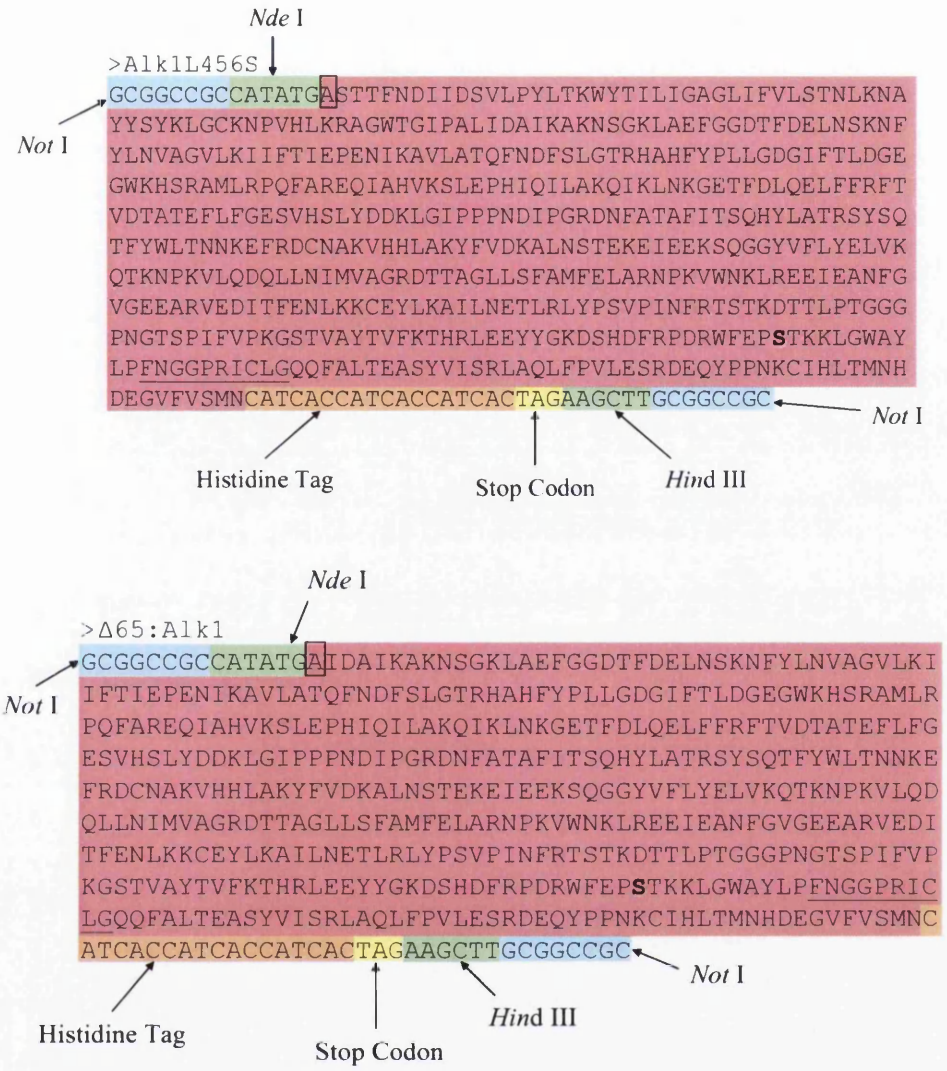


Figure 4.5 - Synthetic genes, *ALK1L456S* and *Δ65:ALK1*. The protein sequence was optimised for expression in *E. coli*. The first A (alanine) in the protein sequence is also highlighted because in the original, native sequence it is a serine but this has been changed to enhance protein expression in *E. coli* (Barnes *et al.*, 1991). The haem-binding site is underlined and the amino acid of interest (serine) is in bold. This protein sequence is flanked by nucleotide sequences, which do not require optimisation for expression in *E. coli* and are highlighted as follows. The 6 codon histidine tag is orange. The *Not* I restriction sites are highlighted in blue and the *Nde* I and *Hind* III sites are green. The stop codon is yellow.

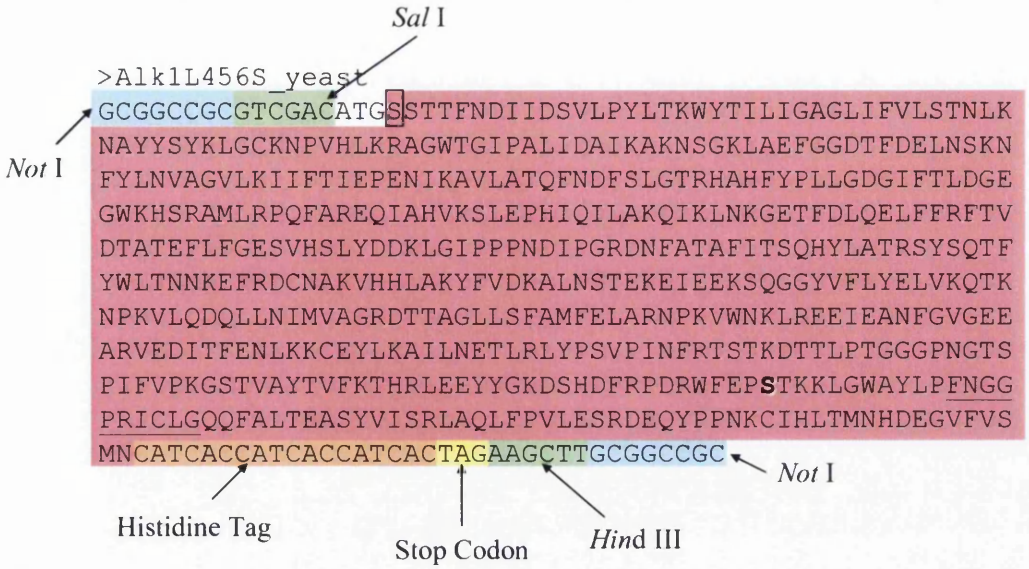


Figure 4.6 - Synthetic gene, *ALK1L456S_yeast*. The protein sequence was optimised for expression in *S. cerevisiae*. The haem-binding site is underlined and the amino acid of interest (serine) is in bold. This protein sequence is flanked by nucleotide sequences, which do not require optimisation for expression in *E. coli* and are highlighted as follows. The 6 codon histidine tag is orange. The *Not* I restriction sites are highlighted in blue and the *Sal* I and *Hind* III sites are green. The stop codon is yellow.

pUC57:Alk1L456S, pUC57:Alk1L456S_yeast and pUC57:Δ65:Alk1 were transformed into DH5α cells (section 2.2.7) and spread onto LB agar plates containing ampicillin. Individual colonies were grown overnight at 37°C in LB ampicillin broth and the plasmid was purified (section 2.2.5). *ALK1L456S* and *Δ65:ALK1* were liberated using a *Nde* I/*Hind* III digest (section 2.2.6.2). The DNA was extracted using the QIAquick Gel Extraction kit (section 2.2.4) and ligated into the expression vector pCWori$^+$ for protein expression (section 4.2.3.4). pCWori$^+$:Alk1L456S and pCWori$^+$:Δ65:Alk1 were transformed into DH5α cells and grown on LB + ampicillin agar plates (section 2.2.7). Individual colonies were grown overnight at 37°C in LB ampicillin broth and the plasmid DNA was extracted and purified (section 2.2.5). A *Nde* I/*Hind* III restriction enzyme digest (section 2.2.6.3) was used to screen for the presence of the *ALK1L456S* and *Δ65:ALK1* genes in each plasmid.

The *ALK1L456S_yeast* gene was ligated into the yeast expression vector, YEp51 and transformed into DH5α cells, which were grown on LB + ampicillin agar plates. Individual colonies were grown at 37°C overnight in LB ampicillin broth and the plasmid DNA was extracted and purified. A *Sal* I/*Hin*d III restriction enzyme digest was used to screen for the presence of the gene in the plasmid. The YEp51:Alk1L456S_yeast plasmid was then transformed into *S. cerevisiae* cells and grown at 30°C on YM plates containing histidine (section 2.2.8).

All genes were verified by DNA sequencing (section 2.2.9).

4.2.3.7 Protein Expression

4.2.3.7.1 Protein Expression using *E. coli* as the Host Organism

To 900ml of TB, 100ml of a solution comprising of 0.17M KH_2PO_4 and 0.72M K_2HPO_4, 1ml of ampicillin (100µl/ml) and 10ml of overnight *E. coli* (LB + ampicillin) culture were added. Cultures were incubated at different temperatures and for different durations and these conditions are shown in tables 4.4, 4.5, 4.6, 4.7 and 4.8. After growth, 1mM Ala and IPTG were added and the culture was incubated further (see tables 4.4, 4.5, 4.6, 4.7 and 4.8). However, the expression of CaCPR required IPTG to be added to the media at the start of the experiment (based on the Park *et al.*, 2010 method).

Experiment	Growth	Expression	
1	230rpm 37°C 7 hours	Ala 85mg IPTG 120mg	190rpm 25°C 18 hours
2	190rpm 30°C Overnight	Ala 170mg IPTG 240mg	160rpm 20°C Overnight
3	230rpm 37°C Overnight	Ala 170mg IPTG 240mg	200rpm 28°C 23 hours
4	230rpm 37°C 5 hours	Ala 85mg IPTG 120mg	200rpm 28°C 21 hours
5	230rpm 37°C 6 hours	Ala 85mg IPTG 120mg	200rpm 28°C 18 hours

Table 4.4 – Protein expression of native Alk1. The table shows the amendments made to the growth and expression conditions whilst trying to express active cytochrome P450. Each attempt resulted in a peak at 420nm, indication of protein that was not folded correctly. Arase *et al.*, 2006 method used.

Experiment	Growth	Expression	
1	230rpm 30°C 7 hours	Ala 170mg IPTG 240mg	190rpm 20°C 48 hours
2	230rpm 30°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40.5 hours
3	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40 hours

Table 4.5 – Protein expression of Alk1L456S. The table shows the changes made to the growth and expression conditions during optimisation. Arase *et al.*, 2006 method used.

Experiment	Growth	Expression		Method
1	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 17 hours	Arase *et al.*, 2006
2	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40 hours	Arase *et al.*, 2006
3	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 30°C 17 hours	Arase *et al.*, 2006
4	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 30°C 40 hours	Arase *et al.*, 2006
5	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 17 hours	Isolation of Membranes and Cytosol
6	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40 hours	Isolation of Membranes and Cytosol
7	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 30°C 17 hours	Isolation of Membranes and Cytosol
8	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 30°C 40 hours	Isolation of Membranes and Cytosol

Table 4.6 – Protein expression of Δ33:Alk1, Δ65:Alk1, Δ69:Alk1, modifiedΔ50:Alk1 and modifiedΔ44:Alk1. The table shows the changes made to the growth and expression conditions, along with the methods used, to try and express active protein.

Experiment	Growth	Expression		Method
1	230rpm 37°C 5 hours	Ala 170mg IPTG 240mg	190rpm 25°C 22 hours	Arase *et al.*, 2006
2	230rpm 37°C 7 hours	Ala 170mg IPTG 240mg	190rpm 25°C 17 hours	Arase *et al.*, 2006
3	230rpm 30°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40 hours	Arase *et al.*, 2006
4	230rpm 35°C 18 hours	Ala 170mg IPTG 240mg	190rpm 25°C 24 hours	Arase *et al.*, 2006
5	230rpm 37°C 5 hours	Ala 170mg IPTG 240mg	190rpm 25°C 22 hours	Isolation of Membranes and Cytosol

Table 4.7 – Protein expression of CYP52A3. The table shows the changes made to the growth and expression conditions, along with the methods used, to try and express active protein.

Growth	Expression	Method
200rpm 37°C 5 hours	200rpm 28°C 19 hours	Isolation of Membranes and Cytosol

Table 4.8 – Protein expression of CaCPR. The table shows the growth and expression conditions used, as based on the method of Park *et al.*, 2010.

4.2.3.7.1.1 Arase *et al.*, 2006 Method for Protein Isolation

Figure 4.7 shows an overview of the method used for protein isolation based on Arase *et al.*, 2006.

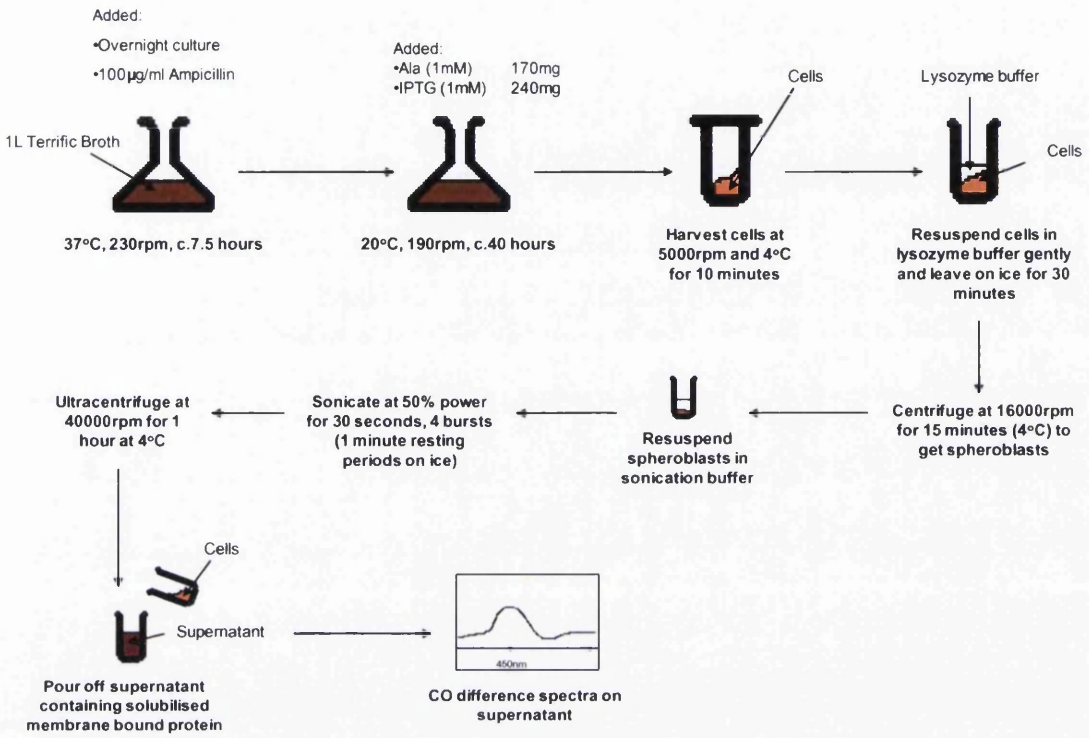


Figure 4.7 – Schematic diagram of the experimental approach used to express Alk1 protein. The figure shows an overview of the method used to express cytochrome P450.

Cells were harvested at 5000rpm (Beckman JLA8.1000 rotor) and 4°C for 10 minutes. The supernatant was discarded and the cells were gently resuspended in lysozyme buffer before being incubated on ice for 30 minutes. The resuspended cells were then centrifuged for 15minutes at 16000rpm (Beckman JA25.5 rotor) and 4°C to obtain spheroplasts. The spheroblasts were resuspended in sonication buffer and sonicated at 50% for 30 seconds in 4 bursts (with a rest period of 1 minute between bursts). The sonicated cells were then placed in the ultracentrifuge for 1 hour at 4°C and 40000rpm (Beckman type 50.2Ti ultracentrifuge rotor). The supernatant was then removed and the carbon monoxide (CO) difference spectrum was determined using the cytochrome P450 assay (section 4.2.3.10.1).

4.2.3.7.1.2 Isolation of Membranes and Cytosol

Cells were harvested as in method 4.2.3.7.1.1, except the spheroplasts were resuspended in 0.1M Tris HCl pH8.1 buffer instead of sonication buffer. Following sonication, the sonicated cells were then centrifuged at 16000rpm (Beckman JA25.5 rotor) for 10 minutes. The pellet was discarded and the supernatant was centrifuged again for 10 minutes at 16000rpm (Beckman JA25.5 rotor). The supernatent was then place in the ultracentrifuge for 1 hour at 40000rpm (Beckman type 50.2Ti ultracentrifuge rotor). The supernatant was removed and the membranes were resuspended in 0.1M Tris HCl, pH8.1 containing 25% (w/v) glycerol using a Potter-Elvehjeim glass homogenizer. The CO difference spectrum was determined for both the membranes and supernatant using the cytochrome P450 assay method (section 4.2.3.10.1). The CaCPR membranes were solubilised (section 4.2.3.8) prior to Ni^{2+}-NTA agarose purification (section 4.2.3.9).

4.2.3.7.2 Protein Expression using *S. cerevisiae* as the Host Organism

Based on the methods of Zimmer *et al.*, 1995; Shyadehi *et al.*, 1996 and Lamb *et al.*, 1999.

To 1 litre of minimal media (containing 2% (w/v) sugar (see table 4.9)), 10ml of 100mg/ml histidine (Acros Organics) and 20mls of overnight AH22 (YM + histidine) culture was added. Cultures were grown under different conditions, which are shown in table 4.9.

Experiment	Sugar Source	Growth	Expression	
1	2% (w/v) Glucose	200rpm 28°C 48 hours	Glucose exhausted ($5x10^7$cells/ml)	Induced with 3% (w/v) galactose
2	2% (w/v) Glucose	200rpm 30°C 48 hours	Glucose exhausted ($6x10^7$cells/ml)	Induced with 3% (w/v) galactose
3	2% (w/v) Raffinose	220rpm 28°C 48 hours	Cells grown to $8.6x10^7$cells/ml	Induced with 2% (w/v) galactose

Table 4.9 – Protein expression of CYP52A3 in AH22 yeast cells. The table shows the changes made to the growth and expression conditions and the sugar used as the initial carbon source in the media. In all experiments expression lasted 24 hours.

Cells grown in minimal media containing 2% (w/v) glucose (section 4.2.1.2.1) were induced with a final concentration of 3% (w/v) galactose (Acros Organics) following glucose exhaustion. Glucose exhaustion was tested using Clinistix reagent strips for urinalysis (Bayer). Alternatively, cells grown in minimal media containing 2% (w/v) raffinose (section 4.2.1.2.2) were grown to a cell density of $8.6x10^7$ cells/ml before induction with 2% (w/v) galactose (final concentration).

The protein was isolated from yeast using the following method. Yeast cells were harvested at 5000rpm (Beckman JLA8.1000 rotor) and 4°C for 10 minutes. The supernatant was discarded and the cells were resuspended in distilled water. The suspension was centrifuged at 3000xg and 4°C for 3 minutes. The supernatant was discarded and the cells were resuspended in 0.1M potassium phosphate, pH7.4 buffer containing 20% (w/v) glycerol. The yeast cells were mechanically disrupted using a Braun homogeniser with six 30 second bursts. Cells were cooled during homogenisation by liquid carbon dioxide. The homogenised mixture was centrifuged for 30 minutes at 4000xg and 4°C. The cell debris was discarded and the supernatant was centrifuged for a further 15 minutes at 16000rpm (Beckman JA25.5 rotor) and 4°C. The cell debris was again discarded and the supernatant centrifugation step was repeated. The supernatant was then ultracentrifuged for 1 hour at 40000rpm (Beckman type 50.2Ti ultracentrifuge rotor) and 4°C. The supernatant was removed and the microsomes were resuspended in 0.1M potassium

phosphate, pH7.4 containing 20% (w/v) glycerol using a Potter-Elvehjeim glass homogenizer. The CO difference spectrum was determined for the microsomes using the cytochrome P450 assay method (section 4.2.3.10.1).

4.2.3.8 Solubilisation of CaCPR Membranes for Ni^{2+}-NTA Agarose Purification

After expression of CaCPR in *E. coli*, the CaCPR membranes were diluted in 0.1M potassium phosphate, pH6.8 containing 25% (w/v) glycerol to a final volume of 50ml and stirred at a steady pace with a magnetic flea at 4°C. To the mix 0.5g of sodium cholate (Acros Organics) was added and the mixture was left to stir for 2 hours. The mixture was ultracentrifuged for 90 minutes at 4°C and 40000rpm (Beckman type 50.2Ti ultracentrifuge rotor). The supernatant was decanted to be used for Ni^{2+}-NTA agarose purification (section 4.2.3.9).

4.2.3.9 Ni^{2+}-NTA Agarose Purification of Protein

Ni^{2+}-NTA agarose (Qiagen) was washed prior to use with equilibration buffer 2-3 times until all traces of ethanol were removed and resuspended in 15ml of equilibration buffer and added to the protein sample and left to mix overnight (using a magnetic flea and stirrer) at 4°C. The solution was then centrifuged for 1 minute. The Ni^{2+}-NTA agarose-bound protein was retained and washed twice in equilibration buffer. The agarose/protein was resuspended in 15ml of equilibration buffer and poured into the column. The resin was washed in 2 column volumes of wash buffer. The protein was then eluted from the agarose using elution buffer and collected in 1ml fractions. The fractions were assayed to determine the concentration of cytochrome P450 or CPR protein (see section 4.2.3.10).

4.2.3.10 Determination of Protein Concentration

4.2.3.10.1 Assay to Determine Cytochrome P450 Concentration

The amount of cytochrome P450 protein in the sample was determined using the method of Omura & Sato, 1964.

The protein sample was diluted to 1ml using 100mM Tris HCl pH8.1, containing 25% (w/v) glycerol. CO was gently bubbled through the sample for 1min before it was split between the reference and sample cuvettes (in equal volumes). The baseline was then recorded between 400 and 500nm. A few grains of sodium

dithionite were added to the sample cuvette and mixed thoroughly by inversion. The spectrum was then recorded. The extinction coefficient, 91M cm^{-1} was used to calculate the concentration of cytochrome P450 in the sample (Omura & Sato, 1964).

4.2.3.10.2 Assay to Determine CaCPR Concentration

The absolute spectra of the protein sample was recorded by diluting the sample five fold in 0.1M potassium phosphate, pH6.8 containing 25% (w/v) glycerol to a final volume of 1ml. The baseline was recorded between 295nm and 705nm using buffer filled cuvettes. In the sample cuvette, the buffer was removed from the chamber and was replaced with the diluted sample (1ml). The spectrum was then recorded. The δAmax of the peak at 462nm and the extinction coefficient 21.4M cm^{-1} (Vermilion & Coon, 1978) were used to calculate the concentration of CaCPR in the sample.

4.3 Results

4.3.1 Design of Alk1 Truncations

To design truncations of Alk1, TMHMM, SignalP and HMMTOP were used to predict the hydrophobic N-terminal region of the protein. The results generated using TMHMM suggested that the membrane anchor region of Alk1 begins at residue 13 and ends at residue 35 (see figure 4.8). This was also shown with SignalP (see figure 4.9). The results from HMMTOP, however, predicted the end of the membrane anchor region at residue 33 (see figure 4.10). These results were used to form the basis of the Δ33:Alk1 truncation designed in this study.

The truncations Δ65:Alk1 and Δ69:Alk1 were designed based on the topology of CYP52A3, which is believed to have two predicted hydrophobic regions in its N-terminus (Menzel *et al.*, 1996).

ModifiedΔ50:Alk1 and modifiedΔ44:Alk1 were designed according to the method used to produce truncated versions of CYP2C3 and CYP2C5 where the amino acids prior to a proline-rich region of the sequence were deleted (Richardson *et al.*, 1993; von Wachenfeldt *et al.*, 1997; Cosme & Johnson, 2000; Wester *et al.*, 2002). The N-termini of these two truncations were also modified to MAKKTSSKGKL as this sequence was previously shown to aid the expression of full-length and truncated CYP2C3 in *E. coli* (Li & Chiang, 1991; Karam & Chiang, 1994).

Each truncation is described in further detail in section 4.4.2.

```
# Alk1 Length: 526
# Alk1 Number of predicted TMHs:  1
# Alk1 Exp number of AAs in TMHs: 22.14645
# Alk1 Exp number, first 60 AAs:  21.60348
# Alk1 Total prob of N-in:        0.88388
# Alk1 POSSIBLE N-term signal sequence
Alk1    TMHMM2.0        inside      1    12
Alk1    TMHMM2.0        TMhelix     13   35
Alk1    TMHMM2.0        outside     36   526
```

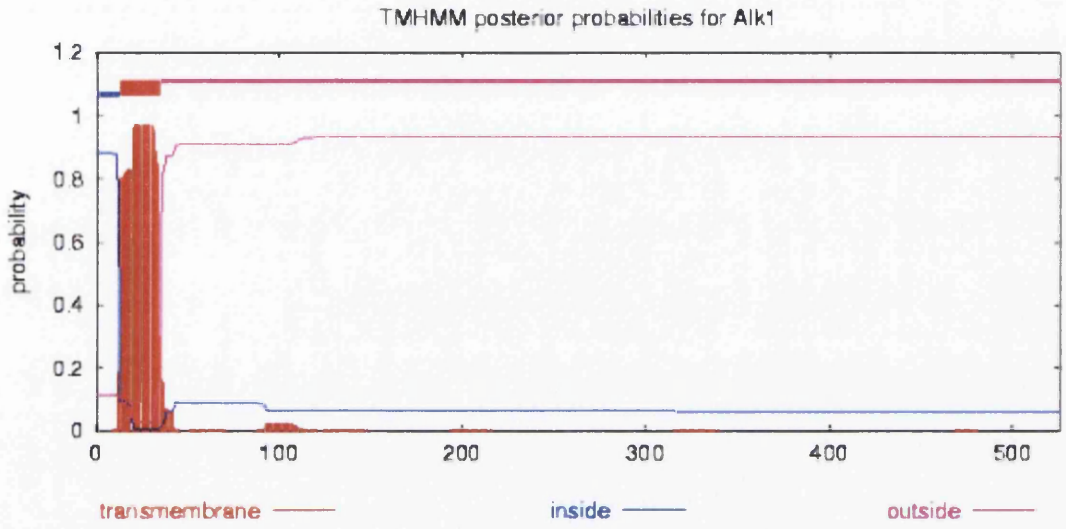
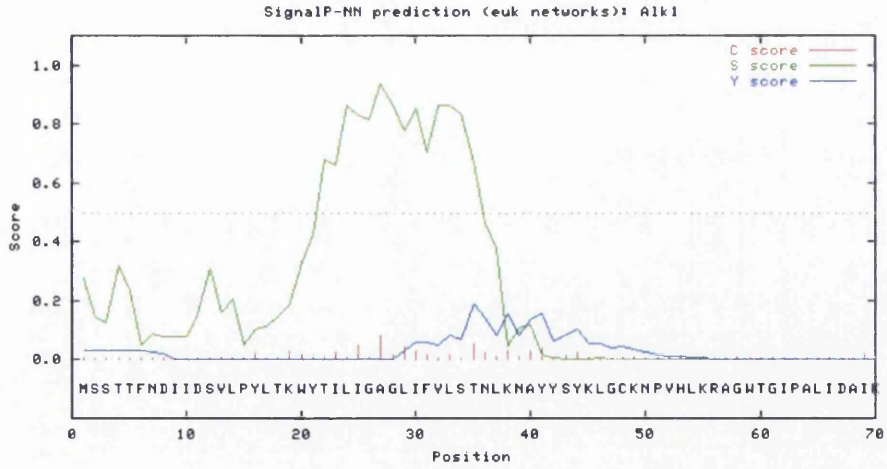
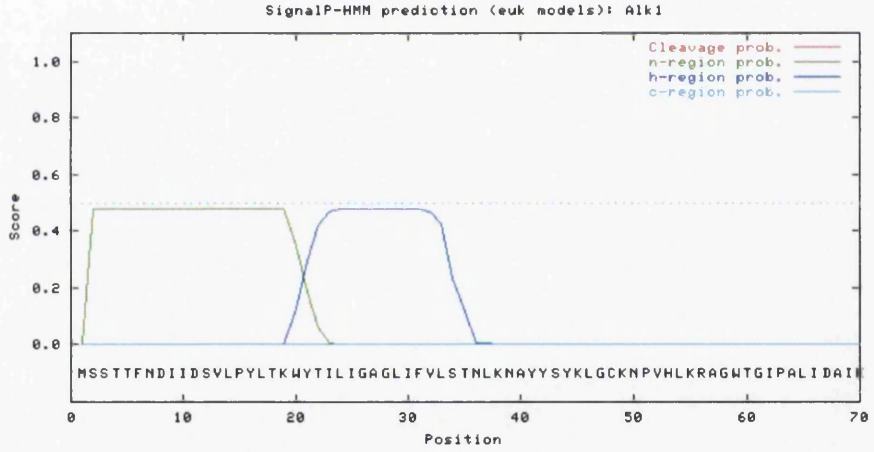


Figure 4.8 – TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) results showing the potential Alk1 membrane anchor region. This plot indicates that the membrane anchor region starts at residue 13 and ends at 35 with the cytosolic region starting at residue 36.



data

```
>Alk1                    length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    27        0.085  0.32    NO
  max. Y    35        0.190  0.33    NO
  max. S    27        0.936  0.87    YES
  mean S     1-34     0.418  0.48    NO
       D     1-34     0.304  0.43    NO
# Most likely cleavage site between pos. 34 and 35: VLS-TN
```



data

```
>Alk1
Prediction: Non-secretory protein
Signal peptide probability: 0.002
Signal anchor probability: 0.477
Max cleavage site probability: 0.001 between pos. 34 and 35
```

Figure 4.9 – SignalP (http://www.cbs.dtu.dk/services/SignalP/) results showing the signal peptide cleavage site of Alk1. The plot indicates that the signal peptide cleavage site of Alk1 is at residue 34.

```
Length:  526
N-terminus:  IN
Number of transmembrane helices: 1
Transmembrane helices: 12-33

Total entropy of the model:  17.0231
Entropy of the best path:  17.0239

The best path:

     seq  MASTTFNDII DSVLPYLTKW YTILIGAGLI FVLSTNLKNA YYSYKLGCKN    50
     pred Iiiiiiiiii iHHHHHHHHH HHHHHHHHHH HHHooooooo ooooooooOO

     seq  PVHLKRAGWT GIPALIDAIK AKNSGKLAEF GGDTFDELNS KNFYLNVAGV   100
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  LKIIFTIEPE NIKAVLATQF NDFSLGTRHA HFYPLLGDGI FTLDGEGWKH   150
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  SRAMLRPQFA REQIAHVKSL EPHIQILAKQ IKLNKGETFD LQELFFRFTV   200
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  DTATEFLFGE SVHSLYDDKL GIPPPNDIPG RDNFATAFIT SQHYLATRSY   250
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  SQTFYWLTNN KEFRDCNAKV HHLAKYFVDK ALNSTEKEIE EKSQGGYVFL   300
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  YELVKQTKNP KVLQDQLLNI MVAGRDTTAG LLSFAMFELA RNPKVWNKLR   350
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  EEIEANFGVG EEARVEDITF ENLKKCEYLK AILNETLRLY PSVPINFRTS   400
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  TKDTTLPTGG GPNGTSPIFV PKGSTVAYTV FKTHRLEEYY GKDSHDFRPD   450
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  RWFEPSTKKL GWAYLPFNGG PRICLGQQFA LTEASYVISR LAQLFPVLES   500
     pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

     seq  RDEQYPPNKC IHLTMNHDEG VFVSMN  526
     pred OOOOOOOOOO OOOOOOOOOO OOOOOO
```

Figure 4.10 – HMMTOP (http://www.enzim.hu/hmmtop/) results for Alk1 showing the potential membrane anchor region of the protein. These results predict the membrane anchor region of Alk1 ends at residue 33.

4.3.2 Cloning of *ALK1*, its Truncations, *CaCPR* and *CYP52A3*

ALK1 was amplified from *C. albicans* genomic DNA (optimal conditions were reaction 7 (see table 4.3b) and thermal cycling conditions #3 (see table 4.3a)). *CaCPR* was also amplified from *C. albicans* and *CYP52A3* from *C. maltosa* genomic

DNA by PCR. The *ALK1* truncations: *Δ33:ALK1*, *Δ69:ALK1*, *modifiedΔ50:ALK1* and *modifiedΔ44:ALK1*, were all amplified by PCR from pCWori^{+}:Alk1L456S plasmid DNA. The genes of interest were visualised on a 1% (w/v) agarose gel and verified by comparing the expected size of that gene (see table 4.2) (see figure 4.11).

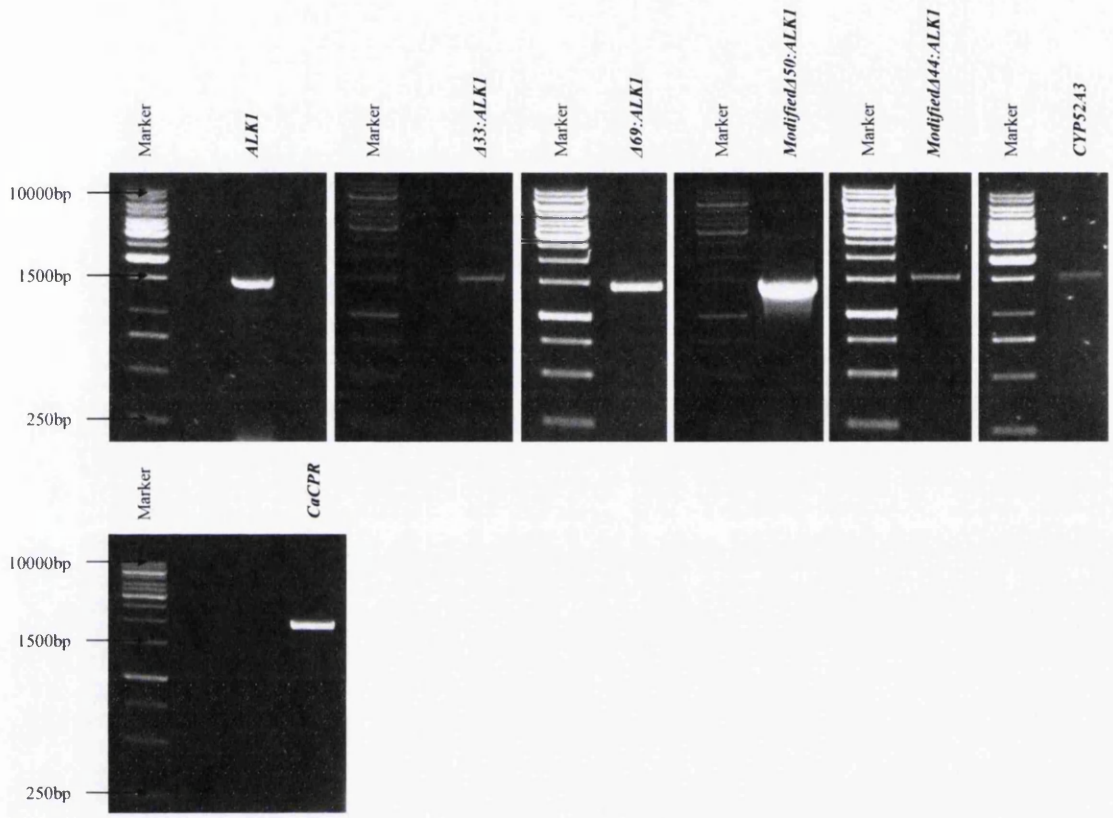


Figure 4.11 – 1% (w/v) agarose gel showing gene fragments following PCR.

Each amplicon was isolated from the gel and the DNA was extracted before ligation into pGEM-T easy and transformation into *E. coli* XL-1 cells. Colonies were screened for the presence of each gene using a *Not* I restriction enzyme digest, which was visualised on a 1% (w/v) agarose gel. The presence of each gene was verified by DNA sequencing.

A *Nde* I/*Hin*d III double digest was used to release each gene from pGEM-T easy and prepare the genes with the appropriate sticky ends for ligation into pCWori^{+}. The genes were visualised on a 1% (w/v) agarose gel. Each band was isolated from the gel and the recovered DNA of each gene, plus *Δ65:ALK1* and *ALK1L456S*, was

then ligated into pCWori^{+}. The ligations were then transformed into DH5α cells. Colonies were screened for each gene using *Nde* I/*Hin*d III digest and visualised on a 1% (w/v) gel. The presence of each gene was verified by sequencing.

4.3.3 Protein Expression of Alk1, Alk1L456S, its Truncations, CaCPR and CYP52A3 Using *E. coli* as a Host Organism

Methods described in section 4.2.3.7.1 were used to express the proteins of Alk1, its truncations and CYP52A3. The Arase *et al.*, 2006 method for the isolation of protein (section 4.2.3.7.1.1) for each gene resulted in the production in a peak at 420nm (see figure 4.12) in the presence of carbon monoxide and sodium dithionite. This is characteristic of the inactive form of cytochrome P450, P420.

Expressing the proteins of the Alk1 truncations using the isolation of membranes and cytosol method (section 4.2.3.7.1.2) also resulted in the production of inactive cytochrome P420 in both the membrane and cytosol fractions.

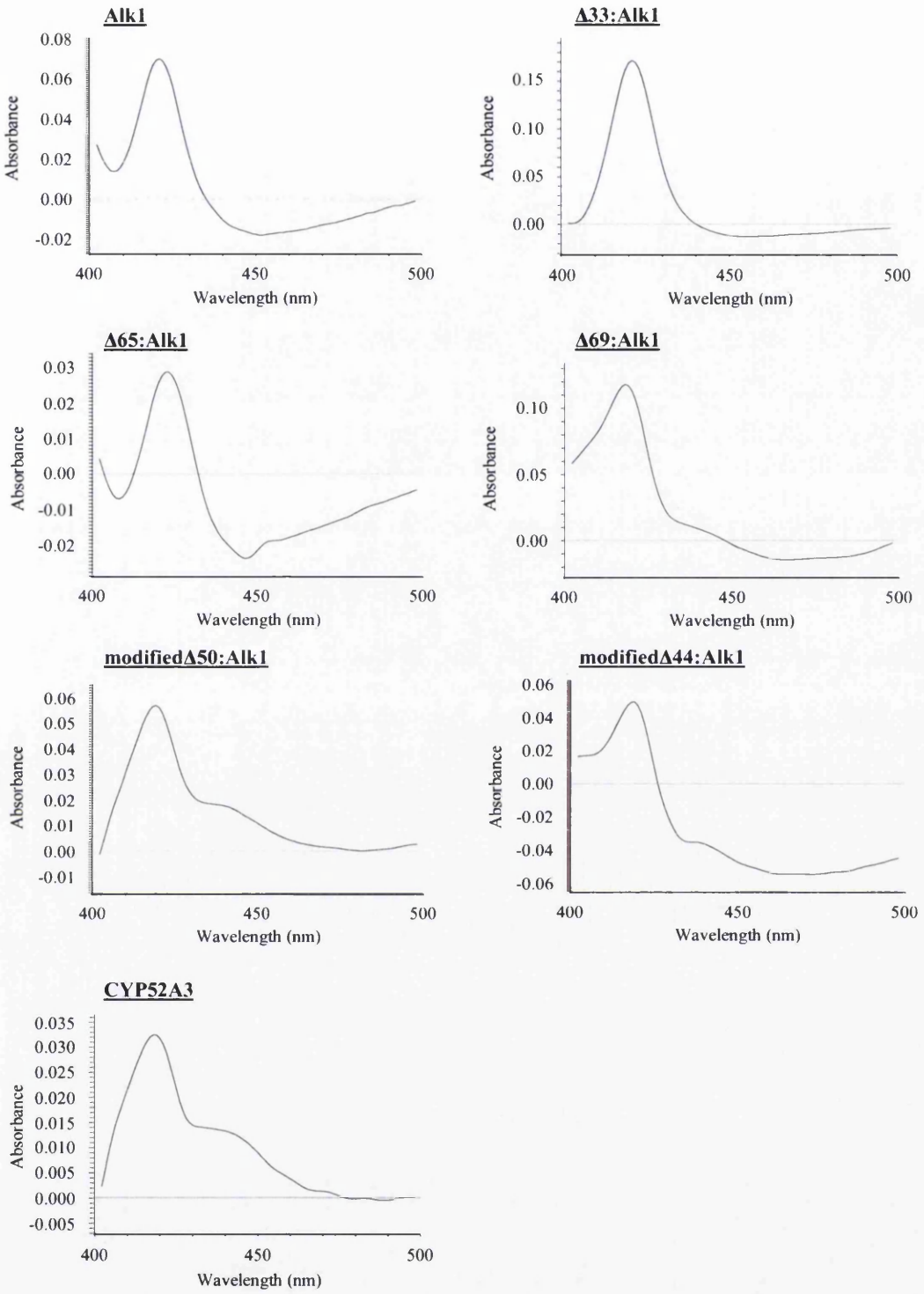


Figure 4.12 – Difference spectra of the reduced/CO-bound forms of Alk1, its truncations and CYP52A3. The spectra produced show peaks at 420nm, which is indicative of the inactive form of cytochrome P450, cytochrome P420.

Using methods described earlier in table 4.5 (section 4.2.3.7.1) it was demonstrated that protein expression of Alk1L456S yielded a peak at 447nm in the presence of carbon monoxide and sodium dithionite, indicating that the protein expressed was a member of the cytochrome P450 superfamily and produced in a correctly folded form (see figure 4.13).

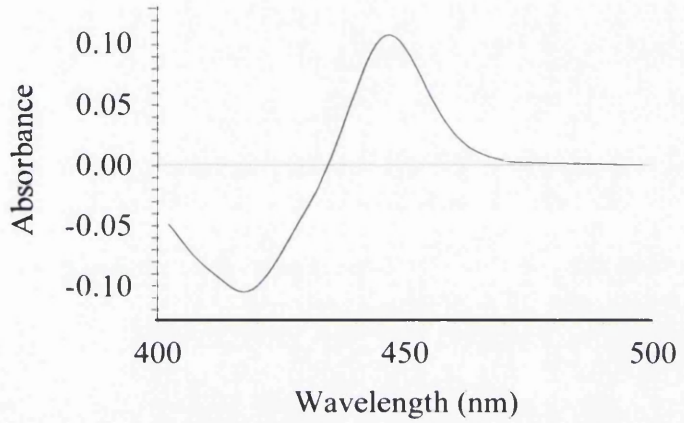


Figure 4.13 – Difference spectrum of the reduced/CO-bound form of Alk1L456S showing a 447nm peak. A peak at 447nm indicates cytochrome P450 protein.

Table 4.10 shows the protein yields obtained from the protein expression experiments undertaken. Increasing the growth temperature from 30°C to 37°C produced almost 2.5 fold more protein.

Experiment	Growth	Expression		Results
1	230rpm 30°C 7 hours	Ala 170mg IPTG 240mg	190rpm 20°C 48 hours	80nmoles/L
2	230rpm 30°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40.5 hours	54nmoles/L
3	230rpm 37°C 7.5 hours	Ala 170mg IPTG 240mg	190rpm 20°C 40 hours	193nmoles/L

Table 4.10 – Protein expression of Alk1L456S. The table shows the changes made to the growth and expression conditions during optimisation and the protein yield obtained.

The Alk1L456S protein was purified using Ni^{2+}-NTA agarose. Initially when the protein fractions were run out on a SDS-PAGE gel, it was found that there were still other proteins in the sample. The Alk1L456S protein was further purified using Ni^{2+}-NTA agarose, but the resin was washed an additional time using 1 column volume of wash buffer containing 10mM of L-histidine, prior to the elution of the protein (see section 4.2.3.9). This extra wash was used to remove the unwanted proteins that were non-specifically bound to the Ni^{2+}-NTA agarose (see figure 4.14) (Wester *et al.*, 2002). The concentration of Alk1L456S protein following purification was 25µM.

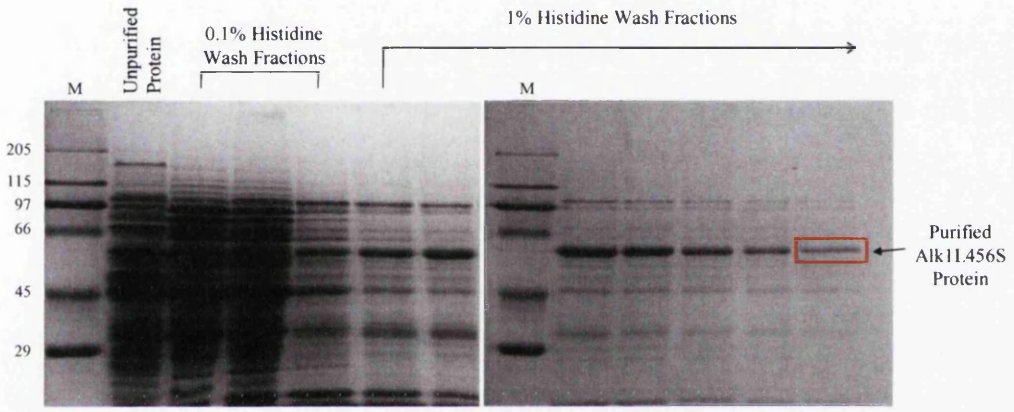


Figure 4.14 – SDS-PAGE gel showing Alk1L456S protein following optimised Ni^{2+}-NTA agarose purification. 10mM histidine step incorporated into the purification method.

CaCPR protein was expressed in *E. coli* using the method in table 4.8 (section 4.2.3.7). Following the solubilisation of the membranes and protein purification (figure 4.15), an absolute spectra of the protein showed peaks at 453nm and 382nm indicating the presence of CaCPR (figure 4.16). The concentration was 19µM.

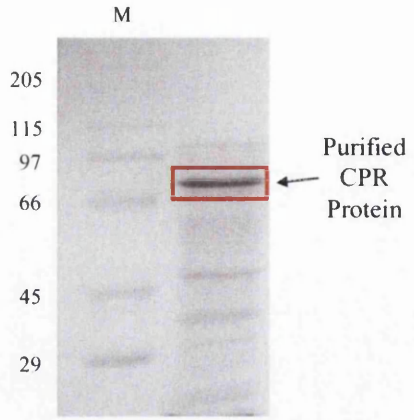


Figure 4.15 – SDS-PAGE gel showing CaCPR following Ni^{2+}-NTA agarose purification.

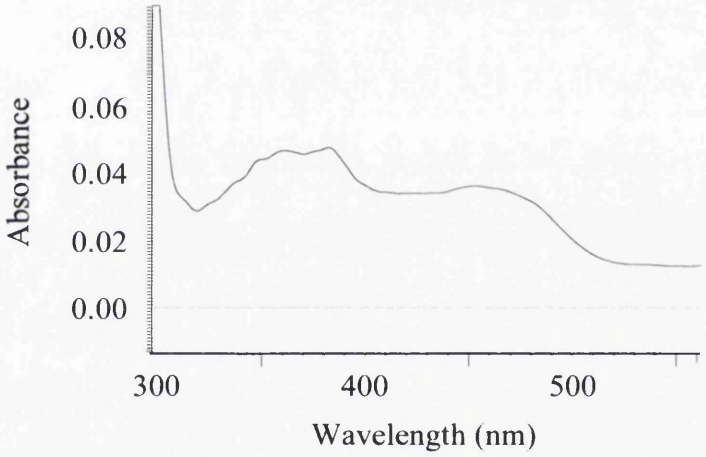


Figure 4.16 – Absorption spectrum of CaCPR, showing peaks at 453nm and 382nm indicating the presence of CPR protein.

4.3.4 Protein Expression of CYP52A3 yeast Using *S. cerevisiae* as the Host Organism

Protein expression of CYP52A3 using *E. coli* as the host organism failed to result in the production of cytochrome P450 protein. Therefore, CYP52A3_yeast was expressed in AH22 cells. Growth in media containing glucose as the initial carbon source before addition of galactose failed to produce microsomes containing cytochrome P450 protein. The expression of CYP52A3_yeast in media containing raffinose (induced with galactose) yielded a peak at 450nm indicating the presence of

cytochrome P450 in the microsomes (figure 4.17). The yield of cytochrome P450 protein was 38nmoles/L.

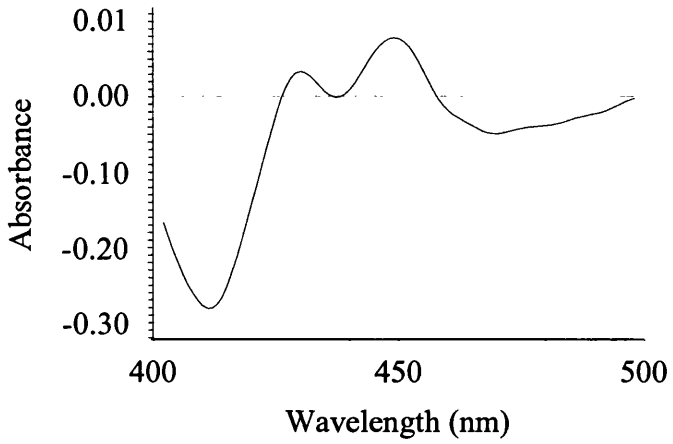


Figure 4.17 – Difference spectrum of the reduced/CO-bound form of CYP52A3_yeast. A peak at 449nm is an indication of the presence of cytochrome P450 protein in the microsome preparations.

4.4 Discussion

4.4.1 Expression of Alk1

All genes were designed to be expressed as protein in *E. coli* because this host organism produces large quantities of protein and is relatively quick and easy to use. However, the expression of Alk1 in *E. coli* resulted in a reduced carbon monoxide difference peak at 420nm rather than a peak at 450nm (which is characteristic of cytochrome P450 (Omura & Sato, 1964)). Initially, this was thought to be due to the formation of insoluble aggregates called inclusion bodies (Sorenson & Mortensen, 2005), which can form during protein expression in *E. coli* (Verma *et al.*, 1998). Inclusion bodies are misfolded proteins, which are formed when cells are open to environmental stresses, such as extremes in temperature, and are generally inactive (Sorenson & Mortenson, 2005; Arié *et al.*, 2006). These inclusion bodies could form if the temperatures used for the protein expression protocol were too harsh or were held for too long. One method to prevent the formation of these protein aggregates is to alter the temperature and duration of expression (Ventura & Villaverde, 2006). However, changes in the temperature and duration of both the growth and expression

phases of Alk1 protein expression (see table 4.4) continued to result in a peak at 420nm, which corresponds to the inactive form of cytochrome P450, P420.

Other approaches that can be employed to prevent the formation of inclusion bodies include the use of chaperones (to aid with the folding of the protein) and alterations to the protein sequence (as altering the amino acids used can alter the integrity/folding of the structure) (Ventura & Villaverde, 2006). With these approaches in mind, the translated nucleotide and protein sequences of Alk1 were aligned for comparison (see figure 4.18). At residue 456, it was shown that the nucleotide sequence (CUG) encodes for the amino acid leucine, but the protein sequence shows a serine. This is because in *C. albicans* there is an alteration in the universal genetic code in which the triplet CUG encodes for serine instead of the expected leucine (Santos & Tuite, 1995). A deviation in the genetic code such as this is not exclusive to *Candida* spp. as other species also show alterations.

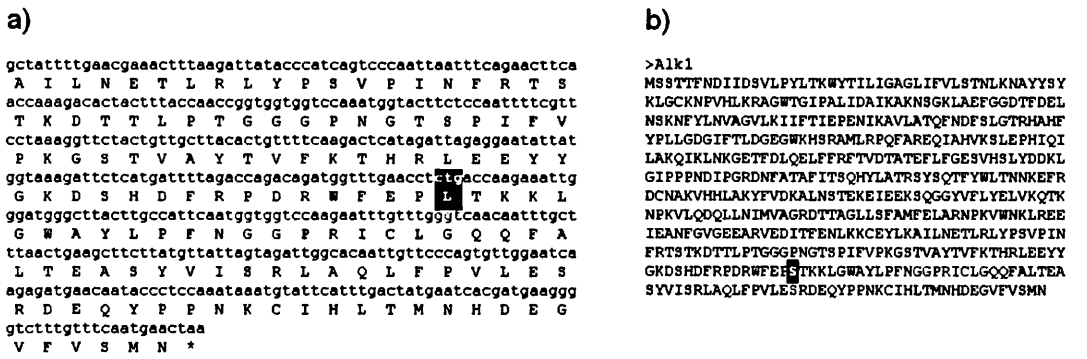


Figure 4.18 - Comparison of the nucleotide and protein sequences of Alk1. a) C-terminal *ALK1* nucleotide sequence with 1 letter amino acid translation (beginning at nucleotide position 1141); b) Alk1 protein sequence. The haem binding domain is highlighted in red. Amino acid differences highlighted by a black box.

The first deviation in the universal genetic code was identified in 1979 in vertebrate (human) mitochondria (Barrell *et al.*, 1979). AUA was shown to encode for the start codon methionine rather than isoleucine and UGA for tryptophan instead of a stop codon as in the universal genetic code (Barrell *et al.*, 1979). This revelation seemed to cast doubt on Crick's "Frozen Accident Theory," which suggests that any changes/alterations in the code could cause potentially lethal mutations in the

organism (Crick, 1968). However, as this codon deviation was in a structure within the organism/cell, it was thought that any possible changes would remain within the organelle and not affect the cell as a whole, i.e. not lead to cell death. This theory was shattered in 1985 when the first organism was shown to have codon deviations within its genetic code (Yamao *et al*, 1985). This was *Mycoplasma capricolum*, a goat pathogen, which exhibited a deviation in the translation of the UGA codon from a stop codon to a tryptophan codon (Yamao *et al.*, 1985). Since then the genetic codes of other organisms have been shown to contain codon deviations, including *Euplotes* spp. (stop codon UGA reassigned as cysteine) (Moura *et al.*, 2010) and *Candida* spp. (Ohama *et al.*, 1993; Pesole *et al.*, 1995).

Several *Candida* spp. have been shown to contain a deviation in the CUG codon whereby it encodes for serine rather than leucine (as in the "universal" genetic code). This deviation was first identified in *C. cylindracea* (Kawaguchi *et al.*, 1989; Yokogawa *et al.*, 1992) before being identified in other species in the *Candida* genus, such as *C. maltosa* (Ohama *et al.*, 1993; Pesole *et al.*, 1995; Sugiyama *et al.*, 1995), *C. albicans*, *C. tropicalis* and *C. parapsilosis* (Ohama *et al.*, 1993; Pesole *et al.*, 1995). Such deviations are thought to arise due to the loss of $tRNA^{Leu}CAG$ corresponding to CUG along with the codon itself (Osawa & Jukes, 1989). This could possibly be due to AT pressure, which can cause CUG to be converted to other codons encoding leucine (e.g. CUA, UUG and UUA) (Jukes & Osawa, 1996). This means neutral changes occur ensuring protein sequences remain unchanged and the codon is completely lost from the organism's genome (Jukes & Osawa, 1996).

Initially, following the loss of CUG and its tRNA, the CUG codon may act as a nonsense (stop) codon, thereby terminating translation (Osawa & Jukes, 1989; Jukes & Osawa, 1996). This indicates there is a transitional period between the loss of the codon and its recapture. During evolution, the codon is recaptured by a newly developed tRNA with the correct anticodon (Osawa & Jukes, 1989). In the case of *Candida*, this is $tRNA^{Ser}CAG$. These recaptured CUG codons are then reincorporated into genes via mutations (Jukes & Osawa, 1996). This process is known as the "Codon Reassignment (Codon Capture) Theory" (Osawa & Jukes, 1989).

The deviations in the "universal" genetic code such as the one described above may have a detrimental effect on active gene and protein expression. Leucine is a non-polar, hydrophobic amino acid, whereas serine is polar and hydrophilic. Such a difference in the properties of the amino acids could cause an alteration in the way the protein folds, leading to inactive expression. An example of the affect of CUG on expression was demonstrated by Sugiyama *et al.* (1995) using *URA3* from *S. cerevisiae*. This gene contains one CUG codon, which is usually translated as leucine in *S. cerevisiae* (Sugiyama *et al.*, 1995). Upon transformation into *C. maltosa* the gene was found to be inactive i.e. it would not complement (Sugiyama *et al.*, 1995). Following mutation of the codon to CTC, which codes for leucine in both *S. cerevisiae* and *C. maltosa*, the gene was found to be active (Sugiyama *et al.*, 1995). Not only does this show that *C. maltosa* encodes CUG as serine, but it also suggests that genes containing CUG in *Candida* need to be altered to a different serine codon for active expression in other organisms. Inactive expression has also been shown to occur within a number of cytochromes P450 from *Candida*. They include: *CYP52A4* (residues 99 and 389) (Scheller *et al.*, 1996; Zimmer *et al.*, 1996), *CYP52A5* (residue 499) and *CYP52A9* (position 249) from *C. maltosa* (Zimmer *et al.*, 1996), *CYP52A17* (residues 261 and 490) and the reductase, *NCPA* from *C. tropicalis* (Eschenfeldt *et al.*, 2003).

In *ALK1*, the CUG codon can be found in close proximity to the haem-binding site (see figure 4.18) suggesting any changes in the protein structure that may occur may affect access to or incorporation of the haem. To rectify this problem in *ALK1*, the gene was synthesised to express UCU (which encodes for serine) in place of the CUG codon. UCU was used because *ALK1* was to be expressed in *E.coli* and this species has a bias towards this particular codon, therefore, giving it greater efficiency when translating (Sharp & Li, 1987).

Expression of Alk1L456S resulted in a peak at 450nm, which is indicative of correctly folded cytochrome P450 (Omura & Sato, 1964). This showed that the problems with protein expression, which occurred with the native *ALK1* gene were due to the amino acid change associated with CUG in *C. albicans*. With this in mind, the sequences of the other (potential) cytochrome P450 genes identified in *C. albicans* (chapter 3) were aligned to check for the codon CUG and the translated

amino acid. Of the remaining cytochrome P450 genes, three did not contain this leucine/serine alteration. These genes are *ALK2*, *ALK3* and *TRI4*. *CYP61* (residue 43), *ALK8* (residue 40) and *DIT2* (residue 122) all contained single changes, whereas multiple alterations were identified in *PAH1* (four – residues 3, 391, 392 and 565) and *ALK6* (three – residues 163, 206 and 491). Thus suggesting these cytochromes P450 may require replacement of the CUG codon with another coding for serine if active protein was not expressed using the native sequence.

After altering CUG to another serine-encoding codon and modifying the N-terminus for optimised expression in the synthesised gene, *ALK1L456S* resulted in the production of 193nmoles/L of protein. Similar modifications that were made to *ALK1L456S* (i.e. alanine at the second position and the first six codons made AT rich) were made to the human *CYP17* sequence. This resulted in 40nmoles/L of protein being expressed (Waterman *et al.*, 1995), which is 5 fold less than Alk1L456S. However, by using the exact N-terminal sequence of modified bovine *CYP17A1* (ATG GCT CTG TTA TTA GCA GTT TTT (Barnes *et al.*, 1991)), 400nmoles/L of CYP17 protein was expressed, which is 10 fold more than what was previously produced, but still almost half the yield of the modified bovine CYP17A1 protein (750nmoles/L (Barnes *et al.*, 1991)) (Waterman *et al.*, 1995). This difference in protein yields could be due to the fact the rest of the sequence also plays a part in the level of protein expression, not just that of the N-terminus alone (Waterman *et al.*, 1995). Other cytochromes P450 have also shown high levels of protein expression by using the modified bovine *CYP17A1* N-terminal sequence. These include rabbit CYP2C3 (400nmoles/L (Richardson *et al.*, 1993)) and human CYP1A2 (700nmoles/L (Fisher *et al.*, 1992) and 245nmoles/L (Sandhu *et al.*, 1994) – differences in yield are potentially due to differences in the *E. coli* strains used (JM109 and DH5α respectively)). This suggests that if the exact bovine *CYP17A1* N-terminal sequence was used in *ALK1L456S*, the level of protein expressed may have been increased.

4.4.2 Expression of Alk1 Truncations

As alluded to in section 4.4.1 modifications can occur in the N-terminus without any effect on the activity of the cytochrome P450 protein, suggesting truncated forms of membrane-bound enzymes can be produced without any detrimental effect on

function. Versions of Alk1 missing various parts of the N-terminus were designed for heterologous expression in *E. coli*. They were designed using the web-based programs: TMHMM, SignalP and HMMTOP. The results of these programs predicted that the membrane anchor region terminated between residues 33 and 36. The decision was made to delete the first 33 amino acids of the Alk1 sequence to produce the truncation Δ33:Alk1. Expression of this protein in *E. coli* resulted in the production of inactive cytochrome P450 protein. This same method was used for the production of a truncated version of CYP2E1 from rabbit liver (Larson *et al.*, 1991). Truncation following the first hydrophobic region of CYP2E1 resulted in the production of active cytochrome P450 protein, which was still membrane-bound (i.e. the protein was found bound to the membrane of the cell rather than soluble in the cytosol) (Barnes *et al.*, 1991; Larson *et al.*, 1991). However, this protein contains two transmembrane loops (Larson *et al.*, 1991) and truncation following this second region may have resulted in a soluble protein expressed in the cytosol (Pernecky *et al.*, 1993; von Wachenfeldt *et al.*, 1997). Another cytochrome P450 with a second hydrophobic membrane anchor region is CYP52A3 (see figure 4.19) (Menzel *et al.*, 1996). Therefore, the second and third truncations were based on the topology of CYP52A3 (figure 4.19).

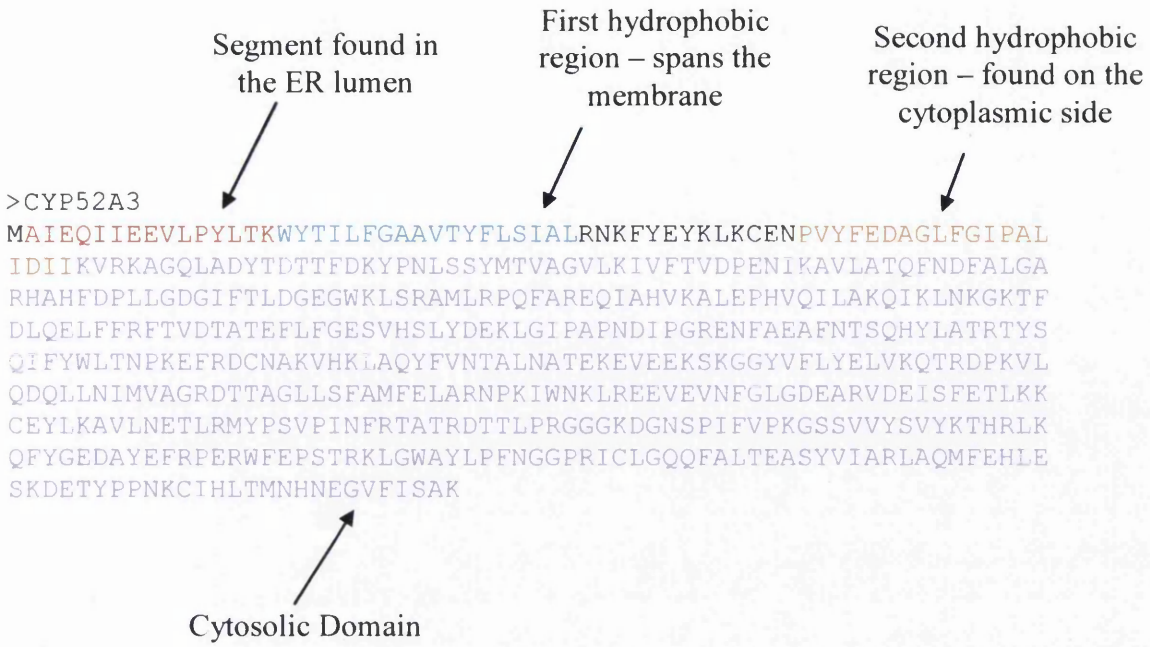


Figure 4.19 – Membrane topology of CYP52A3 (Menzel *et al.*, 1996). Highlighted in red is the portion of gene found in the lumen of the endoplasmic reticulum. The first hydrophobic region is highlighted in blue and the second is in brown. The purple sequence is the cytosolic domain.

Δ69:Alk1 was designed to begin at the end of the second predicted hydrophobic region (and prior to the hydrophilic lysine residue), thus deleting the first 69 residues of the sequence. This region of Alk1 was deleted because an alignment of the CYP52A3 and Alk1 protein sequences showed similarity between the second predicted membrane anchor region of CYP52A3 and Alk1 (figure 4.20).

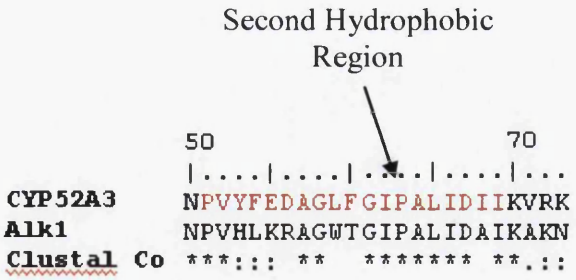


Figure 4.20 – Sequence alignment of CYP52A3 and Alk1. Similarity shown between the second hydrophobic region of CYP52A3 (which is highlighted in red) and Alk1.

Δ65:Alk1 was truncated prior to the end of the membrane anchor region (highlighted in figure 4.20). In CYP52A3, this region was used to create a site-specific region for factor Xa digestion in the production of an active soluble CYP52A3 protein (Scheller *et al.*, 1994). In Alk1, this region was used to create the truncation Δ65:Alk1 as adding four hydrophobic residues to the beginning of the Δ69:Alk1 sequence may act to stabilise the Alk1 protein during protein expression.

However, these truncations resulted in the production of inactive proteins. The production of cytochrome P420 truncated protein with this and the previous method was not due to the CUG problem that plagued the native *ALK1* gene expression as the truncations were amplified from the *ALK1L456S* synthesised sequence or the sequence of the truncation was synthesised. Therefore, the truncated genes contained the required alteration (TCT for CTG (as explained in section 4.4.1)).

The reason for the expression of inactive protein could be due to the instability of the folded structure. A truncated version of the bovine 17α-hydroxylase, CYP17, missing its N-terminal anchor, was shown to be expressed in its inactive form (Clark & Waterman, 1991; Sagara *et al.*, 1993). Insertion of an anchor sequence into this protein resulted in the production of correctly folded, active protein (Sagara *et al.*, 1993). Thus suggesting the N-terminus anchor region is required for correct folding and, therefore, stability of the protein, which is lacking in the truncated versions (Clark & Waterman, 1992; Sagara *et al.*, 1993). This problem can be overcome by designing truncations with regard to the proline-rich regions of the N-terminus as in CYP2C3 (Richardson *et al.*, 1993; von Wachenfeldt *et al.*, 1997) and CYP2C5 (Cosme & Johnson, 2000; Wester *et al.*, 2002). Proline residues are required for the correct folding of the protein, as they are important in the formation of turns and the disruption/formation of α-helices and β-sheets in secondary structure (Kusano *et al.*, 2001; Kusano *et al.*, 2001b). Rabbit CYP2C3 was shown to be successfully expressed in its truncated form using this method of introducing the truncation after the proline rich region following the hydrophobic membrane anchor (von Wachenfeldt *et al.*, 1997). This version of CYP2C3 was associated with the membrane, but could be found in the cytosol following treatment with a high salt buffer (von Wachenfeldt *et al.*, 1997) suggesting the truncated protein produced was not membrane-bound (and did not require detergents for its release). Similar results

were seen with CYP2C5 (von Wachenfeldt *et al.*, 1997). Due to this success, two truncations were designed for Alk1 using this method (although Alk1 does not contain a defined proline-rich region within its protein sequence). ModifiedΔ50:Alk1 was designed using the same N-terminus sequence as CYP2C3. This sequence MAKKTSSKGKL has been shown to maintain folding stability without affecting the cytochrome P450 function and was optimised for the expression of (full-length) CYP2C3 in *E. coli* (Li & Chiang, 1991; Karam & Chiang, 1994). ModifiedΔ44:Alk1 was designed using the CYP2C3 N-terminal leader sequence and more of the Alk1 protein sequence prior to the proline rich region to promote correct folding.

Both modifiedΔ50:Alk1 and modifiedΔ44:Alk1 resulted in inactive cytochrome P420 following expression. Unlike CYP2C3, CYP2C5 and other microsomal P450s (e.g. CYP2D6), Alk1 does not contain a defined proline-rich region, PPGPXPXP (see figure 4.21) (von Wachenfeldt *et al.*, 1997; Kusano *et al.*, 2001b). However, this method was used to see if truncating the sequence near a proline may result in the production of an active, truncated version of Alk1. Therefore, the proline used in the design of the Alk1 truncations was chosen by aligning the Alk1 sequence with the amino acid sequences for CYP2C3 and CYP2C5 (figure 4.21). However, this resulted in the production of inactive cytochrome P420 protein. This indicates that the proline identified may not have been the important residue for folding stability and the truncations of Alk1 may require re-designing at other prolines to produce more soluble, active cytochrome P450 protein.

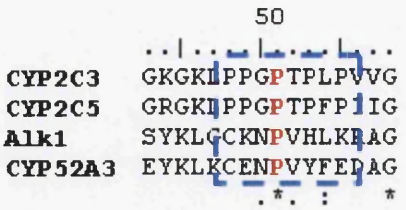


Figure 4.21 – Sequence alignment showing the proline-rich region. Sequence alignment of rabbit CYP2C3, rabbit CYP2C5, CYP52A3 and Alk1. The dashed blue line indicates the proline-rich region of the sequence. Highlighted in red are the proline residues conserved in all four sequences.

Active soluble CYP52A3 protein was produced using sequence-specific proteolysis at a specially designed cleavage site (Scheller *et al.*, 1994). The full-length membrane-bound CYP52A3 protein was expressed in *S. cerevisiae* containing the recognition sequence Ile-Glu-Gly-Arg for factor Xa (Scheller *et al.*, 1994). Upon treatment of the protein with this protease, the cytosolic domain was released from the N-terminus producing a soluble CYP52A3 protein with spectral properties similar to the full-length version, with only a decrease seen in the catalytic activity of the truncation (Scheller *et al.*, 1994). A similar method was used with human CYP1A2 (Dong *et al.*, 1996). In this method, a thrombin-sensitive region was incorporated into the full-length sequence to allow release of the cytosolic domain from the N-terminus upon treatment with thrombin (Dong *et al.*, 1996), thus producing an active, soluble cytochrome P450 protein. This may be an alternative method that could be used to produce active soluble Alk1 protein.

The production of active cytosolic cytochromes P450 is difficult as it depends upon the hydrophobic profile of the protein, the method of design used and the cytochrome P450 protein itself. Many attempts at producing a variety of different cytochromes P450 (e.g. CYP2B6 from human livers) by different laboratories have resulted in either the formation of inactive cytochrome P450 or an active truncated version of the full-length protein, which is still membrane-bound (Scott *et al.*, 2001). Therefore, the difficulties in producing an active truncated protein as seen in this study are not exclusive to Alk1.

4.4.3 Expression of CYP52A3

Using *E. coli* as a system for the expression of CYP52A3 resulted in the production of inactive cytochrome P420 protein. As Alk1 was also unable to express in its native form in this organism a similar approach (as discussed in section was 4.4.1) was implemented to try and produce CYP52A3 in its active form. Alterations in the temperature and duration of both the growth and expression phases of CYP52A3 protein expression continued to result in the production of cytochrome P420 protein, suggesting the formation of inclusion bodies was not the reason for the production of inactive CYP52A3 protein. The protein and translated nucleotide sequences for CYP52A3 were aligned to ascertain whether CYP52A3, like Alk1, had a deviation in the translation of CUG. The *CYP52A3* nucleotide sequence did not contain this

codon and there were no other differences identified between the two sequences. Therefore, the deviation in the "universal" genetic code, which was problematic in the protein expression of native Alk1, was not the reason for the production of inactive CYP52A3 protein.

Previously, CYP52A3 was expressed in its active cytochrome P450 form by Scheller and co-workers (1996) when *S. cerevisiae* was used as the host organism. Therefore, *S. cerevisiae* AH22 cells were used to express CYP52A3_yeast protein in this study. This resulted in the production of protein with a reduced carbon monoxide difference spectrum at 449nm, suggesting the presence of cytochrome P450 protein in the microsomes. This might suggest that CYP52A3 requires eukaryotic posttranslational modifications for it to be expressed in its active cytochrome P450 form or it could be other aspects of eukaryote biology, such as appropriate chaperones.

Despite the expression of proteins in *E. coli* resulting in high yields (compared to other systems described in section 4.1) and being relatively quick and easy, this expression system can result in the production of instable and inactive proteins. This can be due to the inability of *E. coli* to undertake posttranslational modifications on the protein (such as phosphorylation or glycosylation), which are required for the correct folding of the protein and for enzymatic activity (Gonzalez & Korzekwa, 1995). Yeast such as *S. cerevisiae*, are able to perform these posttranslational modifications and are, therefore, ideal host organisms for the expression of proteins that are unable to express in their active form in *E.coli*.

In conclusion, *ALK1L456S*, *CaCPR* and *CYP52A3_yeast* were all shown to produce a reduced carbon monoxide cytochrome P450 spectrum upon expression of the protein in either *E. coli* or *S. cerevisiae*. The Alk1L456S protein will be used in further studies to identify possible substrates of this protein through substrate binding and reconstitution assays. CYP52A3_yeast will be used as a control in these assays. CaCPR will be used in these reactions for the transfer of electrons from NADPH.

Chapter 5: Alk1L456S Binding and Metabolism Studies Using Alkanes and Fatty Acids as Substrates

5.1 Introduction

Cytochromes P450 produce a Soret peak at 450nm when carbon monoxide binds to the sodium dithionite reduced protein (Omura & Sato, 1964) and exhibit other spectral peaks in the native state. Cytochromes P450 also exhibit other spectra on binding with potential substrates and the spectral changes which occur can be classified into two types: type I and type II depending on the spectra produced (figure 5.1).

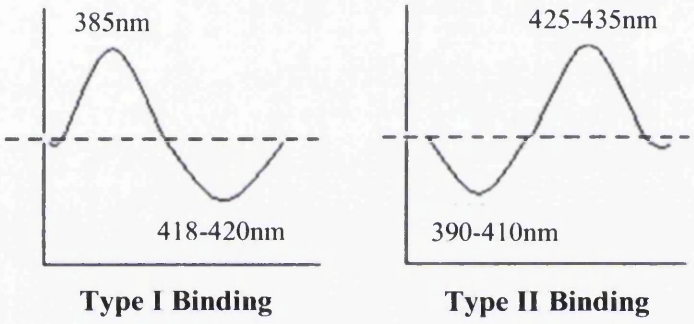


Figure 5.1 – Schematic overview of the two classes of binding spectra: type I and type II. Indicated on each spectrum is the wavelength at which the peak and trough occur, characterising the type of spectra shown.

In its substrate-free (resting) state, the cytochrome P450 haem iron is in its ferric low-spin form (figure 5.2). In this state, water acts as the sixth ligand and an absolute spectra of 416-418nm is produced (Yoshida, 1988). In the presence of substrate, the substrate coordinates with the haem iron at the sixth coordination position, but does not adhere as the sixth ligand. This prevents water accessing the haem causing a spin shift change from low- to high-spin (figure 5.2) (Yoshida, 1988). This shift is characteristic of type I binding and is represented as a peak at 385nm and a trough at 418-420nm (Schenkman, 1970; Schenkman *et al.*, 1972).

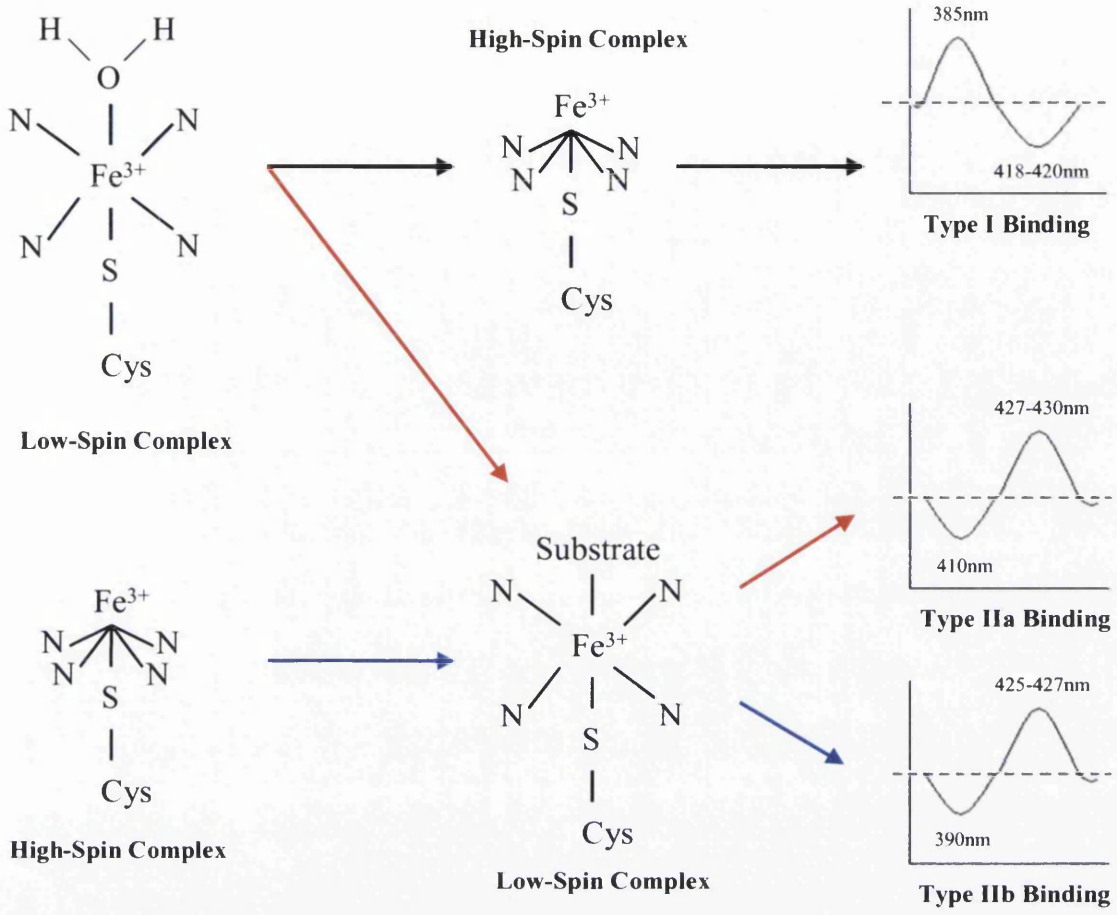


Figure 5.2 – Alterations in the spin-state of the cytochrome P450 haem iron (caused by substrate interaction) and the binding spectra produced. Type I binding is caused by a shift in the spin-state of the haem iron from low to high, which occurs when the substrate prevents the water molecule binding as the sixth ligand (black arrows). Type II binding occurs when the substrate coordinates with the haem iron acting as the sixth ligand. Type IIa binding spectra occurs when the haem iron remains in a low-spin state (red arrows). Type IIb binding spectra is produced when there is a shift in the spin-state from high to low (blue arrows).

However, if a substrate interacts with the haem iron acting as the sixth ligand, then the cytochrome P450 acquires an artifical hexa-coordinated low-spin configuration (figure 5.2). This results in the formation of type II binding spectra, of which there are two types: type IIa and type IIb. Type IIa binding occurs when the substrate-free form of the cytochrome P450 is in a high-spin state resulting in a high- to low-spin state shift when the substrate binds as the sixth ligand (figure 5.2) (Yoshida, 1988). This type of binding is characterised by a peak at 427-430nm and trough at 410nm (Schenkman, 1970; Schenkman *et al.*, 1972; Yoshida, 1988). Type IIb binding

occurs when the substrate-free form is already in a low-spin state and is represented as a peak at 425-427nm and a trough at 390nm (figure 5.2) (Schenkman, 1970; Schenkman *et al.*, 1972; Yoshida, 1988). Type II binding can be indicative of cytochrome P450 inhibition. However, some substrates containing nitrogen or sulphur atoms can bind as the sixth ligand producing weak type II binding spectra as these atoms can readily interact with the haem iron (Locuson *et al.*, 2007).

Binding studies are important to identify potential substrates of a cytochrome P450 enzyme. These substrates can then be used in *in vivo* and *in vitro* experiments to identify how the cytochrome P450 enzyme catalyses their metabolism.

Reconstitution assays are a common method to evaluate the *in vitro* activity of a purified cytochrome P450 protein on a substrate. To allow the cytochrome P450 monooxygenase cycle to progress and for metabolites to be produced redox partners need to be added to the assay (see sections 1.2.1 and 1.2.2). In this chapter, purified Alk1L456S protein was used to catalyse substrate metabolism and as this is a membrane-bound eukaryotic protein, it belongs to the class II group of cytochrome P450 proteins (see section 1.2.2.2). Cytochromes P450 belonging to this class require NADPH-cytochrome P450 reductase (CPR) for the transfer of electrons from the electron donor NADPH to the haem centre of the protein (Werck-Reichhart & Feyereisen, 2000; Paine *et al.*, 2005). CPR is a flavoprotein consisting of equal amounts of the cofactors ferredoxin reductase (FAD - flavin adenine dinucleotide) and flavodoxin (FMN - flavin mononucleotide), which are required for the transfer of electrons (He & Chen, 2005; van Bogaert *et al.*, 2007). The hydride ion of NADPH transfers electrons to the FAD domain of the CPR, as FAD has a lower oxidation potential than NADPH (van Bogaert *et al.*, 2007). A single electron is then transferred from FAD to the FMN domain, which is then transferred to the cytochrome P450 to reduce the haem centre (Groves, 2005; van Bogaert *et al.*, 2007). A second electron is then transferred to cleave oxygen for incorporation into the substrate (Groves, 2005). Therefore, these components need to be added to the reconstitution assay to aid Alk1L456S activity.

However, not all cytochromes P450 require CPR and NADPH for the transfer of electrons to the haem centre, so alternative components may need to be added to the

reconstitution assay for functional enzymatic activity to occur. For instance, CYP101D1 from *Novosphingobium aromaticivorans* belongs to the class I of cytochromes P450 and requires ferredoxin reductase and iron sulphur redoxin (as well as NADH) to catalyse the conversion of camphor to 5-exo-hydroxycamphor (Bell *et al.*, 2010). Alternatively, CYP102A1 from *Bacillus megaterium* is a naturally occurring cytochrome P450-CPR fusion protein (Warman *et al.*, 2005), which requires the addition of NADPH to the assay, but not an additional redox partner.

A regeneration system can also be added to ensure an adequate supply of the cofactor in the assay. For example, glucose-6-phosphate and glucose-6-phosphate dehydrogenase can be added to the assay to regenerate NADPH (Urlacher & Eiben, 2006). The transfer of electrons to the cytochrome P450 results in the reduction of NADPH to $NADP^+$, which can no longer be used to drive the cytochrome P450 monooxygenase cycle. $NADP^+$ can be regenerated in the presence of glucose-6-phosphate and glucose-6-phosphate dehydrogenase as the dehydrogenase catalyses the removal of hydrogen (dehydrogenation) from the glucose-6-phosphate molecule. The hydrogen released can react with $NADP^+$ to form NADPH, which can then be used to drive the reaction. The addition of a regeneration system is not necessary for cytochrome P450 activity to occur. However, NADPH is an expensive compound to purchase and a continuous supply to the assay would prove costly so the addition of a regenerating system is not only beneficial in regenerating consumed NADPH, but also reducing the overall cost of the reaction (Urlacher & Eiben, 2006).

A lipid component can also be added to the reconstitution assay mix if a membrane-bound cytochrome P450 protein is used. This is required for protein stability and to ensure the cytochrome P450 and CPR proteins are at the required orientation for electron transfer by ensuring the N-terminus of the protein is attached to the membrane and the C-terminus is facing the cytoplasm (Paine *et al.*, 2005; van Bogaert *et al.*, 2007). Phospholipids, dilaurylphosphaticholine (DLPC) and yeast microsomes have all been used as the lipid component in reconstitution assays (Guengerich *et al.*, 1991; Reed *et al.*, 2006). However, if yeast microsomes are used then purified cytochrome P450 protein does not need to be added to the assay as microsomes can be produced containing the protein of interest. These microsomes

also contain their own CPR and so additional reductases are not required to be added to the assay (Guengerich *et al.*, 1991). However, some cytochromes P450 are not able to utilise the yeast CPR for electron transfer in the reaction and, therefore, require the addition of CPR protein to the assay (Pompon *et al.*, 1996).

Each of the components described above can be added in different concentrations to the reconstitution assay mix to aid the enzymatic activity of the cytochrome P450 protein and to enhance product formation, thus making reconstitution assays a favourable method for evaluating cytochrome P450 activity. However, reconstitution assays require a quantifiable amount of cytochrome P450 protein for experiments to be undertaken. Expressed cytochrome P450 protein can be difficult to obtain (see chapter 4), therefore, *in vivo*, whole-cell experiments can be undertaken as alternatives to reconstitution assays. Not only do these experiments not require the addition of protein, they do not require the addition of expensive cofactors, such as NADPH, and they also contain additional proteins, which may be required for enzymatic activity and stability (Urlacher & Eiben, 2006). However, these experiments can be limited by poor substrate uptake and potential substrate toxicity as well as the possibility that the products formed could be subjected to further metabolism by other enzymes occurring naturally in the cell (Urlacher & Eiben, 2006). These limitations associated with *in vivo* experiments make *in vitro* reconstitution assays a more attractive method for the evaluation of cytochrome P450 activity.

The objective of this chapter is to attempt to deorphanise Alk1L456S using the purified protein produced in chapter 4 in binding studies and reconstitution assays. As Alk1L456S shares homology with CYP52A3 from *Candida maltosa*, a variety of alkanes and fatty acids will be used in binding studies to establish those hydrocarbons, which bind to Alk1L456S producing a type I binding spectra. These substrates will then be used in reconstitution assays to identify if and how Alk1L456S metabolises these hydrocarbons, potentially resulting in the production of diacids. CYP52A3 will be used as a control. This study will attempt to assign function to Alk1L456SS, but will also increase the understanding of proteins belonging to the CYP52 family of cytochromes P450. To date very few of these CYP52 enzymes have been characterised at the protein level despite 52 genes being

identified in a number of yeast/fungal strains (http://drnelson.uthsc.edu/CytochromeP450.html).

5.2 Materials and Methods

5.2.1 Buffer D

Based on the buffer used by Scheller *et al.*, 1996.

1M sodium citrate buffer, pH7.4, 30% (w/v) glycerol and 0.1% (w/v) *n*-dodecylmaltoside (Sigma-Aldrich). Stored at 4°C.

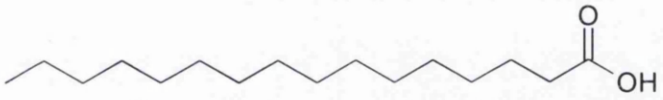
5.2.2 Standards

All standards (figure 5.3) were made up to a final concentration of 2mg/ml in chloroform. 5µl of each standard was evaporated to complete dryness in a SpeedVac, derivatised and run on the GCMS.

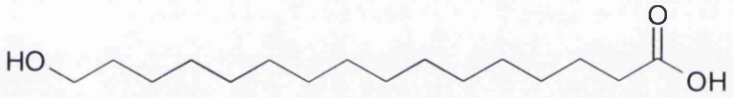
Hexadecane

Hexadecanol

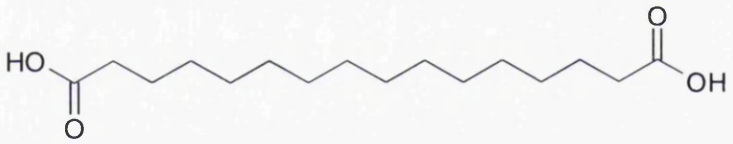
Palmitic Acid



16-Hydroxypalmitic Acid



Thapsic Acid



Palmitic Acid-16,16,16-d_3

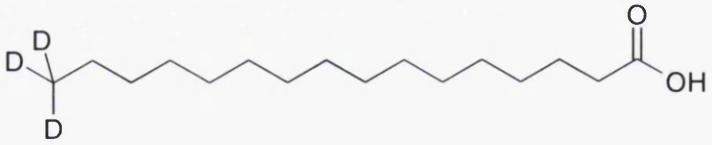


Figure 5.3 – Structures of hexadecane, hexadecanol, palmitic acid, 16-hydroxypalmitic acid, thapsic acid and palmitic acid-16,16,16-d_3. Hexadecanol (alcohol) is the hydroxylated version of hexadecane (alkane). Palmitic acid (fatty acid) is hydroxylated at the ω- (16^{th}) carbon to produce 16-hydroxypalmitic acid (fatty alcohol). Thapsic acid is a diacid. All structures are 16 carbons in length. Palmitic acid-16,16,16-d_3 is the deuterated form of palmitic acid. Three heavy hydrogen atoms (D) are attached to carbon 16 increasing the molecular weight from 256 to 259.

5.2.3 Minimal Salts Media

Based on the recipe of Dickinson & Wadforth, 1992. Tween 80 was added according to Tanaka *et al.*, 1977.

7g potassium dihydrogen phosphate, 2g sodium phosphate, 1.5g magnesium sulphate, 1.5g yeast extract, 0.1g calcium chloride, 0.008g ferric (iron) chloride, 0.0001g zinc sulphate and 15g diammonium tartrate were made to a final volume of 1 litre using distilled water. The solution was adjusted to pH6.5 and autoclaved. To this solution the following were added immediately before use: 2% (w/v) galactose, 0.1% (v/v) hexadecane and 0.05% (v/v) Tween 80. 5μg/ml histidine was added to all media.

5.2.4 Methods

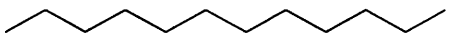
5.2.4.1 Substrate Binding

All binding studies of lipid to protein were undertaken using a Hitachi U-3310 UV-Vis scanning spectrophotometer.

5.2.4.1.1 Alkane Binding

The alkanes used for binding studies with purified Alk1L456S protein are shown in figure 5.4.

Dodecane



Hexadecane

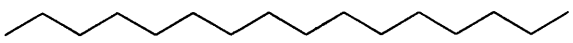


Figure 5.4 – Structural overview of dodecane (C12) and hexadecane (C16).

5.2.4.1.1.1 Titration of Alkane Against Purified Alk1L456S Protein

Alkanes were diluted in DMF (dimethylformamide) (5mg/ml stock solution) or ethanol (100mM stock solution) and titrated against 2μM Alk1L456S protein in a split cuvette. Depending on the solvent used to dissolve the alkanes, the same solvent (minus the hydrocarbon) was added to the protein in the reference cuvette. For example, if DMF was used to dissolve the alkane, then DMF would have been added to the reference cuvette. The following buffers were used for all spectral determinations: buffer D, 0.1M Tris HCl, pH8.1 containing 25% (w/v) glycerol and 0.1M potassium phosphate, pH7.4 containing 20% (w/v) glycerol. Spectra were measured between 350nm and 500nm.

5.2.4.1.1.2 Scheller *et al.*, 1996 Method

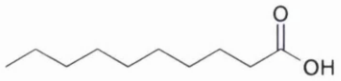
Based on the method of Scheller, *et al.*, 1996.

A 100mM stock solution of hexadecane was made up in ethanol. This stock was diluted to a 2mM solution in buffer D at 37°C. The solution was then sonicated in a sonic bath for 1 minute to mix. The 2mM solution was diluted further in buffer D to produce different concentrations of hexadecane. The concentrations used were: 7.8μM, 15.6μM, 62.5μM, 250μM and 1mM. The baseline was recorded using buffer D. Alk1L456S was added to the hexadecane to produce a final protein concentration of 1μM. The difference spectra were then recorded. This procedure was repeated for each hexadecane concentration.

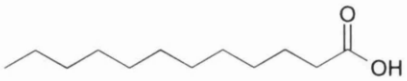
5.2.4.1.2 Fatty Acid Binding

A range of fatty acids (from C10:0 to C18:0 (see figure 5.5)) were dissolved in DMF for use in this study. These fatty acids were individually titrated against Ni^{2+}-NTA agarose purified Alk1L456S protein (2μM) in a split cuvette. A reference sample was also set-up in a second split cuvette with DMF added to the cytochrome P450 protein. The fatty acid (dissolved in DMF) was added to the buffer in the back compartment of the reference cuvette to prevent precipitation interfering with the spectra produced. Buffer containing 0.1M Tris HCl, pH8.1 and 25% (w/v) glycerol was used for all spectral determinations. Spectra were measured until saturation with the fatty acid was reached.

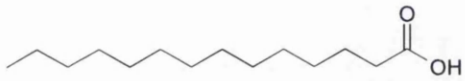
Capric Acid (C10:0)



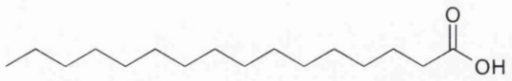
Lauric Acid (C12:0)



Myristic Acid (C14:0)



Palmitic Acid (C16:0)



Stearic Acid (C18:0)

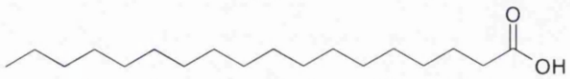


Figure 5.5 – Structural overview of the saturated fatty acids used in binding studies with purified Alk1L456S protein.

5.2.4.2 Reconstitution Assays

5.2.4.2.1 Reconstitution Assays Using Purified Alk1L456S Protein to Catalyse the Reaction

Micelles were created according to the method in section 2.2.11.1 using 10µl of 2mg/ml of palmitic acid dissolved in chloroform because this fatty acid was shown to have the greatest binding affinity of all the hydrocarbons used for binding studies in this chapter.

All assays were set-up according to the method described in section 2.2.11.2 with the addition of 2.5µM purified Alk1L456S protein and 10µM ScΔ33:CPR, unless otherwise stated. Negative controls were set-up without NADPH. 100µl of 0.2% (w/v) sodium bicarbonate was added instead to ensure the pH of the assay remained the same in both reactions.

To check whether buffer pH affected product formation, 100µl of 1M Tris-HCl, pH6.4 was substituted in the assay with the same buffer at pH6.6, pH6.8, pH7.0 or pH7.2. Negative controls were set-up with NADPH omitted.

Different CPRs were also added to the assays as some cytochromes P450 are unable to utilise different CPR for electron transfer during the monooxygenase cycle (Pompon *et al.*, 1996). ScΔ33:CPR was replaced with either 3.8μM CaCPR or 0.052μM human reductase (hCPR) (Sigma-Aldrich). Negative controls were also set-up with CPR omitted from the assay.

CPR titrations were set-up using different rations of CPR to Alk1L456S. ScΔ33:CPR was added to the assays at a ratio of 1:1 and in excess at 2:1, 4:1 and 10:1. The concentration of ScΔ33:CPR added was, therefore, 2.5μM, 5μM, 10μM and 25μM respectively. The concentration of Alk1L456S was reduced from 2.5μM to 0.5μM for CaCPR titrations. CaCPR was set-up in excess of Alk1L456S at ratios of 2.5:1, 5:1 and 10:1. Therefore, the concentration of CaCPR added were 1.25μM, 2.5μM and 5μM respectively.

CaCPR titrations were repeated increasing incubation at 37°C from 1hour to 18 hours.

Micelles were created using 10μl of 2mg/ml palmitic acid-16,16,16-d_3 dissolved in chloroform rather than palmitic acid. CaCPR titrations were repeated using these micelles. Assays were incubated at 37°C for 1 hour and 18 hours before extraction.

In all titrations, negative controls were set-up without CPR.

All assays were terminated on ice and extracted with 5ml of a chloroform:methanol (2:1, v/v) solution. Samples were evaporated to complete dryness in a SpeedVac prior to derivatisation (sections 5.2.4.3 and 2.2.12) and analysed on the GCMS (section 2.2.13.1).

All reactions were set-up in duplicate.

5.2.4.2.2 Reconstitution Assays Using Microsomal Preparations of CYP52A3 yeast to Catalyse the Reaction

2.2μM of CYP52A3 microsomal preparations were added to 10μl of 2mg/ml palmitic acid-16,16,16-d_3 (dissolved in ethanol), 100μl of 1M potassium phosphate, pH6.4, 50μl of 40mM glucose-6-phosphate and 50μl glucose-6-phosphate dehydrogenase. The solution was made up to a final volume of 900μl with distilled water and mixed thoroughly. The solution was incubated on ice for 30 minutes prior to pre-incubation at 30°C for 5 minutes. The reaction was initiated by the addition of 100μl of 10mM NADPH dissolved in 0.2% (w/v) sodium bicarbonate. The solution was mixed thoroughly before incubation at 30°C for 1 hour. Samples were terminated on ice, extracted with 5ml of chloroform:methanol (2:1, v/v) and evaporated to complete dryness in a SpeedVac. Samples were then derivatised (sections 5.2.4.3 and 2.2.12) and analysed on the GCMS (section 2.2.13.1).

3.8μM of CaCPR was added to some reactions to aid product formation.

Reactions were also set up with alternative substrate. Either 10μl of 10mg/ml palmitic acid-16,16,16-d_3 or hexadecane dissolved in DMF was added to the samples.

Negative controls were set up as above with either the substrate or NADPH removed. Microsomal preparations were also set up as blank controls. 2.2μM of CYP52A3 microsomes was diluted to a final volume of 1ml with 1M potassium phosphate, pH6.4 with the reconstitution system omitted. All controls were extracted and analysed as above.

All reactions were set-up in duplicate.

5.2.4.2.3 Metabolism of Hexadecane by *S. cerevisiae* Transformed with YEp51:Alk1L456S yeast Plasmid

Based on the methods of Ohkuma *et al.*, 1998; Kogure *et al.*, 2007 and Uppuluri & Chaffin, 2007. The extraction method was based on the protocols of Bligh & Dyer, 1959 and Iverson *et al.*, 2001.

10ml of YEPD was inoculated with one *S.* cerevisiae colony (containing either YEp51:Alk1L456S_yeast, YEp51:CYP52A3_yeast (positive control) or YEp51 only (negative control)) and incubated overnight in a shaker set at 30°C and 150rpm. Following overnight growth, the number of cells in the culture was counted using a light microscope and haemocytometer. The cells were washed in distilled water and added to 50ml of minimal salts media (containing 2% (w/v) galactose, 0.1% (v/v) hexadecane and 0.05% (v/v) Tween 80) at a final cell density of $1x10^5$ cells/ml. Cultures were grown at 30°C and 150rpm and samples were taken at regular intervals. Experiments were set-up in duplicate.

5ml of the culture was extracted from the growth sample at each interval and centrifuged for 3 minutes at 3000g. The supernatant was discarded and the cell pellet was resuspended in 1ml of distilled water. 3.75ml of a chloroform:methanol (1:2) mixture was added to the suspension followed by 1.25ml of chloroform and 1.25ml of distilled water. The mixture was thoroughly mixed on a vortex after each addition. The lower chloroform phase was extracted with a glass Pasteur pipette and evaporated to complete dryness in a SpeedVac. Samples were derivatised (section 5.2.4.3) and run on the GCMS (section 2.2.13.1).

5.2.4.3 Derivatisation of Samples

5.2.4.3.1 Base Catalysed Esterification

Based on method C from Basconcillo & McCarry, 2008.

1ml of 0.5M sodium methoxide in methanol (Sigma-Aldrich) was added to the dried sample and mixed on the vortex thoroughly. The mixture was then incubated in a waterbath for 15 minutes at 50°C. The reaction was stopped by adding 100µl of glacial acetic acid. This was followed by the addition of 1ml of hexane and 1ml of distilled water. The mixture was thoroughly mixed and the top phase was extracted with a glass Pasteur pipette. The hexane extraction step was repeated twice (2 x 1ml) and the extracted layers combined. To these combined layers 0.5g of sodium sulphate (anhydrous) (Sigma-Aldrich) was added and mixed thoroughly. The sample was left to settle for 5 minutes before the liquid layer was extracted carefully with a glass Pasteur pipette. The extracted mix was then evaporated to complete dryness in

GC vials using the SpeedVac. Samples were resuspended in methanol and the lid was crimped on.

5.2.4.3.2 Acid Catalysed Esterification

Based on method B from Basconcillo & McCarry, 2008.

To the dried sample 2ml of 3% (v/v) sulphuric acid dissolved in methanol was added and the mixture was thoroughly mixed on a vortex. The suspension was then incubated in a waterbath set at 50°C for 12-24 hours. Following the period of incubation, the samples were removed from the waterbath and left to cool for 10 minutes. 1ml of hexane was added to the mixture, mixed and the top phase was extracted with a glass Pasteur pipette. The hexane extraction steps were repeated twice and the extracted layers combined. To these combined layers 0.5g of sodium sulphate (anhydrous) (Sigma-Aldrich) was added and mixed thoroughly. The sample was left to settle for 5 minutes before the liquid layer was extracted carefully with a glass Pasteur pipette. The extracted mix was then evaporated to complete dryness in GC vials using the SpeedVac. Samples were resuspended in methanol and the lid was crimped on.

5.3 Results

5.3.1 Alkane Binding

The alkanes dodecane (C12) and hexadecane (C16) were dissolved in DMF and titrated against 2μM of purified Alk1L456S protein. Dodecane binding produced no discernable binding spectra, but hexadecane showed a slight trough at 418nm, which suggests possible type I binding (figure 5.6). Therefore, hexadecane was used for further binding studies with Alk1L456S.

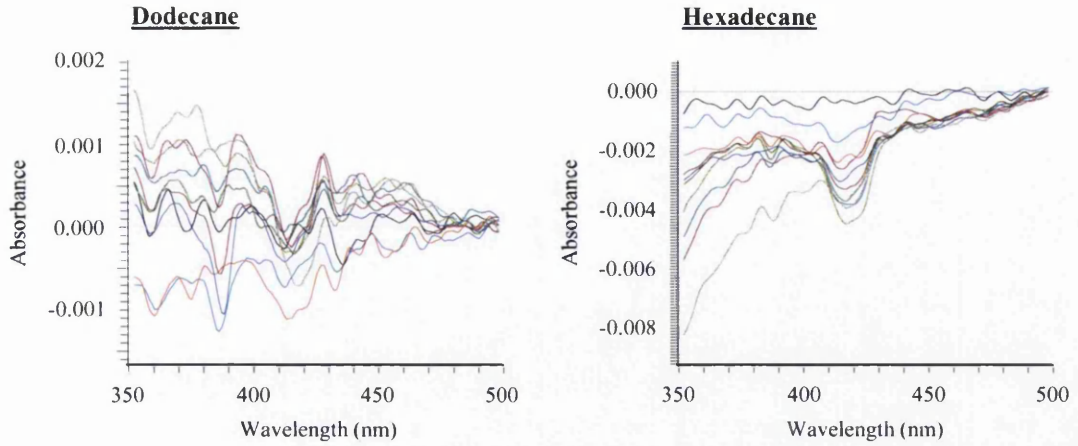


Figure 5.6 – Spectra produced by dodecane and hexadecane in binding studies with Alk1L456S. Dodecane produced no discernable binding spectra. Spectra produced by hexadecane showed a peak at 418nm, suggesting type I binding.

To evaluate whether the solvent used to dissolve the hexadecane was preventing the production of a clear binding spectra, hexadecane was dissolved in ethanol to a stock concentration of 100mM. Titration of this solution against 2μM of Alk1L456S produced neither type I nor type II binding spectra, but did show a degree of precipitation. To evaluate whether the ethanol was causing this precipitation, the 100mM ethanol stock was diluted in buffer D. Diluting the 100mM stock to a 10mM working solution in this buffer resulted in a cloudy suspension, which could not be used for binding studies. A 2mM concentration of hexadecane in this buffer was then produced and was used for binding as the suspension remained clear. Binding with this solution again resulted in no discernable binding spectra. Increasing the concentration of purified Alk1L456S protein from 2μM to 8μM also failed to show any evidence of substrate binding. Figure 5.7 shows an example of the spectra obtained.

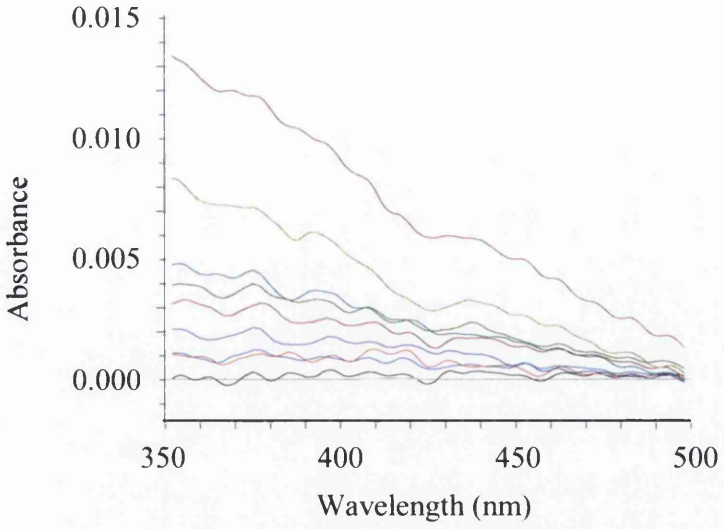


Figure 5.7 – An example of the spectra produced when using hexadecane resuspended in buffer D as the substrate probe in binding studies with Alk1L456S.

To increase the concentration of hexadecane used in these binding studies, the stock 100mM ethanol solution of this hydrocarbon was diluted in different buffers (table 5.1). Table 5.1 shows that a 5mM solution of hexadecane could be produced using 0.1M potassium phosphate, pH7.4 containing 20% (w/v) glycerol. Binding studies undertaken using this hexadecane solution produced no binding spectra.

Buffer	Hexadecane Concentration				
	2mM	3mM	4mM	5mM	6mM
Buffer D	✓	✓	X	-	-
0.1M Tris HCl pH8.1, 25% (w/v) glycerol	✓	✓	X	-	-
0.1M potassium phosphate pH7.4, 20% (w/v) glycerol	✓	✓	✓	✓	X

Table 5.1 – Buffers used to dilute 100mM hexadecane (in ethanol) for use in binding studies with Alk1L456S. The table shows the concentrations of hexadecane achieved by diluting a stock solution of this alkane. Clear suspensions are indicated as ✓ in the table. X shoes the concentrations at which hexadecane produced a cloudy suspension. Concentrations not tried are indicated by -.

To show the binding of hexadecane to CYP52A3, Scheller and co-workers (1996) added protein to different concentrations of substrate rather than titrating the substrate against the protein (the method used in this study). Therefore, this alternative method was used for hexadecane binding studies with Alk1L456S. Rather than producing typical type I or type II binding spectra this method produces an absolute binding spectra where binding is indicated by a shift in the peak to different wavelengths. If hexadecane were to bind to Alk1L456S there would be a shift in the peak from 419nm towards 390nm representing a shift in the conformation of the haem from a low spin complex to a high spin complex. However, no shift in the peak was observed suggesting no binding had taken place between Alk1L456S and hexadecane. The concentration of Alk1L456S protein was increased from 2μM to 8μM and the experiment was repeated. This, again, failed to show a shift in the peak towards 390nm, suggesting no binding of hexadecane to Alk1L456S. An example of the spectra obtained is shown in figure 5.8.

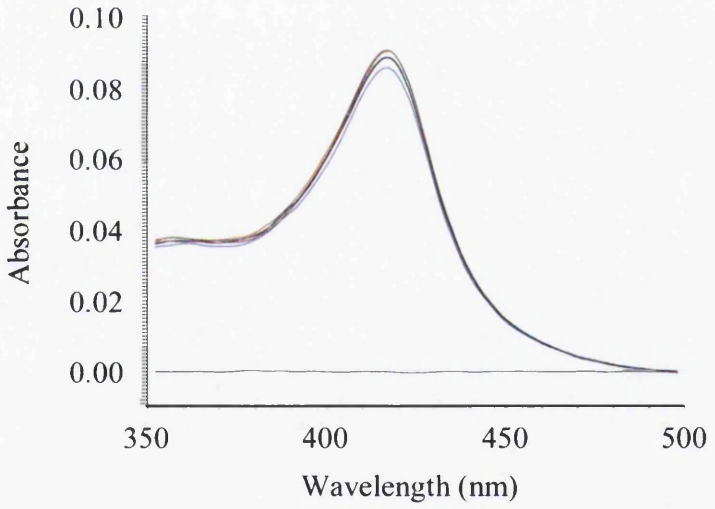


Figure 5.8 – Example of spectra produced using the Scheller *et al.*, 1996 method for binding studies involving Alk1L456S and hexadecane.

5.3.2 Fatty Acid Binding

Various concentrations (5mg/ml and 10mg/ml) of palmitic acid (C16:0) were dissolved in DMF for fatty acid binding studies. 5mg/ml palmitic acid dissolved in DMF showed type I spectra, but failed to reach saturation in titration studies (figure 5.9). Using 10mg/ml, saturation of palmitic acid was reached (figure 5.9).

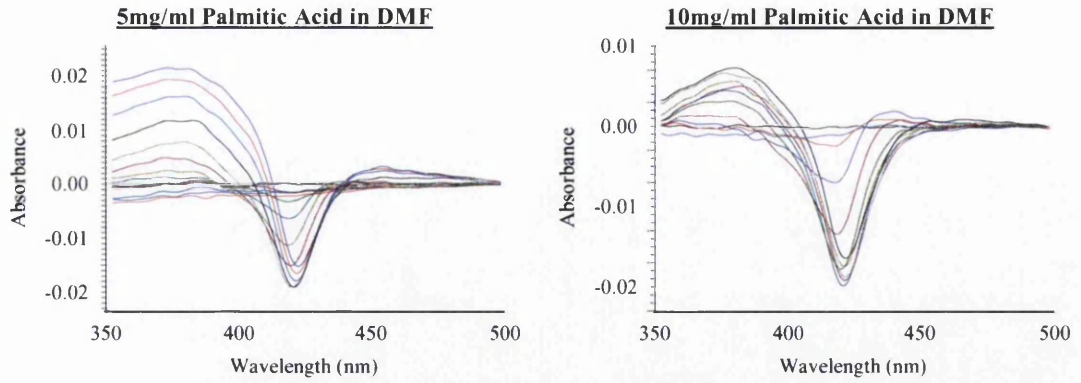


Figure 5.9 – Type I binding spectra produced by 5mg/ml and 10mg/ml of palmitic acid dissolved in DMF.

As palmitic acid showed type I binding spectra with Alk1L456S, other saturated fatty acids were titrated against the protein. Lauric acid (C12:0) and myristic acid (C14:0) both showed type I binding spectra (figure 5.10). Stearic acid (C18:0) produced no discernable spectra and capric acid (C10:0) showed neither a type I nor type II binding spectra (figure 5.10). However, a trough at 416nm was shown, but no identifiable peak at either 385nm or 425-435nm was seen.

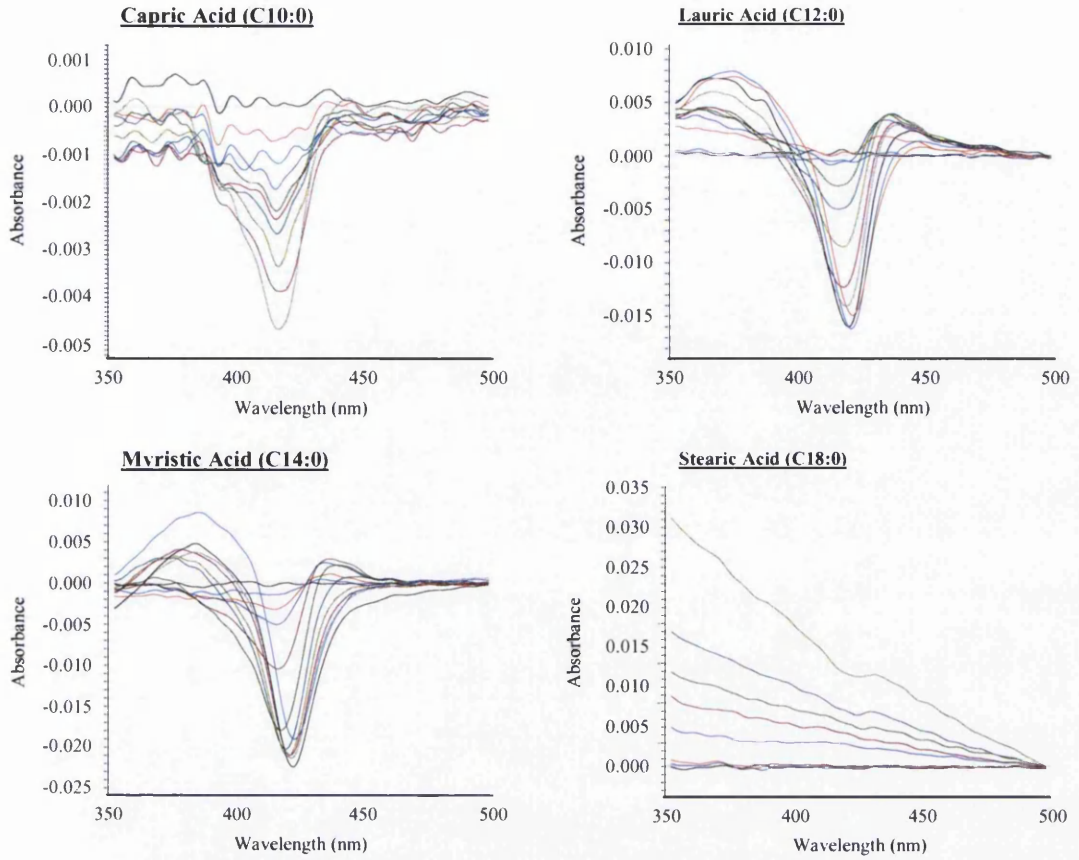


Figure 5.10 – Spectra produced by fatty acids used in binding studies with Alk1L456S. Binding with lauric acid and myristic acid produced type I spectra. Stearic acid showed no discernable spectra and capric acid produced a trough at 416nm, which represents neither type I nor type II binding spectra.

As lauric acid, myristic acid and palmitic acid all showed type I binding with Alk1L456S, the substrate binding constant (K_s) for each fatty acid was calculated using the Lineweaver-Burke equation (see table 5.2).

Fatty Acid	Ks (μM)
Lauric Acid	3466.94±260.55
Myristic Acid	934.05±42.05
Palmitic Acid	139.89±52.97

Table 5.2 – Substrate binding constants (K_s) of the fatty acid-Alk1L456S complexes.

Using the data shown in table 5.2, it is possible to rank the fatty acid binding affinity for Alk1456S from the strongest to the weakest as the greater the binding affinity the smaller the binding constant:

Palmitic acid > Myristic acid > Lauric acid

Palmitic acid was shown to have the smallest K_s value, thus giving it the greatest binding affinity to Alk1L456S. Therefore, this fatty acid was to be used for reconstitution assays with Alk1L456S.

5.3.3 Detection of Hexadecane, Palmitic Acid and their Derivatives on the GCMS

To ensure that hexadecane, palmitic acid and their derivatives could be detected on the GCMS for analysis, standards were produced. These standards were derivatised by a number of methods (silylation (section 2.2.12), base catalysed esterification (section 5.2.4.3.1) and acid catalysed esterification (section 5.2.4.3.2)) and run on the GCMS using the method printed in section 2.2.13.1.

5.3.3.1 Hexadecane

The retention time for hexadecane is 11.304min and the mass ion is 226 (figure 5.11).

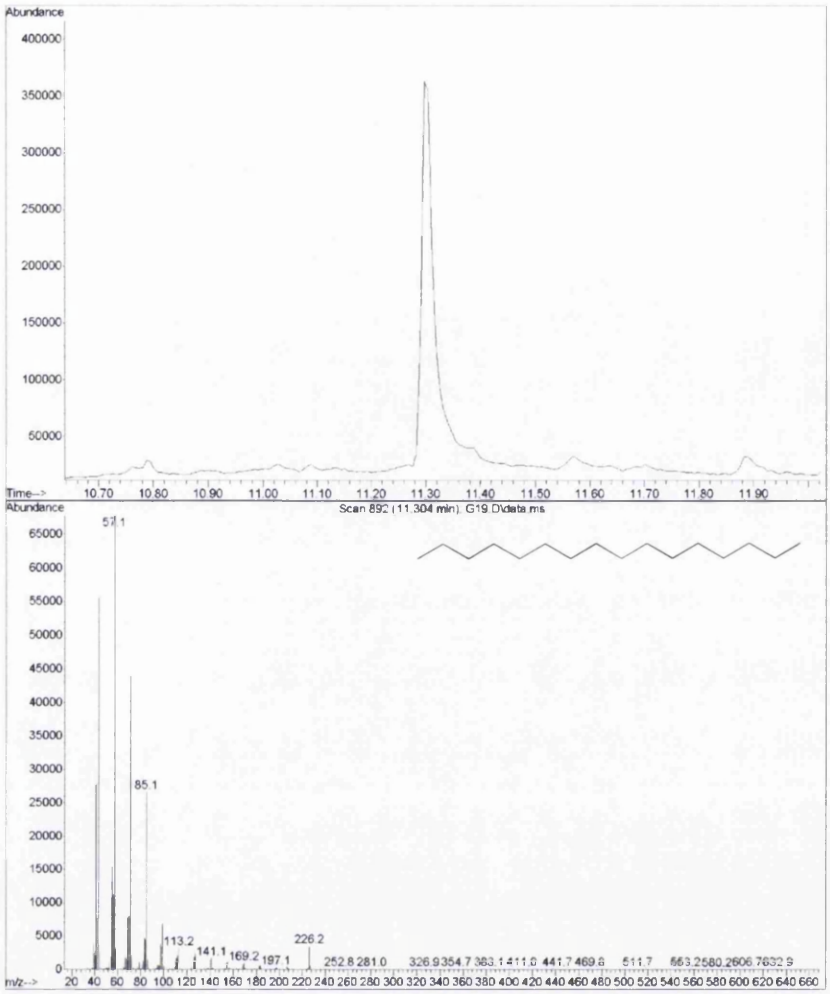


Figure 5.11 – Hexadecane. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) for hexadecane. Inset is the structure of this alkane.

5.3.3.2 Palmitic Acid

To identify fatty acids using the GCMS, they first require derivatisation. Therefore, palmitic acid was derivatised using silylation, base catalysed esterification and acid catalysed esterification methods to identify the optimal protocol for identification of this standard on the GCMS.

5.3.3.2.1 Silylation

Silylation did not result in the full conversion of palmitic acid to its derivatised form. 76% of the palmitic acid standard used was derivatised to its TMS version, while 24% remained underivatised. The retention time for underivatised palmitic acid is 15.176min and the mass ion is 256 (figure 5.12).

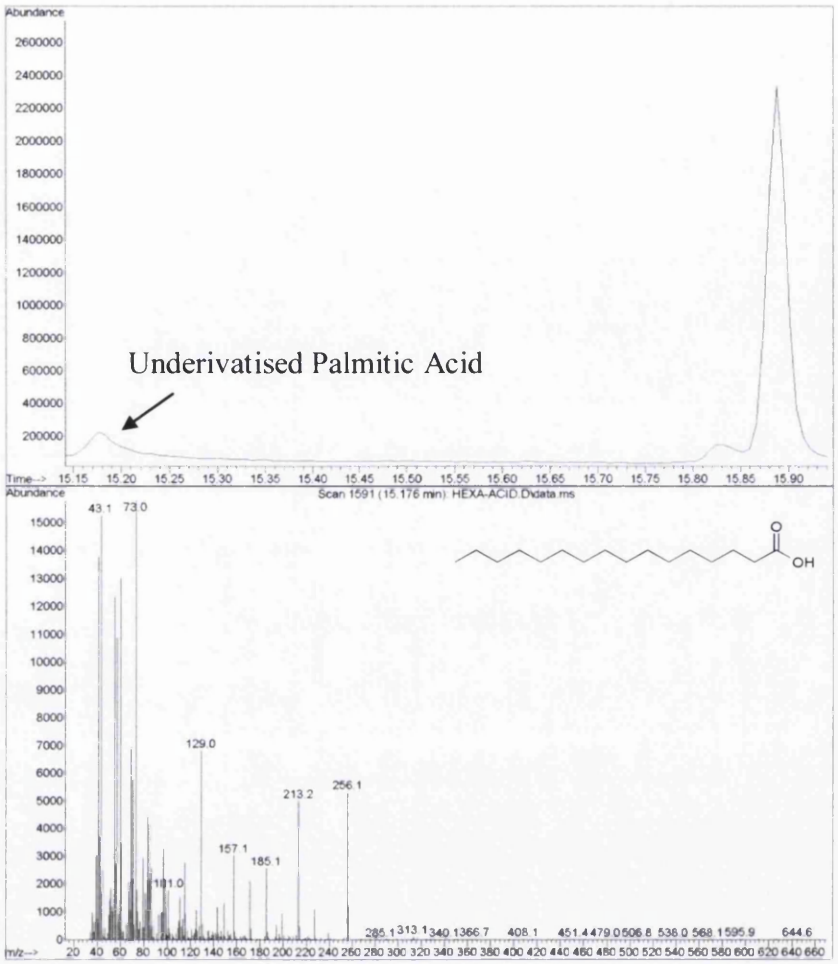


Figure 5.12 – Underivatised palmitic acid. GCMS results produced following the silylation of palmitic acid. The underivatised palmitic acid peak is indicated on the GC chromatograph (top). The MS spectrum (bottom) shows the fragmentation pattern of this underivatised fatty acid. The structure is shown inset.

The addition of $Si(CH_3)_3$ to the carboxyl side-chain of palmitic acid during silylation adds 71 to the fatty acids molecular weight (underivatised molecular weight = 256). Therefore, the mass ion for TMS derivatised palmitic acid is 327 and the retention time is 15.8min (figure 5.13).

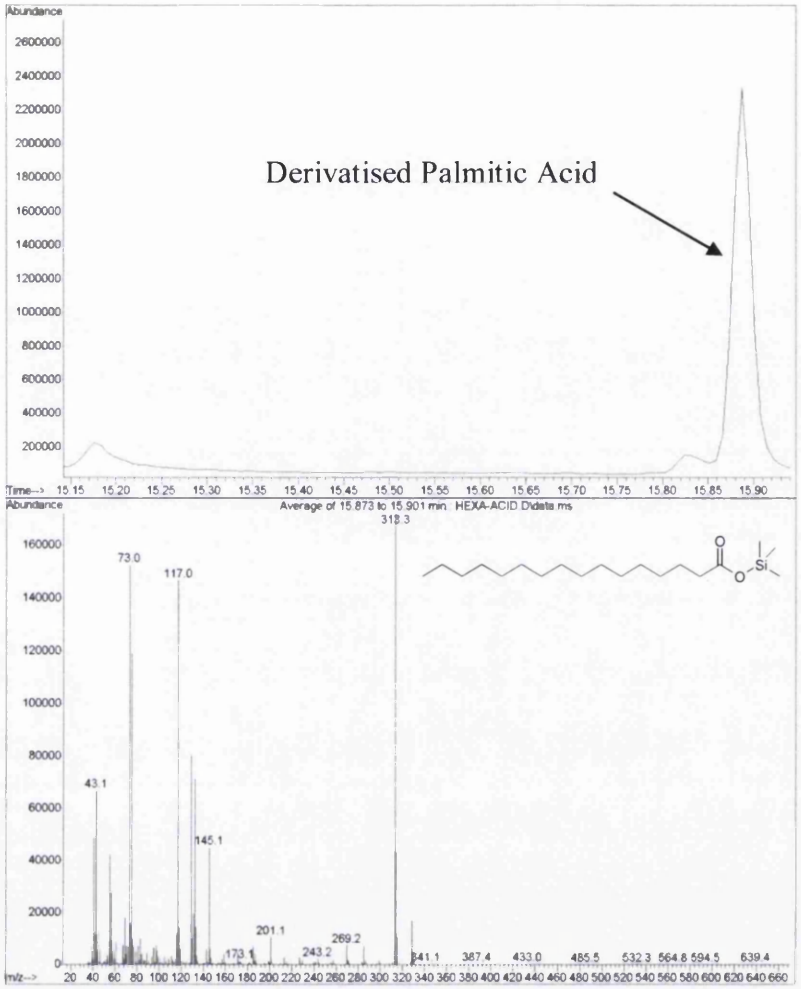


Figure 5.13 – TMS-derivatised palmitic acid. The GC chromatograph (top) shows the peak produced by TMS derivatised palmitic acid. The fragmentation pattern for this hydrocarbon is shown in the MS spectrum (bottom). Inset is the structural overview of the TMS-derivatised palmitic acid molecule.

5.3.3.2.2 Base Catalysed Esterification

The entire palmitic acid standard was derivatised to its methyl ester form using this method. The addition of CH_3 to the hydroxyl group of the carboxyl side-chain increased the molecular weight of the fatty acid from 256 to 270. Therefore, the mass ion of palmitic acid methyl ester is 270 and the retention time is 13.953min (figure 5.14).

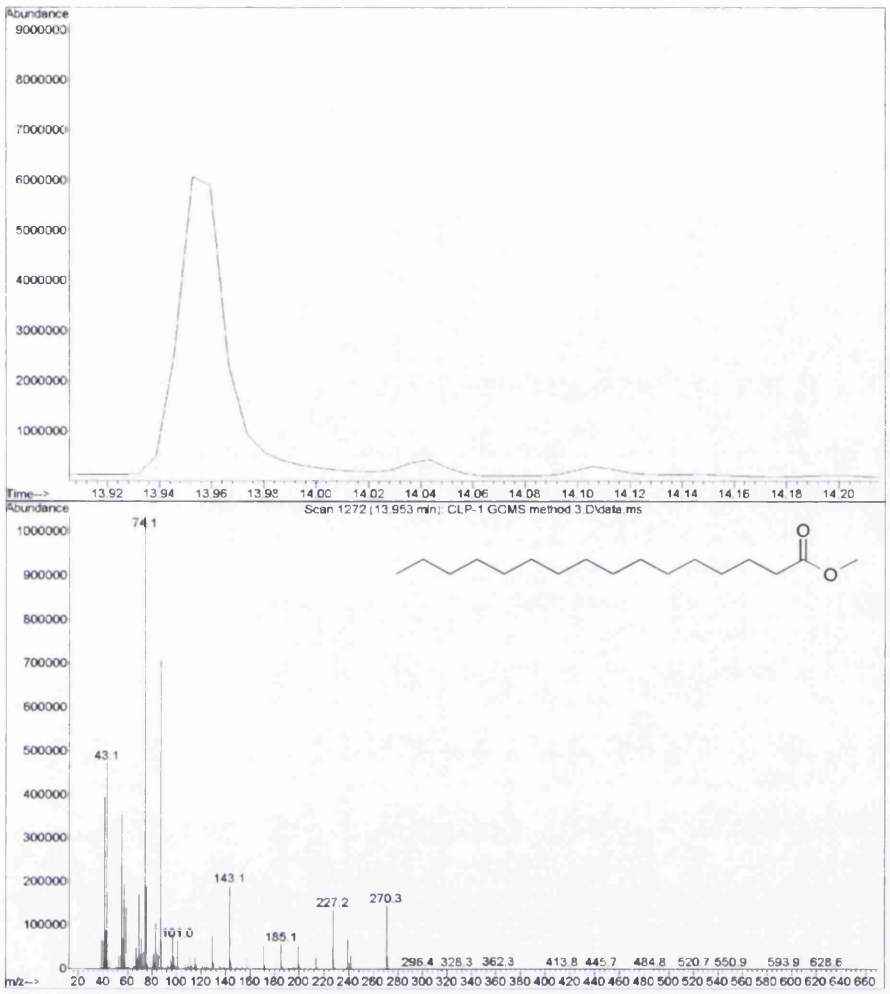


Figure 5.14 – Palmitic acid, methyl ester produced using sodium methoxide as the catalyst in esterification. The GC chromatograph (top) shows a peak corresponding to palmitic acid, methyl ester. The MS spectrum (bottom) shows the fragmentation pattern. Inset is the esterified structure of palmitic acid.

5.3.3.2.3 Acid Catalysed Esterification

As with base catalysed esterification (section 5.3.3.2.2), the entire palmitic acid standard was converted to its methyl ester form. The retention time for palmitic acid methyl ester is 13.102min and the mass ion is 270 (figure 5.15).

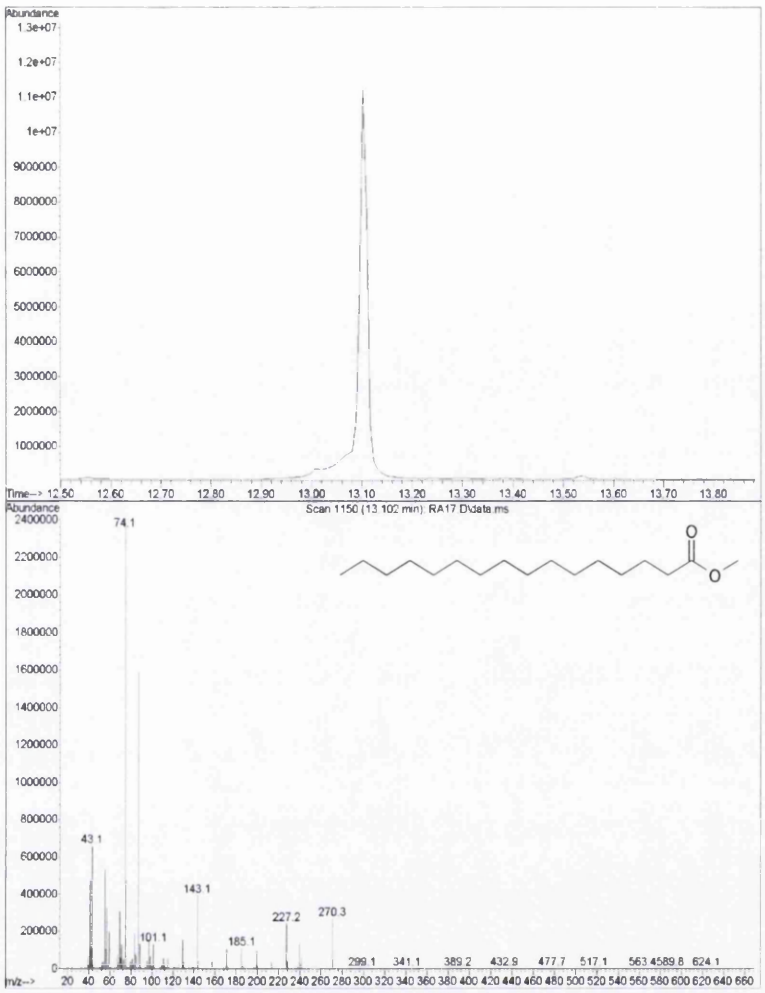


Figure 5.15 – Palmitic acid, methyl ester produced using 3% (v/v) sulphuric acid as the catalyst in esterification. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) of palmitic acid, methyl ester. Inset is the structure of this esterified standard.

5.3.3.3 Thapsic Acid

5.3.3.3.1 Base Catalysed Esterification

Using the base catalysed esterification method, thapsic acid was unable to be detected on the GCMS (figure 5.16).

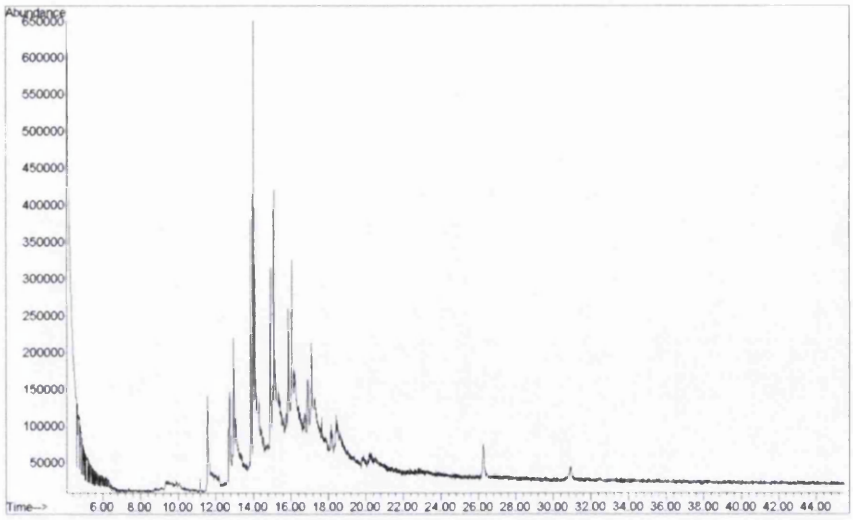


Figure 5.16 – GC chromatograph following the base catalysed esterification of the thapsic acid standard. No peak representing thapsic acid can be identified on the GC chromatograph.

5.3.3.3.2 Acid Catalysed Esterification

Thapsic acid was able to be identified on the GCMS using this method. The retention time is 15.5min and the mass ion is 283 (figure 5.17).

As the acid catalysed method can both esterify palmitic acid and allow thapsic acid to be detected by the GCMS, it is the method of choice to esterify the products of this study.

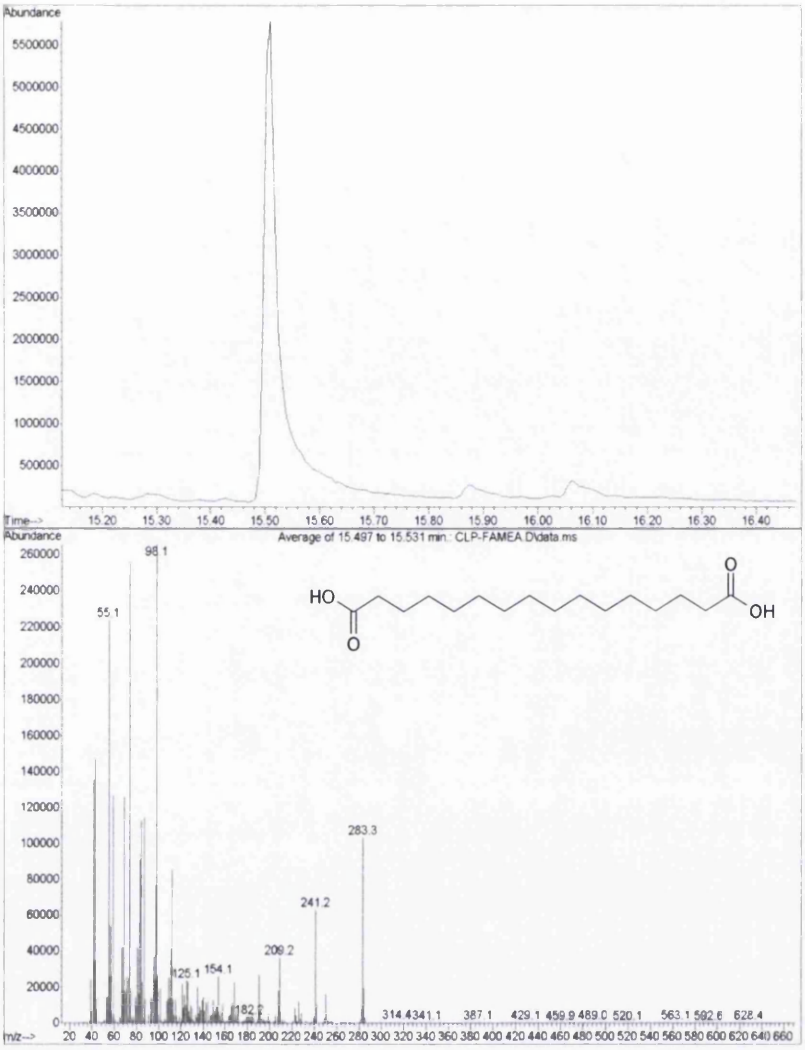


Figure 5.17 – Thapsic acid. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) of this standard. The structure of the standard is shown inset.

5.3.3.4 Hexadecanol

The retention time of hexadecanol is 12.886min and the mass ion is 242 (figure 5.18).

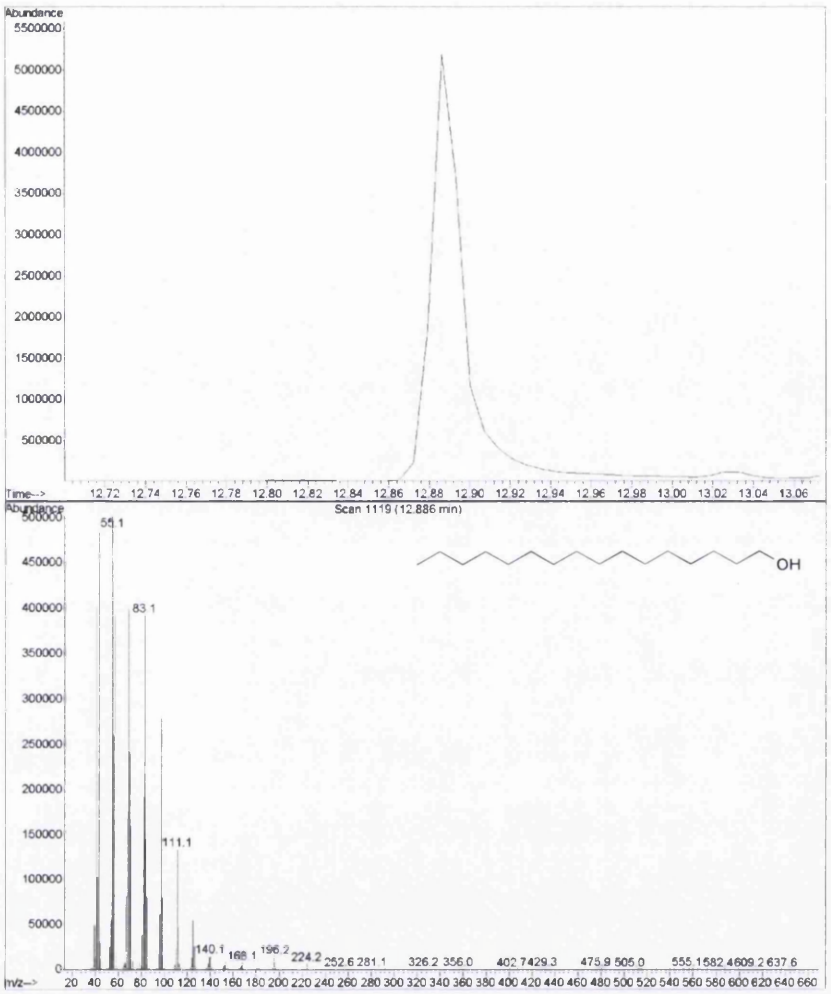


Figure 5.18 – Hexadecanol. The GC chromatograph (top) shows a peak corresponding to hexadecanol. The MS spectrum (bottom) shows the fragmentation pattern and inset is the alcohol structure.

5.3.3.5 16-Hydroxypalmitic Acid

16-hydroxypalmitic acid could not be detected on the GCMS following acid catalysed esterification alone. The substrate required both esterification and silylation. This resulted in a retention time of 14.705min and a mass ion for 16-hydroxypalmitic acid of 358 (figure 5.19).

As both derivatisation techniques are required to detect 16-hydroxypalmitic acid on the GCMS, both of these methods will be used to identify the hydrocarbons in this study.

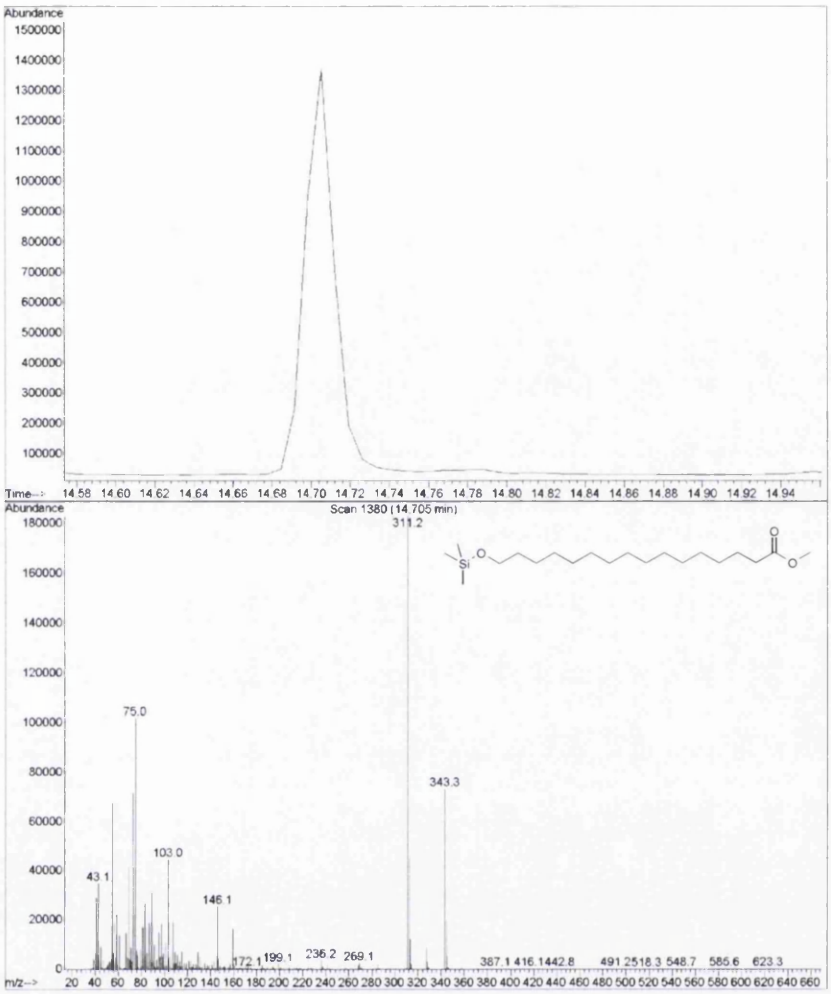


Figure 5.19 - 16-Hydroxypalmitic acid, TMS, methyl ester. The GCMS results for 16-hydroxypalmitic acid, TMS, methyl ester showing the GC chromatograph (top) and MS spectrum (bottom) produced. Inset the structure of the silylated and esterified fatty alcohol is shown.

5.3.3.6 Palmitic Acid-16,16,16,d_3

Acid catalysed esterification resulted in the production of palmitic acid-16,16,16-d_3, methyl ester. This hydrocarbon has a retention time of 13.066min and a mass ion of 273 (figure 5.20). The additional silylation step had no effect on the methyl ester molecule produced by esterification.

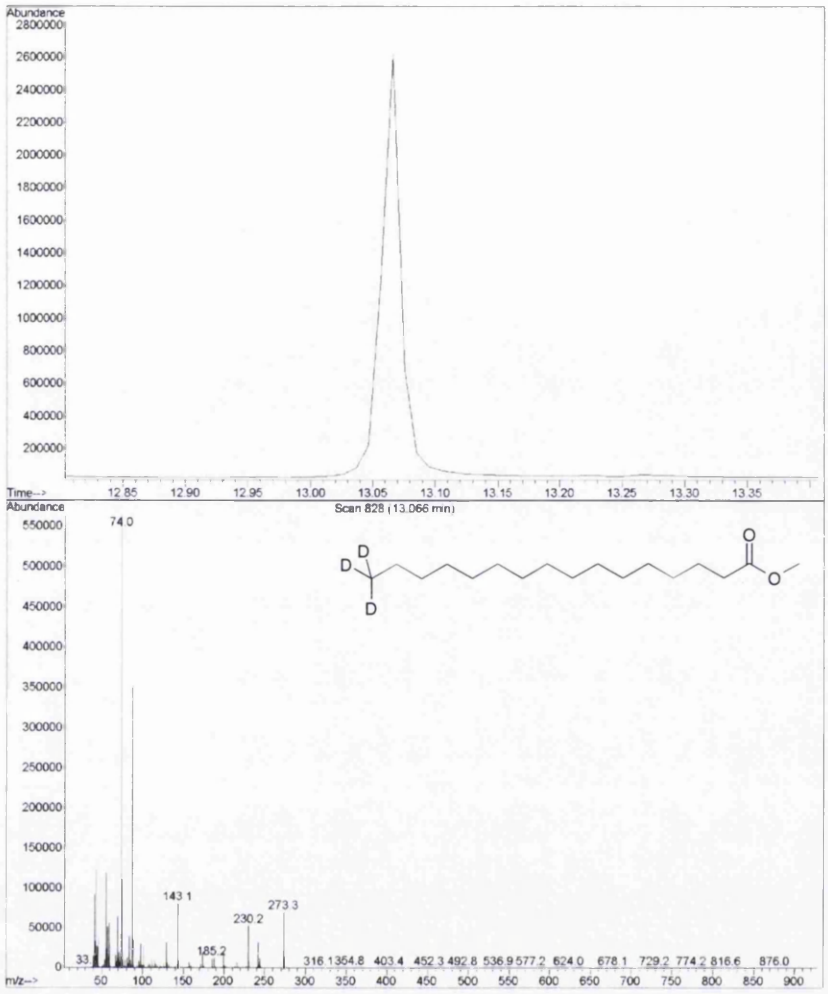


Figure 5.20 - Palmitic acid-16,16,16-d_3, methyl ester. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) for the methyl ester form of palmitic acid-16,16,16-d_3. The structure is shown inset.

5.3.4 Reconstitution Assays

5.3.4.1 Reconstitution Assays Using Purified Alk1L456S Protein to Catalyse the Reaction

As palmitic acid was shown to have the greatest binding affinity to Alk1L456S of the hydrocarbons used for binding studies in this chapter, this fatty acid was used as the substrate in reconstitution assays. Initially, assays were set-up using ScΔ33:CPR as the electron donor and a buffer of pH6.4. Negative controls were set-up with NADPH omitted from the assay as NADPH is required for electron donation. No product (16-hydroxypalmitic acid or thapsic acid) was detected in either of these reactions (figure 5.21).

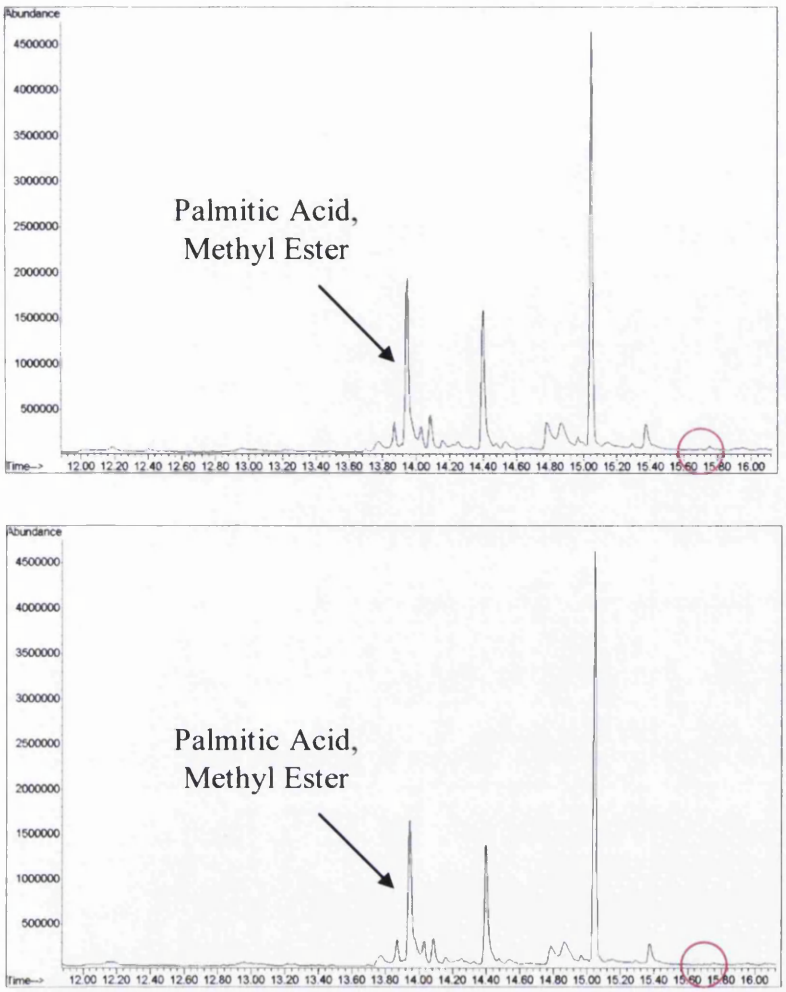


Figure 5.21 – GC chromatographs for reconstitution assays undertaken in the presence and absence of NADPH. Both GC traces show similar peaks. The peak representing palmitic methyl ester is indicated above. The pink ring identifies the retention time at which a peak representing 16-hydroxypalmitic acid, TMS, methyl ester should be seen if detected by the GCMS. This ring shows no peak in this area. The other large peaks detected on this GC trace correlate to silane molecules, which are indicative of column bleed (degradation of the column).

To identify whether the failure of Alk1L456S to produce product from palmitic acid in these assays was due to pH, a number of 1M potassium phosphate buffers were used each with a different pH (pH6.4, pH6.6, pH6.8, pH7.0 and pH7.2). Negative controls were set-up for each pH without NADPH. Again, no product was detected on the GCMS (figure 5.22) suggesting product formation in these assays was not affected by the pH of the buffer used.

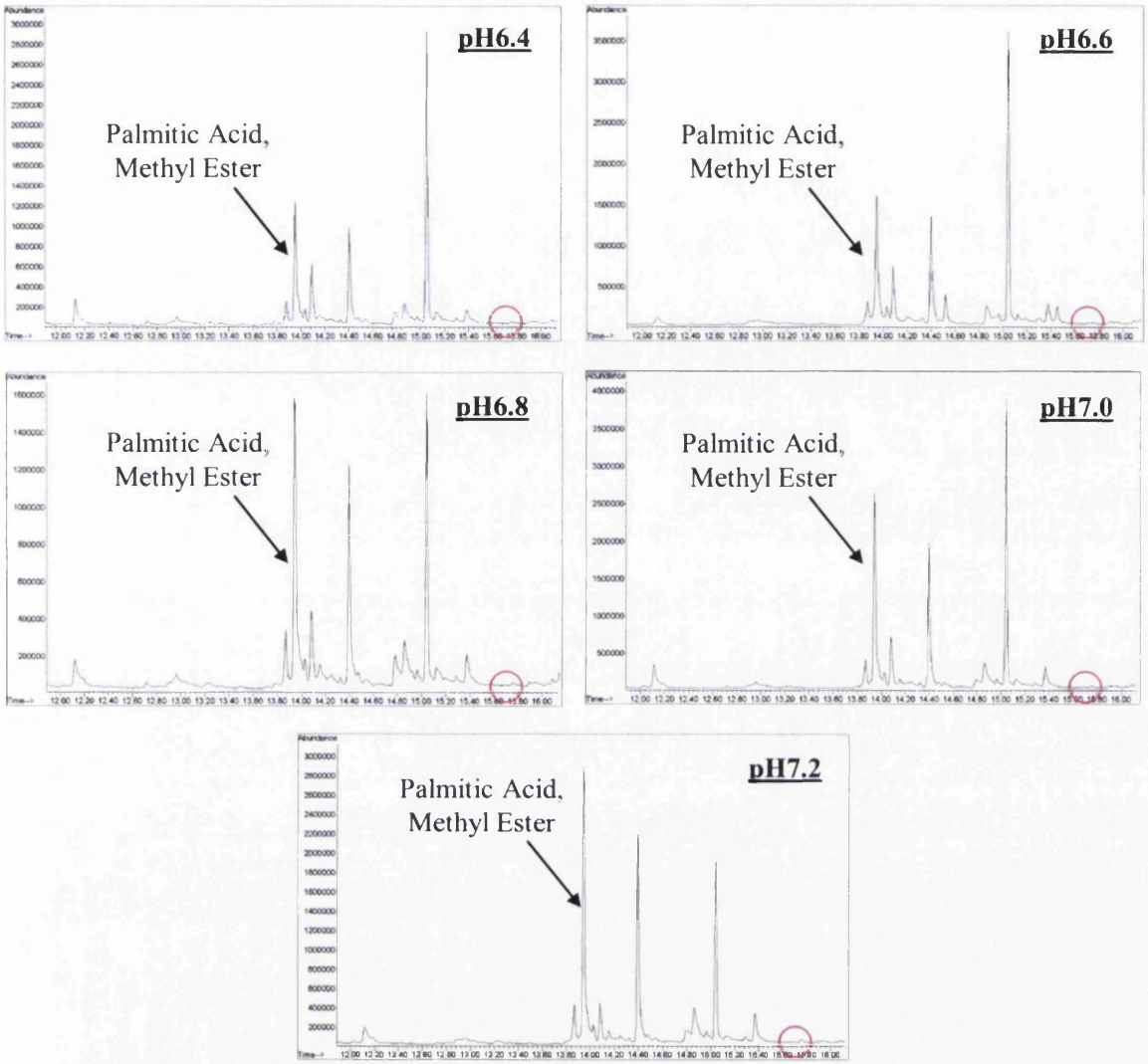


Figure 5.22 – GC chromatographs for reconstitution assays using different 1M potassium phosphate buffers, each with a different pH. The substrate peak is indicated. A pink ring indicates where a peak representing 16-hydroxypalmitic acid, TMS, methyl ester should be found on the GC trace. There is no peak indicating no product detection. The other large peaks are again due to column bleed.

To discover if the CPR used (ScΔ33:CPR) in these assays with Alk1L456S was affecting product formation (possibly due to poor electron transfer between the reductase and Alk1L456S), CaCPR and hCPR were also used. Negative controls were also set-up with CPR omitted from the assay. Product was not detected in any of the GC chromatographs produced in this experiment (figure 5.23), suggesting the reductase used was not the limiting factor in product formation.

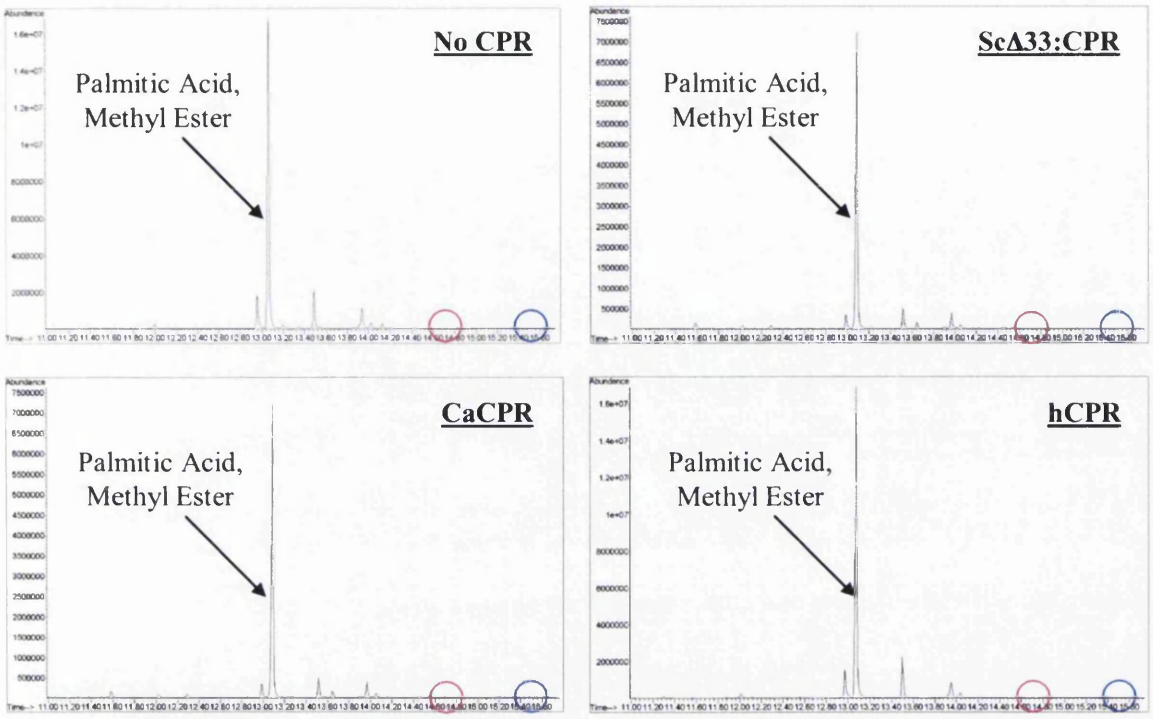


Figure 5.23 – GC chromatographs for reconstitution assays using different reductases for electron transfer. The substrate is indicated. Pink ring represents the retention time at which a peak representing 16-hydroxypalmitic acid, TMS, methyl ester should be seen. The blue ring represents thapsic acid. Peaks representing either metabolite were not identified.

The lack of product formation in these assays could be due to the concentration of CPR used in these reactions, rather than the type of CPR being used. Other cytochrome P450 enzymes have been shown to require increased ratios of reductase to cytochrome P450 protein for substrate metabolism to occur. For instance, CYP4A1 requires reductase to be present in the assay in excess at a ratio of 10:1 for product formation to occur (He *et al.*, 2005). Therefore, CPR titrations were set-up using various ratios of CPR to Alk1L456S. Assays containing ScΔ33:CPR were set-up at ratios of 1:1 with Alk1L456S and in excess at 2:1, 4:1 and 10:1. Therefore, in assays containing 2.5μM of Alk1L456S protein, 2.5μM, 5μM, 10μM and 25μM of ScΔ33:CPR was added respectively. Negative controls were set-up with CPR omitted. These titration experiments, again, resulted in a failure to detect any product on the GC chromatographs produced (figure 5.24).

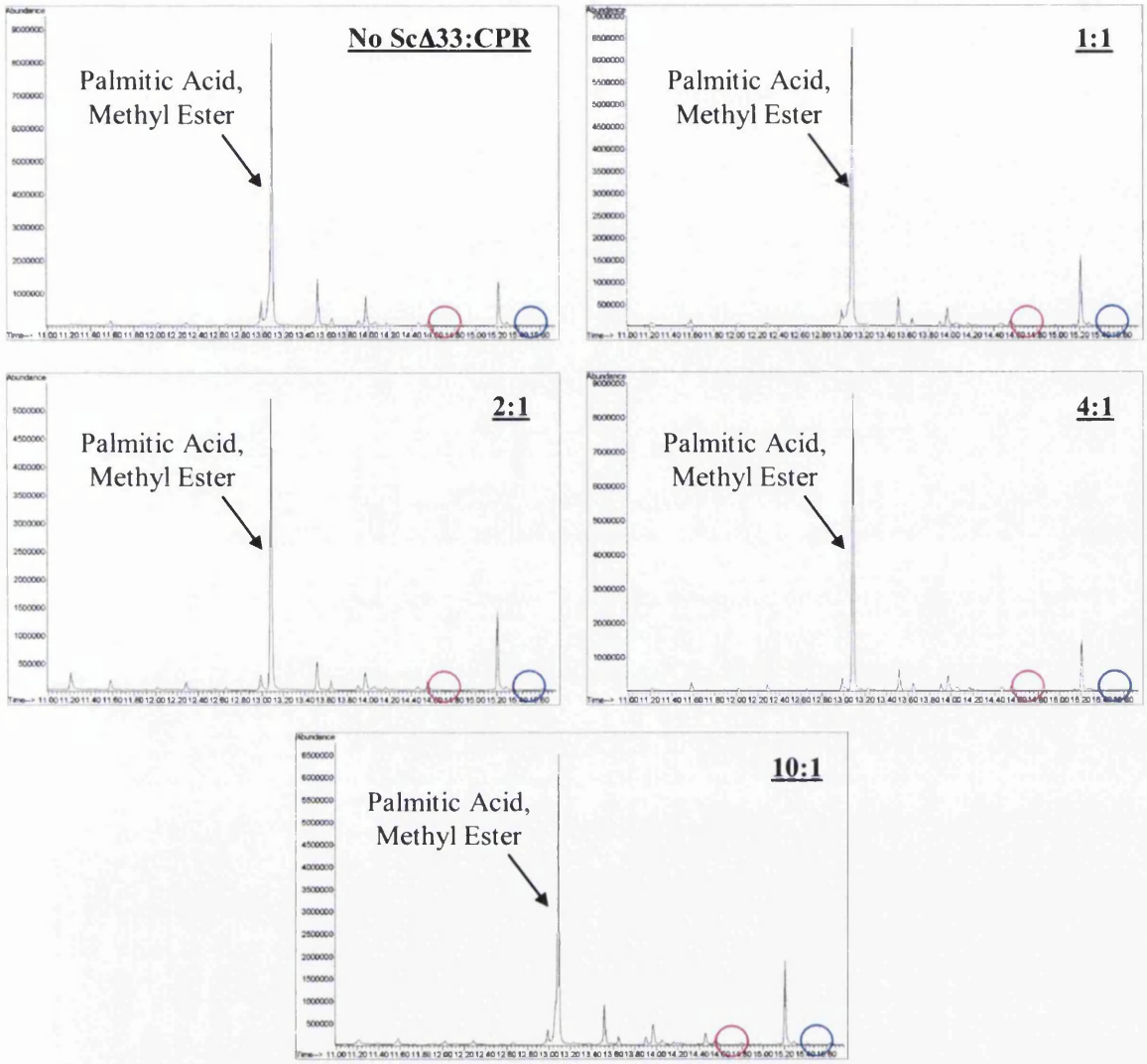


Figure 5.24 – GC chromatographs for reconstitution assays using different ratios of ScΔ33:CPR to Alk1L456S. The substrate peak is identified. The pink and blue rings represent the location at which peaks representing 16-hydroxypalmitic acid, TMS, methyl ester and thapsic acid (respectively) should be detected. No peaks corresponding to these metabolites were identified.

CaCPR was used for similar CPR titration experiments with Alk1L456S. Ratios were set-up with CaCPR in excess at 2.5:1, 5:1 and 10:1. To be able to titrate CaCPR at a ratio of 10:1 with Alk1L456S, the concentration of Alk1L456S needed to be reduced from 2.5μM (as used in ScΔ33:CPR titration experiments) to 0.5μM. Therefore, this concentration of Alk1L456S protein was used in all assays with the addition of 1.25μM, 2.5μM and 5μM of CaCPR respectively. CPR was not added to the negative controls. As with the ScΔ33:CPR titration experiments, no product was detected on the GCMS (figure 5.25).

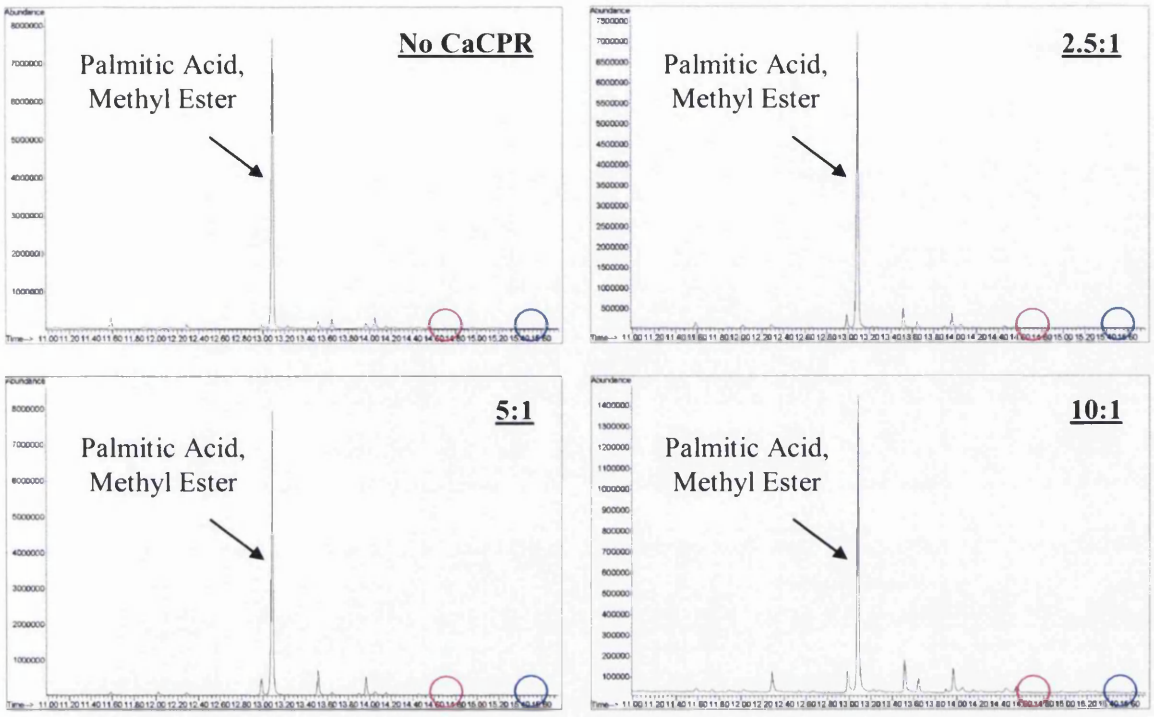


Figure 5.25 – GC chromatographs for reconstitution assays using different ratios of CaCPR to Alk1L456S. The substrate peak is identified. Pink and blue rings identify the retention times at which the alcohol and diacid (respectively) metabolites should be detected. No products were identified.

The incubation period (1 hour) at 37°C may be too brief for the reaction to take place and for product formation to occur. Therefore, the CaCPR titration experiments mentioned above were repeated, except the period of incubation at 37°C was increased from 1 hour to 18 hours to ensure the experiment was of sufficient length for the reaction to take place. No product was detected in GC chromatographs produce as a result of each assay (figure 5.26).

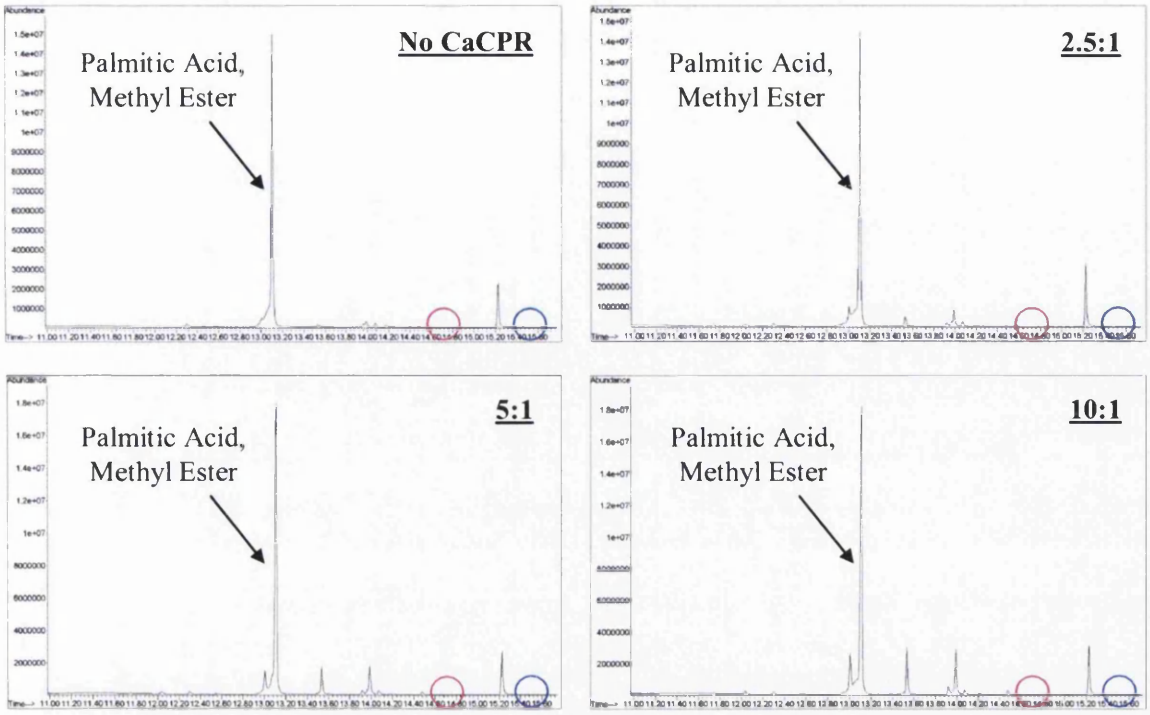


Figure 5.26 – GC chromatographs for reconstitution assays using different ratios of CaCPR to Alk1L456S and an increased incubation period at 37°C. The substrate is identified on the GC chromatograph. The pink and blue rings show the potential locations of peaks representing 16-hydroxypalmitic acid, TMS, methyl ester and thapsic acid (respectively). Peaks representing these metabolites were not identified.

Kim and co-workers (2007) utilised a deuterated version of lauric acid in reconstitution assays with Alk8 from *C. albicans*, therefore a similar version of palmitic acid, palmitic acid-16,16,16-d_3, was used in assays with Alk1L456S to identify whether product formation could be detected on the GCMS. CaCPR titrations were set-up as mentioned previously using palmitic acid-16,16,16-d_3 as the substrate. Samples were initially incubated for 1 hour. Product was not detected on the GCMS (figure 5.27).

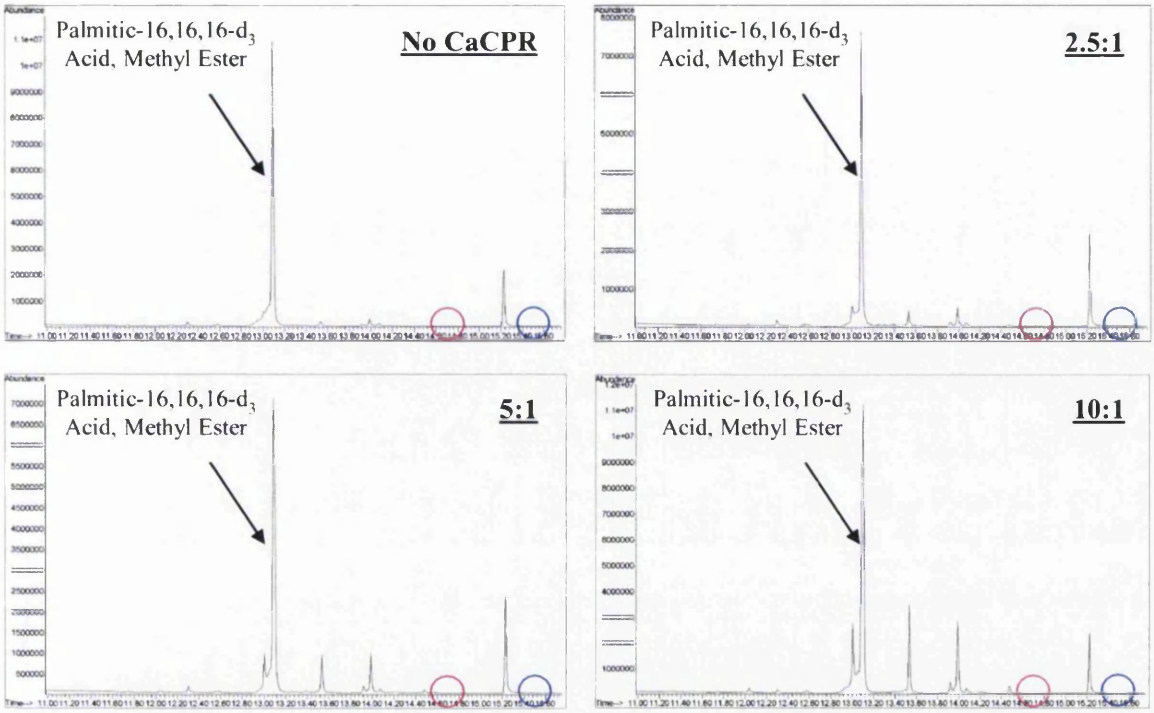


Figure 5.27 – GC chromatographs for reconstitution assays using different ratios of CaCPR to Alk1L456S and palmitic acid-16,16,16-d_3 as the substrate. The peak representing palmitic acid-16,16,16-d_3 acid, methyl ester was identified on the GC trace. The pink and blue rings highlight the area where peaks representing 16-hydroxypalmitic acid, TMS, methyl ester and thapsic acid (respectively) should occur. No peaks were detected on this GC chromatograph.

As no product was detected after 1 hour the period of incubation was increased to 18 hours. No product was detected as a result of these assays (figure 5.28).

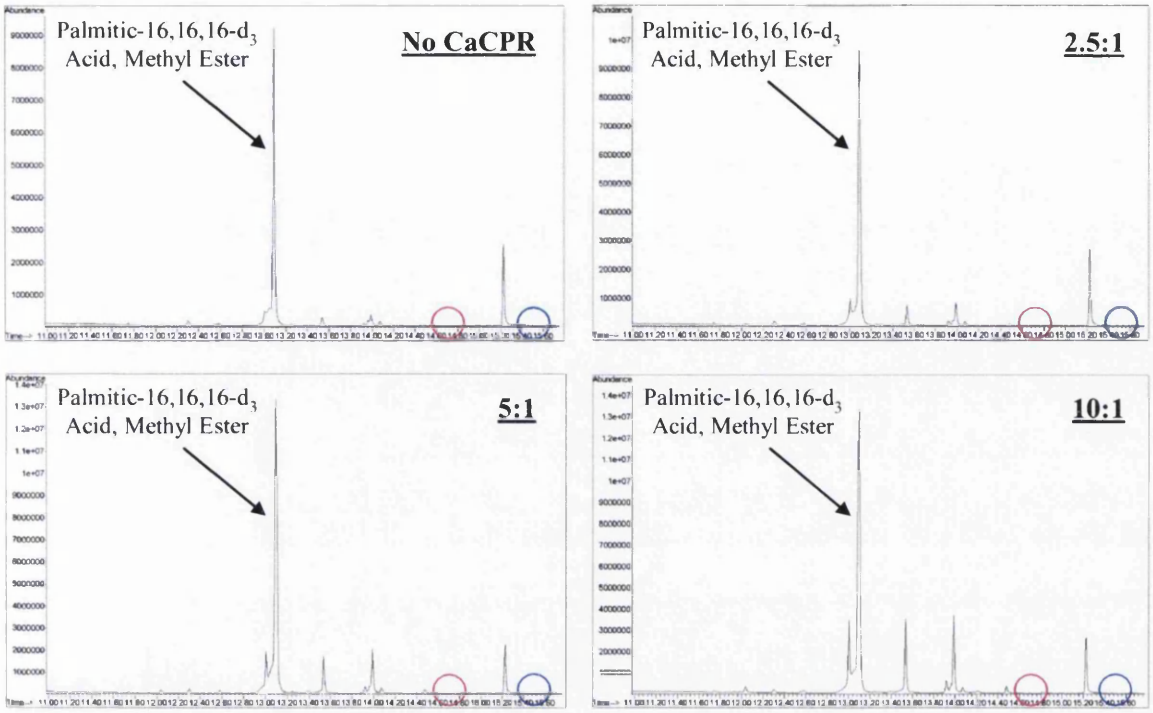


Figure 5.28 – GC chromatographs for reconstitution assays using different ratios of CaCPR to Alk1L456S, an increased incubation period at 37°C and palmitic acid-16,16,16-d_3 as the substrate. The substrate is identified on the GC chromatograph. The pink and blue rings show the potential locations of peaks representing 16-hydroxypalmitic acid, TMS, methyl ester and thapsic acid (respectively). No peaks were identified.

5.3.4.2 Reconstitution Assays Using Microsomal Preparations of CYP52A3_yeast to Catalyse the Reaction

Microsomal preparations of CYP52A3_yeast were used in reconstitution assays to catalyse the reactions. Initially 10µl of 2mg/ml of palmitic acid-16,16,16-d_3 was dissolved in ethanol and used as the substrate for this reaction. This substrate was used as microsomal controls set-up with the substrate and reconstitution system omitted showed the presence of palmitic acid, methyl ester on the GC chromatograph. The use of deuterated palmitic acid will allow the two peaks to be distinguished from one another on the GC chromatograph and MS spectum. Ethanol was used as the solvent to prevent the precipitation of microsomal proteins, which would occur with chloroform. These assays showed no product formation (with or without the presence of CaCPR) and the substrate, palmitic acid-16,16,16-d_3, could not be detected on the GC chromatograph (see figure 5.29).

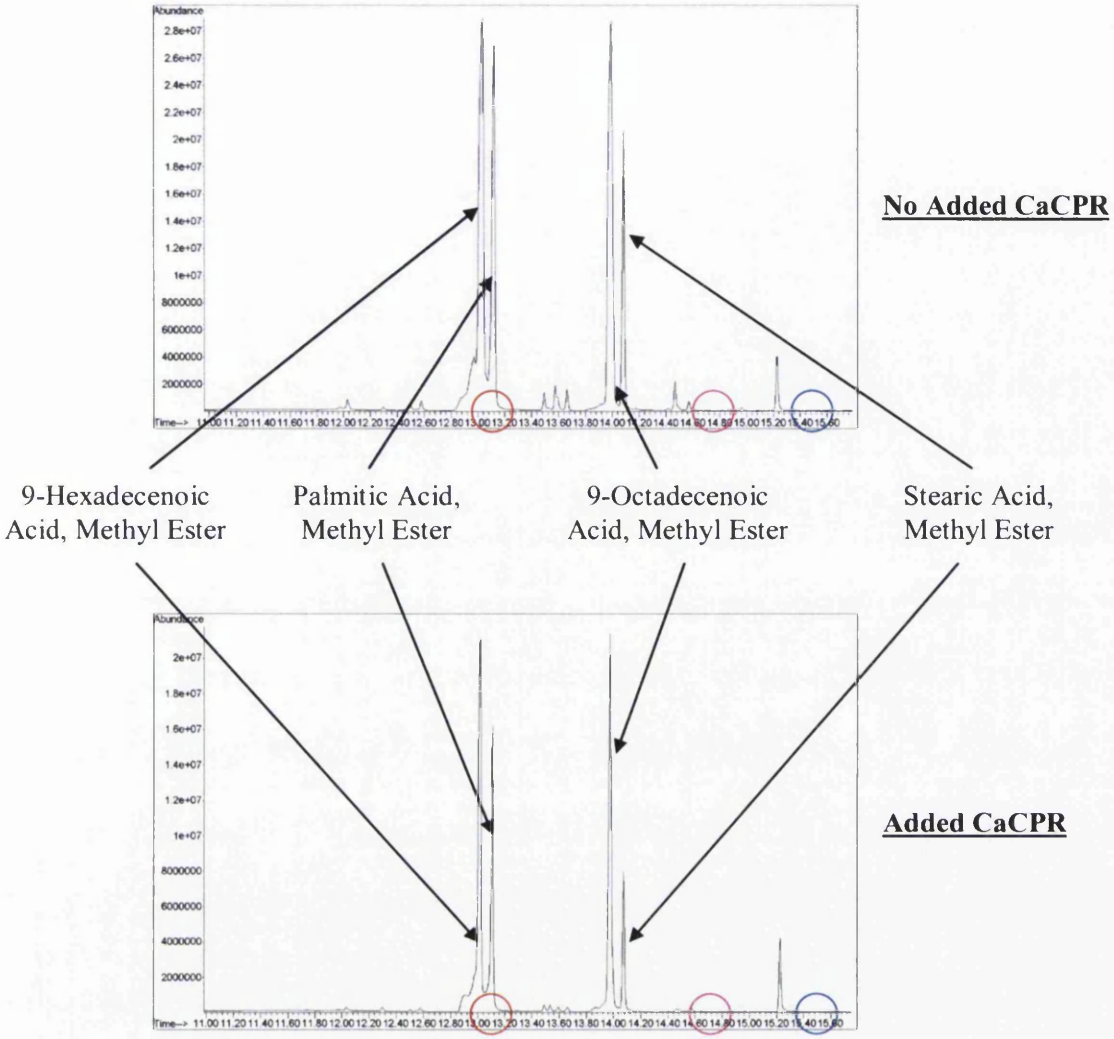
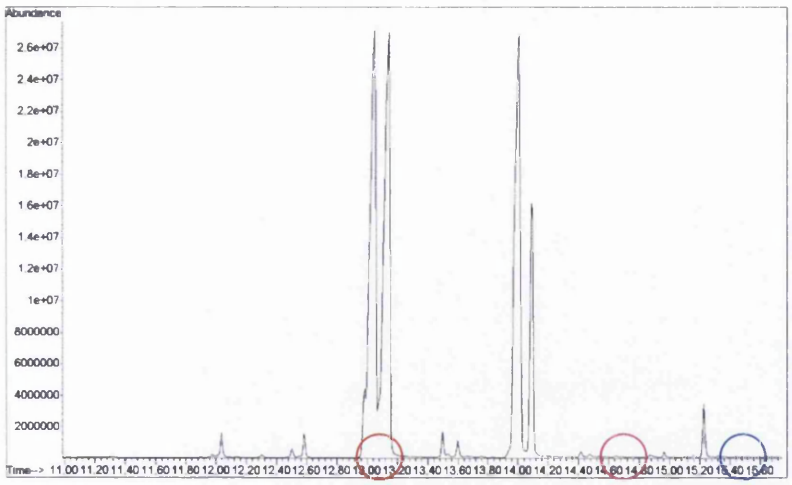
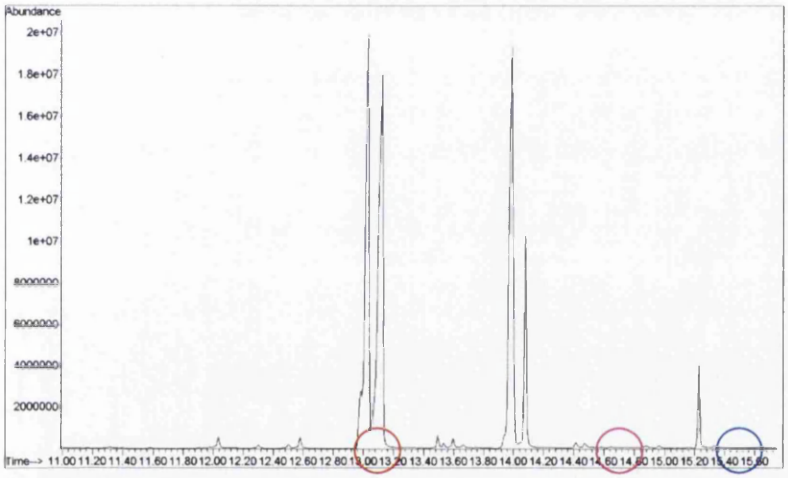


Figure 5.29 – GC chromatographs for reconstitution assays using CYP52A3_yeast microsomes and palmitic acid-16,16,16-d_3 dissolved in ethanol as substrate. The probable membrane lipids are identified. The red ring indicates the location/retention time at which a peak representing palmitic acid-16,16,16-d_3, methyl ester should be identified. The pink ring represents 16-hydroxypalmitic acid, TMS, methyl ester. The blue ring correlates to thapsic acid.

The inability to identify palmitic acid-16,16,16-d_3 could either be due to the large peaks representing the membrane lipids (including palmitic acid, methyl ester) obscuring the peak produced by the substrate or the solvent used to dissolve the substrate was not compatible for this assay. The reconstitution assays were repeated using an increased amount of palmitic acid-16,16,16-d_3 (10µl of 10mg/ml) dissolved in DMF rather than ethanol. These reactions produced similar results as neither the substrate nor the product was detected (figure 5.30). Again, the addition of CaCPR had no effect on product formation (figure 5.30).



No Added CaCPR



Added CaCPR

Figure 5.30 – GC chromatographs for reconstitution assays using CYP52A3_yeast microsomes and palmitic acid-16,16,16-d_3 dissolved in DMF as substrate. The red ring indicates the location/retention time at which a peak representing palmitic acid-16,16,16-d_3, methyl ester should be identified. The pink ring represents 16-hydroxypalmitic acid, TMS, methyl ester. The blue ring correlates to thapsic acid.

To check whether the solvent being used in this assay was preventing the detection of the hydrocarbon, 10µl of 10mg/ml hexadecane dissolved in DMF was used as an alternative substrate. Hexadecane was detected on the GC chromatograph (figure 5.31), but as a small peak on the trace as the peaks representing the membrane lipids were large and obscuring the other peaks produced. This made it difficult to identify any other peaks, including those that may have been produced by any metabolites formed by these assays. Therefore, microsomal preparations containing Alk1L456S were not produced to act as the lipid component of reconstitution assays instead of DLPC.

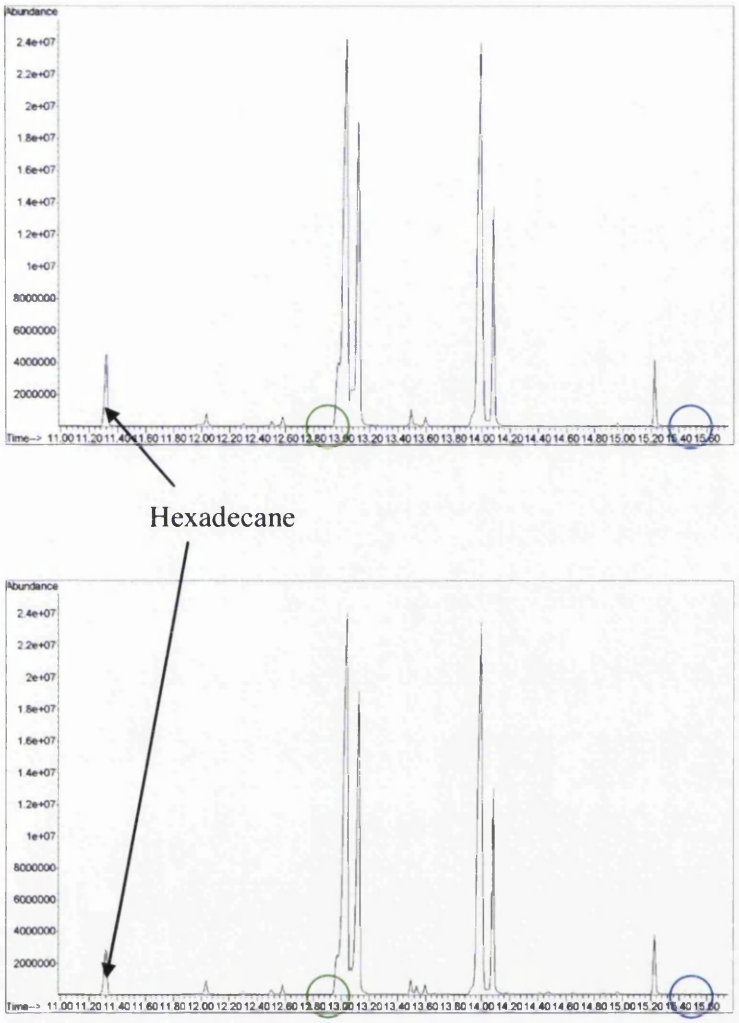


Figure 5.31 – GC chromatographs for reconstitution assays using CYP52A3_yeast microsomes and hexadecane as substrate. The substrate is indicated on the GC traces. The green and blue rings show the potential locations of peaks representing hexadecanol and thapsic acid (respectively). No peaks were identified.

5.3.4.3 Metabolism of Hexadecane by *S. cerevisiae* Transformed with YEp51:Alk1L456S_yeast Plasmid

As reconstitution assays failed to show product formation in reaction catalysed by Alk1L456S, *in vivo* experiments were set up. *S. cerevisiae* cells transformed with YEp51:Alk1L456S_yeast plasmid were grown on media containing hexadecane, despite hexadecane producing no binding spectra in binding studies with Alk1L456S. This is because *S. cerevisiae* is able to utilise fatty acids for growth, but not alkanes (Rehm & Reiff, 1981), thereby allowing *S. cerevisiae* containing only the YEp51 vector to be used as a negative control in this study.

S. cerevisiae cells containing YEp51:Alk1L456S_yeast, YEp51:CYP52A3_yeast (positive control) and YEp51 only (negative control) were grown in minimal salts media containing 2% (w/v) galactose and 0.1% (v/v) hexadecane (figure 5.32). Galactose was added to the media to induce the transcription of Alk1L456S_yeast and CYP52A3_yeast in these cells.

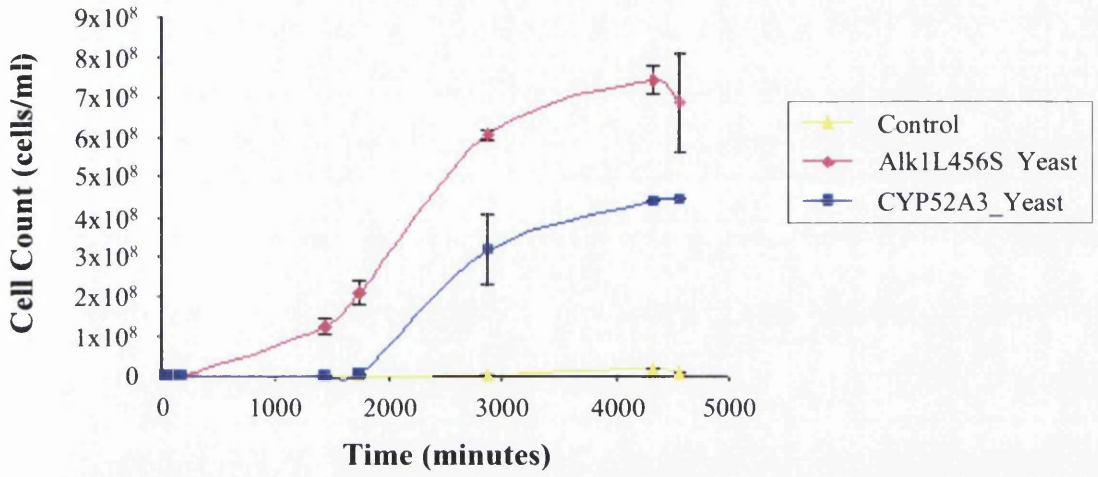
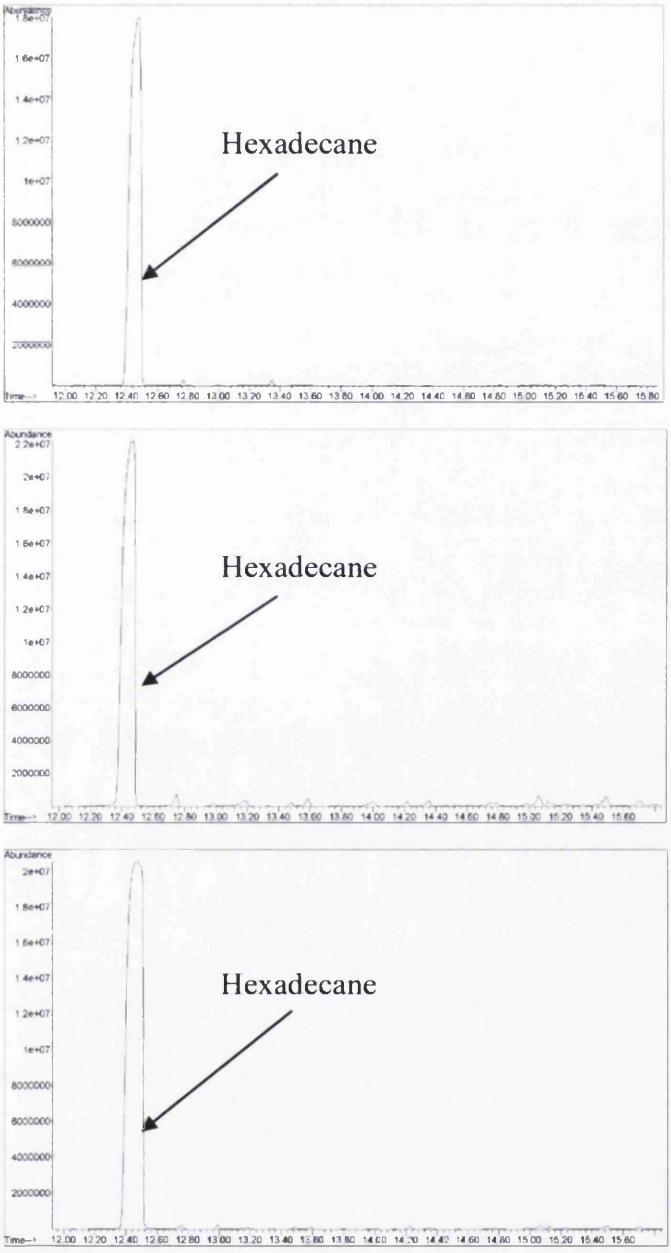


Figure 5.32 – Growth curve showing the growth of *S. cerevisiae* cells containing YEp51:Alk1L456S_yeast, YEp51:CYP52A3_yeast (positive control) or YEp51 only (negative control) in media containing hexadecane.

Both Alk1L456S_yeast and CYP52A3_yeast *S. cervisiae* cells were able to grow on media containing hexadecane suggesting that despite the inability of Alk1L456S protein to produce binding spectra with alkane, in this experiment Alk1L456S_yeast is able to convert hexadecane into a substrate that could be used by the yeast cell for growth. The negative control showed minimal growth in this media compared to the cells containing YEp51:Alk1L456S_yeast and YEP51:CYP52A3_yeast plasmids. This suggests the growth seen in these plasmid-containing cells was not due to the galactose in the media, but rather the hexadecane as *S. cerevisiae* is normally unable to grow on alkanes (Rehm & Reiff, 1981).

At an average cell count of 2.02×10^5 cells/ml, the GC chromatographs produced showed a large peak at 12.5minutes representing hexadecane (figure 5.33).



Alk1L456S yeast

CYP52A3 yeast

(Positive Control)

Negative Control

Figure 5.33 – GC chromatographs produced when *S. cerevisiae* cells were extracted after reaching an average cell count of 2.02×10^5 cells/ml. All GC chromatographs show a peak correlating to hexadecane.

All three types of *S. cerevisiae* cells showed a visibly detectable hexadecane peak on the chromatographs produced (figure 5.34). YEp51:Alk1L456S_yeast and YEp51:CYP52A3_yeast containing cells show large peaks representing fatty acid methyl ester versions of palmitic acid, 9-hexadecenoic acid, 9-octadecenoic acid and stearic acid (figure 5.34). These fatty acids probably correlate to the lipids found in the membranes of the cell. However, the negative control does not show similarly large fatty acid peaks, but do show the presence of other fatty acids in the chromatograph (figure 5.34). The GCMS detected the same fatty acid methyl esters in the negative control as in the Alk1L456S_yeast and CYP52A3_yeast samples with the additional identification of the fatty acid methyl esters: 12-methyl myristic acid, 15-methyl palmitic acid, 14-methyl palmitic acid and 14-methyl stearic acid (figure 5.34). This suggests CYP52A3_yeast and Alk1L456S_yeast were able to metabolise hexadecane resulting in the alternative lipid composition of the membranes of the cell.

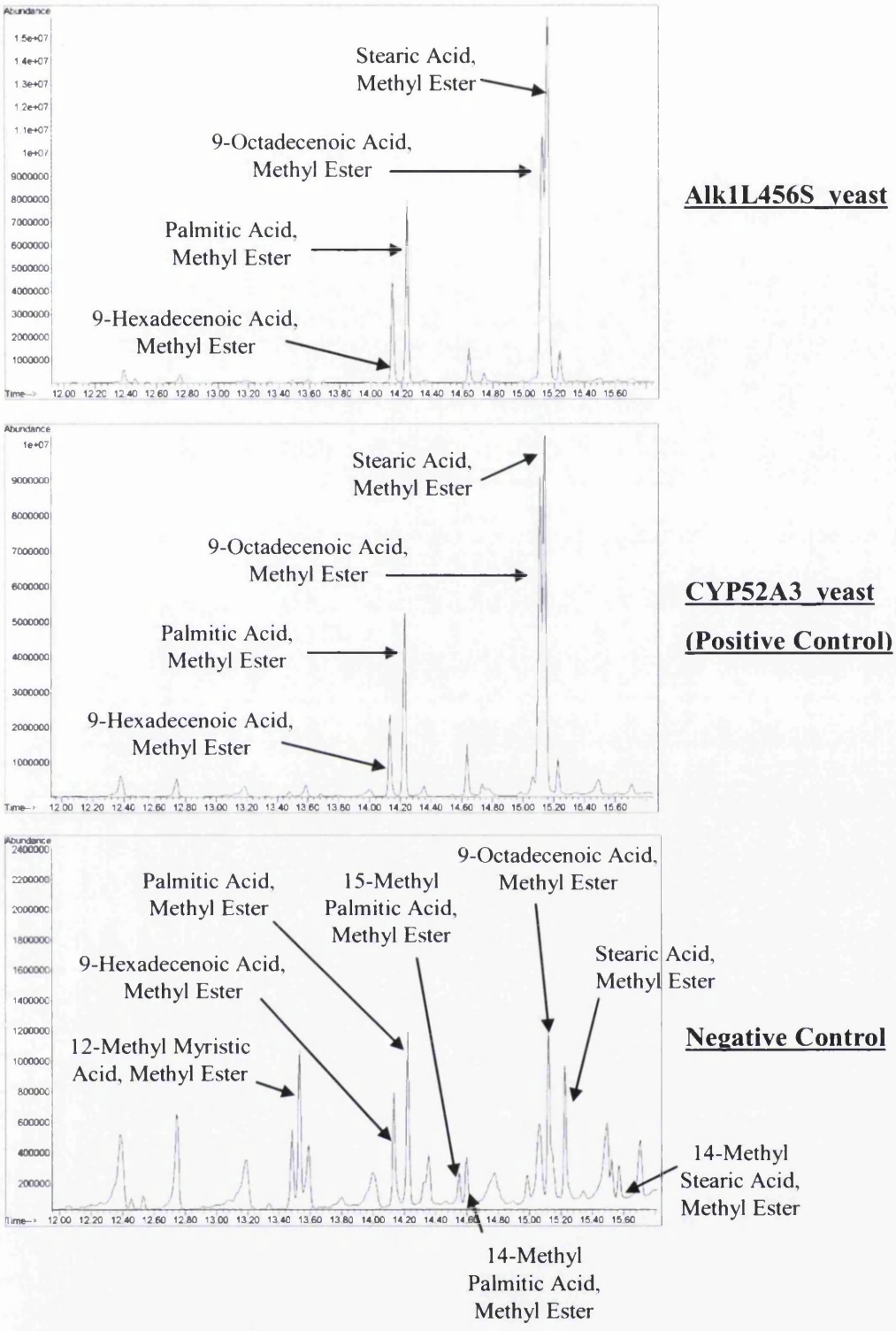


Figure 5.34 – Lipid composition of the *S. cerevisiae* cells following growth on hexadecane. Cells transformed with YEp51:Alk1L456S_yeast and YEp51:CYP52A3_yeast plasmids show similar lipid profiles, suggesting Alk1L456S was involved in the metabolism of hexadecane. The unlabelled fatty acid peaks represent silane molecules, which are indicative of the degradation of the column.

5.4 Discussion

5.4.1 Binding Studies

Despite the CYP52 family of cytochrome P450 proteins being regarded as alkane-assimilating enzymes, Alk1L456S failed to show type I binding spectra with these substrates. However, this protein did produce type I binding spectra in binding studies with fatty acids. This is not unusual as Alk8 from *C. albicans* has been shown to bind to (and metabolise) lauric acid, but not alkanes (Kim *et al.*, 2007). Although no binding spectra have been published for CYP52A13 and CYP52A17 from *C. tropicalis*, they have also been shown to preferentially metabolise fatty acids (Eschenfeldt *et al.*, 2003). This suggests that some members of the CYP52 family may preferentially bind to and metabolise fatty acids over alkanes.

Alk1L456S was shown to be able to bind to lauric acid, myristic acid and palmitic acid. Palmitic acid was shown to have the lowest substrate binding constant (K_s) value for Alk1L456S and, therefore, the greatest binding affinity of these fatty acids. Stearic acid and capric acid produced no discernable spectra with this protein. This data suggests palmitic acid is the fatty acid of choice for Alk1L456S and should be used in reconstitution assays with this protein.

Table 5.3 shows cytochrome P450 enzymes from *C. albicans* and other organisms that have been shown to bind fatty acids.

Enzyme	Organism	Ks (μM)			Reference
		Lauric Acid (C14:0)	Myristic Acid (C14:0)	Palmitic Acid (C16:0)	
CYP52A4	*C. maltosa*	>100	>100	>200	Scheller *et al.*, 1996
Alk8	*C. albicans*	5.1±0.4			Kim *et al.*, 2007
CYP4F11	Human			40±11	Tang *et al.*, 2010
CYP4A1	Rat	15.5			Chaurasia *et al.*, 1995
CYP124	*Mycobacterium tuberculosis*	>100		>100	Johnston *et al.*, 2009
CYP152A1	*Clostridium acetobutylicum*	88±15	9±0.5	10±0.5	Girhard *et al.*, 2007
CYP152A2	*C. acetobutylicum*	221±22	36±5	30±5	Girhard *et al.*, 2007
CYP102A7	*Bacillus licheniformis*	184.0±7.0	16.7±5.5	0.7±0.1	Dietrich *et al.*, 2008
Alk1L456S	*C. albicans*	3466.94±260.55	934.05±42.05	139.89±52.97	This study

Table 5.3 – Comparison of the fatty acid binding constant (K_s) values produced by other cytochromes P450 and Alk1L456S.

Of the enzymes in table 5.3, Alk1L456S was shown to bind to lauric acid, myrisitic acid and palmitic acid more weakly than any of the other cytochromes P450. However, CYP52A4 from *C. maltosa* (Scheller *et al.*, 1996), CYP124 from *M. tuberculosis* (Johnston *et al.*, 2009), CYP152A2 from *C. acetobutylicum* (Girhard *et al.*, 2007) and CYP102A2 from *B. licheniformis* (Dietrich *et al.*, 2008) all have a K_s value of over 100μM for at least one of these fatty acids (table 5.3), but have still been shown to be involved in the hydroxylation of these substrates. Interestingly, lauric acid binds to Alk8 more tightly than it does to CYP52A4 (table 5.3), despite both enzymes belonging to the CYP52 family and being able to metabolise this fatty acid. This suggests that despite the high K_s values produced by palmitic acid, myristic acid and lauric acid in binding studies with Alk1L456S, this protein is still a candidate for investigation in the metabolism of these fatty acids in further reconstitution assays.

Interestingly, many of the cytochrome P450 enzymes described above are also able to bind to other substrates with greater binding affinity. For instance, CYP102A7 is able to bind to methylated fatty acids including 12-methyl myrisitic acid (K_s = 3.5±0.6μM) and 13-methyl myristic acid (5.1±0.9μM) (Dietrich *et al.*, 2008). CYP124 is also involved in the binding of similar forms of fatty acid substrates, including 15-methyl palmitic acid (1.01±0.07μM), phytanic acid (0.22±0.006μM), farnesol (1.04±0.05μM) and geranylgeraniol (0.48±0.06μM) (Johnston *et al.*, 2009). Both enzymes were also involved in the metabolism of these hydrocarbons (Dietrich *et al.*, 2008; Johnston *et al.*, 2009). This suggests Alk1L456S may be able to interact with substrates other than alkanes and saturated fatty acids.

5.4.2 Detection of Hexadecane, Palmitic Acid and their Derivatives on the GCMS

Before some chemicals, such as fatty acids, can be detected on a GCMS they require derivatisation to make them more volatile. Making the compound more volatile means it is able to vaporise more readily, so it can travel along the column. This can be done by silylation or esterification.

In silylation, a silyl group ($Si(CH_3)_3$) can bind to the compound displacing a hydrogen atom (see figure 5.35) (Pierce, 1968). This addition makes the compound

more volatile so it can be run on a GCMS and increases the molecular weight by 71. The technique can be used to derivatise a number of involatile chemicals, including sterols and fatty acids (Pierce, 1968). However, hydroxyl side-chains are more readily silylated than carboxyl side-chains meaning silylation can result in incomplete derivatisation of this group and the TMS versions produced can be relatively unstable (Pierce, 1968). Therefore, the silylation of palmitic acid is not practical in this experiment as any product produced would not be completely derivatised, resulting in multiple peaks on the GC trace (as seen in figures 5.12 and 5.13).

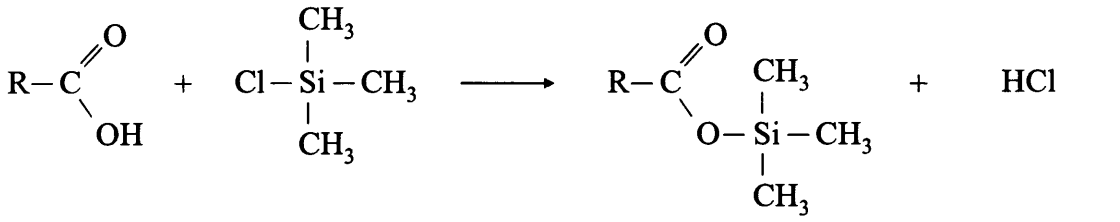


Figure 5.35 – Silylation of fatty acid by TCMS. The TCMS binds to the fatty acid at the carboxyl group displacing the hydrogen and the Cl^- on the TCMS itself. This results in the production of a TMS derivatised fatty acid molecule and hydrochloric acid, which is scavenged by pyridine added to the sample reaction.

Another method for the derivatisation of fatty acids is esterification, most commonly protocols which result in the formation of methyl esters. This reaction involves methanol interacting with the compound in a reaction catalysed by either bases (such as sodium methoxide) or acids (such as sulphuric acid) (Christie, 1989) (see figure 5.36). As with silylation, a hydrogen atom in the carboxyl side-chain is displaced as a methyl group (CH_3) binds to the oxygen (Christie, 1989) (see figure 5.36). This results in the production of a volatile compound, with an increased molecular weight (standard molecule weight plus 14), which is suitable for analysis using GCMS. Although, both sodium methoxide (base) and 3% (v/v) sulphuric acid (acid) (in the presence of methanol) catalysed the formation of palmitic acid methyl esters in this study, the acid catalysed method was used as thapsic acid could not be identified on the GCMS when the basic conditions were used. The base catalysed esterification

method can, also, result in the production of methoxy derivatives (Carrapiso & García, 2000; Murrieta *et al.*, 2003; Juárez *et al.*, 2008) and if water is not completely removed from the sample before esterification can take place, then hydrolysis can occur (Juárez *et al.*, 2008). Therefore, acid catalysed esterification was the method of choice for this study as it resulted in the complete esterification of palmitic acid and allowed the identification of thapsic acid on the GCMS.

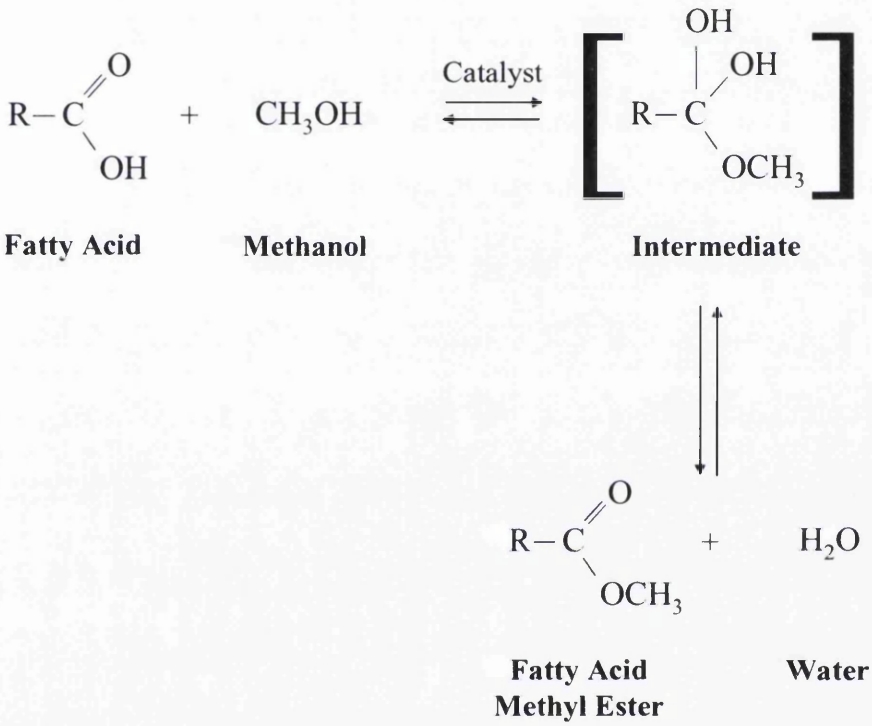


Figure 5.36 – Esterification of fatty acids. Fatty acid methyl esters are produced when methanol is used in the reaction. Other alcohols can be used to esterify fatty acids. For instance, if ethanol is used then ethyl esters are produced. This mechanism is the same for both acid (e.g. sulphuric acid) and base (e.g. sodium methoxide) catalysed esterification.

However, esterification alone was not sufficient for the analysis of 16-hydroxypalmitic acid on the GCMS. Acid catalysed esterification was required to add a methyl group to the carboxyl side chain of the fatty alcohol and silylation for the derivatisation of the hydroxyl group. Neither technique could be used as the sole method for derivatisation in this study as either palmitic acid or 16-hydroxypalmitic acid would not be detected on the GCMS if only one of these methods was used.

Therefore, all samples in this study required both acid catalysed esterification and silylation before analysis on the GCMS.

5.4.3 Reconstitution Assays

5.4.3.1 Reconstitution Assays Using Purified Alk1L456S Protein or CYP52A3_yeast Microsomes to Catalyse the Reaction

Reconstitution assays using purified Alk1L456S protein to catalyse the reaction failed to result in the detection of metabolised palmitic acid product (i.e. either 16-hydroxypalmitic acid or thapsic acid). This was either due to the inability of the enzyme to hydroxylate palmitic acid or low level product formation occurred, which could not be detected on the GCMS.

To aid detection of the substrate and any products formed in these assays, labelled substrate or chromatographic techniques, such as TLC (thin layer chromatography) and HPLC (high-pressure liquid chromatography) could be used. In this experiment, deuterated palmitic acid (palmitic acid-16,16,16-d_3) was used as substrate. This substrate contains three heavy hydrogen atoms at the ω-carbon of the palmitic acid chain producing a stable isotope of this substrate. The mass of these heavy hydrogen atoms is 2, which is twice that of the unlabelled hydrogen atoms, thus increasing the molecular mass of the palmitic acid molecule. Normally palmitic acid has a molecular weight of 256 (270 when in its methyl ester form), but when in its deuterated form the molecular weight increases to 259 (or 273). This slight increase in molecular weight is sufficient to allow different mass spectrums to be produced for each substrate, thereby allowing both substrates to be identified on the GCMS. In experiments with Alk1L456S, no substrate formation was seen when palmitic acid-16,16,16-d_3 was used as the substrate. These results suggest either Alk1L456S is not able to catalyse the hydroxylation of palmitic acid-16,16,16-d_3 or the amount of substrate used in these reactions was not sufficient for product detection. This deuterated hydrocarbon was also used in reconstitution assays with microsomal preparations of CYP52A3_yeast, however, product formation was not detected neither was the substrate. This was due to the presence of large peaks (possibly) correlating to membrane lipids, obscuring the detection of smaller peaks on the GC chromatograph. This suggests the concentration of palmitic acid-16,16,16-d_3 in the reactions was insufficient for the detection of any products produced, despite the

concentration of this substrate being increased. The lack of product detection in these assays was probably not due to the CYP52A3_yeast protein being unable to catalyse the reaction as previous experiments have shown CYP52A3 to be involved in the hydroxylation of palmitic acid to the alcohol, 16-hydroxypalmitic acid (Scheller *et al.*, 1998). A similar deuterated fatty acid, lauric acid-12,12,12-d_3, has been used in experiments with purified Alk8 protein and was shown to catalyse the hydroxylation of this substrate to 12-hydroxylauric acid (Kim *et al.*, 2007), thus suggesting the deuterated palmitic acid used in assays with Alk1L456S and CYP52A3_yeast was not the reason why product was not detected in these experiments.

Alternatively, radiolabelled substrate could be used in reconstitution assays to aid substrate detection. Many of the cytochrome P450 enzymes that have been shown to be involved in alkane/fatty acid hydroxylation have used carbon 14 (^{14}C) labelled radioactive substrate in enzymatic reactions. These enzymes include CYP52A3 (Scheller *et al.*, 1996; Scheller *et al.*, 1998), CYP94A5 from *Nicotiana tabacum* (Le Bouquin *et al.*, 2001), human CYP4F3B (Fer *et al.*, 2008) and human CYP4F11 (Tang *et al.*, 2010). This list is not exhaustive as many other cytochrome P450 proteins have also been shown to use this type of substrate. Product formation can then either be detected by HPLC or on TLC plates followed by detection on an autoradiogram or a scintillation counter (examples include Scheller *et al.*, 1996; Kikuta *et al.*, 1999 and Le Bouquin *et al.*, 2001). This method could be used to identify product formation with Alk1L456S.

The chromatographic techniques HPLC and TLC could be used to detect substrate formation not only with radiolabelled substrates, but also with unlabelled substrates. Although, TLC and HPLC are less sensitive than GCMS, they may be more ameneable to using radio-labelled substrates.

Fatty acids are ubiquitous in nature and these substrates can be detected by the GCMS from the reconstitution assay mix, although they do not interfere with the reaction. This was evident in the reconstitution assays where microsomal preparations of CYP52A3_yeast were used to catalyse the reaction as large peaks representing membrane lipids obscured the identification/detection of the peaks

correlating to the substrate and any products produced. Using TLC prior to running the samples on a GCMS would allow the fatty acids/substrates to be separated on a silica plate (Waxman, 1991) and products identified by comparison of the R_f value of known standards. The R_f value is determined by measuring the distance of the analyte from the origin (where the sample was loaded onto the plate). Solvent is used to separate the components of the substrate on the TLC plate (mobile phase), which can then be visualised using a dye that binds to the lipid fraction (e.g. 2,7-dichlorofluorescein, which binds to the lipid producing yellow spots under UV light) (Fuchs *et al.*, 2011). Using the correct mobile phase solvents (e.g. hexane/diethyl ether/acetic acid (40:60:1, v/v/v (Scheller *et al.*, 1996) would ensure the membrane lipids separate from the substrates/products depending on polarity, thus allowing each type of fatty acid to be distinguished. Usually membrane lipids are identified as fatty acids on the GCMS as fractionation occurs separating the polar head of the molecule and the fatty acid chain allowing membrane lipids to be identified as fatty acids (Chrisite, 1989).

TLC can be used on its own to identify substrates/products or alternatively the substrates/products can be removed from the TLC plate and run on a GCMS (or other mass spectrometric detectors, such as LCMS (liquid chromatography mass spectrometry)) for identification (Fuchs *et al.*, 2011). However, substrates identified using TLC may not always be detected on a GCMS as the concentration of the substrate may be too low for detection. Following reconstitution assays with CYP92B1 from *Petunia hybrida* and linoleic acid (C18:2) or linolenic acid (C18:3), metabolites were detected on a TLC plate, but they could not be identified on the GCMS (Petkova-Andonova *et al.*, 2002). Comparing these products to standards representing ω-hydroxylinoleic acid and ω-hydroxylinolenic acid on the TLC plate, it was shown that CYP92B1 is not involved in the ω-hydroxylation of these fatty acids as the R_f values for these standards and the products produced were not similar (Petkova-Andonova *et al.*, 2002). To date, these metabolites have not been identified.

Alternatively, HPLC can be used for the analysis of lipid metabolites. The samples can be loaded on the HPLC and separated on the column by (high) pressure rather than temperature (as with GCMS). The retention time for each molecule is recorded

and can be used to identify the metabolite produced. HPLC can also be coupled with a detector (such as a MS detector) to further identify the lipid fractions (Lima & Abdalla, 2002). The advantages of using this technique over GCMS are increased sensitivity and increased specificity when coupled with radio-labelled substrate, which allows the metabolites produced at low levels to be detected (Lima & Abdalla, 2002). This would be useful for the analysis of reactions catalysed by Alk1L456S as it may determine if the products formed were at low levels or if the enzyme was unable to catalyse the metabolism of the substrate. Therefore, HPLC or TLC (either coupled with the labelled or unlabelled substrate) could be used as an alternative to GCMS for the detection of metabolites following reconstitution assays using purified Alk1L456S protein or CYP52A3_yeast microsomes as the catalysts. This would then identify whether the failure to detect product in these assays was due to the inability of the protein to hydroxylate palmitic acid or if product was formed at undetectable levels.

Alternatively, modifications to the reconstitution system could be made to enhance product formation. In this study, alterations in the pH, CPR, CPR concentration and an increase in reaction time occurred, but no product was identified in reconstitution assays catalysed by Alk1L456S. Further, modifications to the reconstitution system could be made, including alteration to the lipid component. As described in section 5.1, a lipid component is required for the protein stability and correct orientation of membrane-bound cytochrome P450 and CPR proteins in the assay (Paine *et al.*, 2005; van Bogaert *et al.*, 2007). In reconstitution assays with Alk1L456S micelles were created using DLPC. These micelles have been used in experiments with other cytochromes P450 (including CYP4F11 (Tang *et al.*, 2010)) and have resulted in the formation of metabolites. However, some cytochromes P450 are unable to couple to the micelles preventing product formation. For example, rat CYP3A1 was unable to metabolise testosterone when micelles were used as the lipid component, but was able to catalyse the production of 2β-, 6β- and 15β-hydroxytestosterone in the presence of phospholipids (Eberhart & Parkinson, 1991). This suggests some cytochromes P450 are unable to couple with micelles and that this coupling is dependent on the fatty acid composition of the lipid component used (Duppel *et al.*, 1973; Eberhart & Parkinson, 1991).

In this study, yeast microsomal preparations of CYP52A3 were also used in reconstitution assays to catalyse the hydroxylation of palmitic acid and hexadecane. However, no products of the reaction were detected on the GCMS despite CYP52A3 previously being shown to metabolise these substrates (Scheller *et al.*, 1996; Scheller *et al.*, 1998). This was due to the presence of large peaks representing membrane lipids (on the GC chromatograph) obscuring the identification of other products produced. Thus microsomal preparations of Alk1L456S were not used in this study (instead of DLPC micelles) as labelled substrates would be required to distinguish between the peaks. Therefore, the inability of Alk1L456S to catalyse the metabolism of palmitic acid may be due to the protein being unable to couple with the micelles and, in turn, interact with the CPR and fatty acid correctly preventing hydroxylation. Further experiments could be undertaken using phospholipids or yeast microsomes as the lipid component of the assay.

5.4.3.2 Metabolism of Hexadecane by *S. cerevisiae* Transformed with YEp51:Alk1L456S_yeast Plasmid

As reconstitution assays containing recombinant Alk1L456S protein were unable to show enzymatic activity, YEp51:Alk1L456S_yeast plasmid was transformed into *S. cerevisiae* cells for *in vivo* experiments. These cells were shown to be able to grow on hexadecane, despite the alkane not producing any spectra in binding studies with the recombinant Alk1L456S protein. Hexadecane was used as the substrate in these experiments as *S. cerevisiae* is able to utilise fatty acids for growth, but is unable to grow on alkanes. Therefore, using hexadecane in this experiment ensured any growth of the transformed cells was due to the Alk1L456S protein being expressed rather than normal cell growth.

YEp51:CYP52A3_yeast transformed *S. cerevisiae* cells were used as a positive control as previous experiments involving CYP52A3 have shown the protein to be involved in hexadecane hydroxylation (Scheller *et al.*, 1996). In this experiment, these cells were able to grow on hexadecane and produced similar GC chromatographs as YEp51:Alk1L456S_yeast cells following extraction, suggesting Alk1L456S is involved in the hydroxylation of hexadecane as both types of cells share similar lipid membrane compositions following growth on this substrate.

When fatty acids are used for growth in *S. cerevisiae*, they are taken up into the cell and transported to the peroxisome where they are subjected to β-oxidation and are used for growth (van Roermund *et al.*, 1998; Gurvitz *et al.*, 2001). In other yeast species, which can utilise alkanes, such as the soil-dwelling *C. maltosa*, CYP52 enzymes are required to catalyse the initial conversion of the alkane to its respective alcohol (Scheller *et al.*, 1998). The alcohol can then be further converted to an aldehyde and, in turn, a fatty acid, which can then be subjected to β-oxidation (Scheller *et al.*, 1998; Fickers *et al.*, 2005). As *S. cerevisiae* is unable to utilise alkanes as it does not contain any *CYP52* genes within its genome, the growth seen with cells transformed with YEp51:Alk1L456S_yeast and YEp51:CYP52A3_yeast plasmids suggests that these genes were translated and transcribed into functional CYP52 proteins, which were involved in the conversion of hexadecane into palmitic acid. This fatty acid could then be utilised by the cell for growth as normal.

Fatty acids are not only utilised by yeast for growth, they can also be used for other cell functions, including membrane lipid formation (e.g. phospholipids and sphingolipids) because the natural hydrophobic nature of these hydrocarbons allows them to be involved in the production of membrane bilayers (Tehlivets *et al.*, 2007). Therefore, growth on a particular fatty acid can result in a membrane lipid conformation that contains a high percentage of that particular hydrocarbon. This could explain why a large peak representing palmitic acid (methyl ester) was seen in the GC traces produced by YEp51:Alk1L456S_yeast and YEp51:CYP52A3_yeast transformed cells, but not in the negative control. The plasmid transformed cells also showed large peaks on the GC chromatograph depicting (methyl ester versions of) stearic acid, 9-hexadecenoic acid and 9-ocatadecenoic acid.

In yeast, fatty acids can be subjected to elongation in the endoplasmic reticulum (Tehlivets *et al.*, 2007). This process involves two carbon atoms being added to the fatty acid molecule extending its chain length (Tehlivets *et al.*, 2007). Palmitic acid, which was produced in excess in this experiment, can be elongated to form the 18 carbon saturated fatty acid stearic acid and be used in the formation of membrane lipids. Saturated fatty acids can also be subjected to Δ^9-desaturase in the endoplasmic reticulum to produce unsaturated forms of the hydrocarbon (Tehlivets *et al.*, 2007). At carbon 9 in the saturated fatty acid chain, hydrogen atoms are removed

and a double bond is formed producing a monounsaturated version of the fatty acid that can be used to form membrane lipids (Tehlivets *et al.*, 2007). The palmitic acid and stearic acid molecules formed in this study were subjected to this process producing 9-hexadecenoic acid and 9-octadecanoic acid respectively, which were then identified on the GCMS.

The production of these saturated and unsaturated (16 and 18 carbon chain length) membrane lipids in plasmid transformed *S. cerevisiae* cells grown on hexadecane suggests functional CYP52A3 and Alk1L456S proteins were expressed in the cells and were involved in the conversion of hexadecane to palmitic acid. This is the first time Alk1L456S has been shown to be involved in alkane assimilation.

In conclusion, Alk1L456S was shown to be implicated in the hydroxylation of alkanes, despite binding studies and reconstitution assays failing to show any interaction between this protein and these substrates. *In vivo* experiments showed that *S. cerevisiae* cells containing the YEp51:Alk1L456S_yeast plasmid were able to grow in media containing hexadecane despite this organism being unable to utilise these hydrocarbons normally. The lipid profile of the plasmid containing cells was similar to the positive control (*S. cerevisiae* cells containing the plasmid YEp51:CYP52A3_yeast) showing the formation of palmitic acid from hexadecane. Not only does this suggest that Alk1L456S is involved in the hydroxylation of hexadecane, it also suggests that this alkane is a substrate of this protein despite binding studies and reconstitution assays implying the contrary.

Chapter 6: The Effects of Azole Antifungal Drugs and Sterol Substrates on Alk1L456S Protein in Binding and Metabolism Studies

6.1 Introduction

Azole antifungal drugs are used in the treatment of fungal infections, such as candidosis. The azoles work by targeting the 14α-demethylase, CYP51 in the ergosterol biosynthetic pathway. This causes inhibition of the enzyme's activity preventing the demethylation of lanosterol and in turn the production of ergosterol (Strushkevich *et al.*, 2010) (see figure 6.1). Ergosterol is required to maintain the fluidity and permeability of fungal membranes (Ghannoum & Rice, 1999; Strushkevich *et al*, 2010) and treatment with azoles depletes the concentration of this sterol in the cell. This depletion, in conjunction with an increase in the concentration of 14α-demethylated sterols, disrupts the membrane structure, preventing membrane transport (and other functions) resulting in the inhibition of fungal growth (also known as fungistasis) (Sheehan *et al.*, 1999).

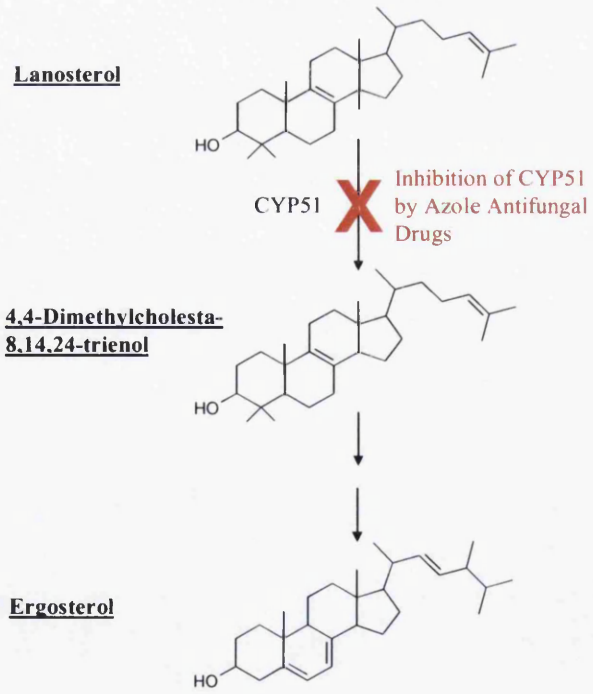
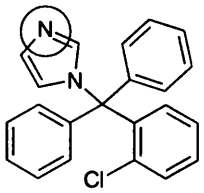


Figure 6.1 – Role of CYP51 in the ergosterol biosynthetic pathway. CYP51 is involved in the conversion of lanosterol to 4,4-dimethylcholesta-8,14,24-trienol, which can be metabolised further to produce ergosterol. Azole antifungal drugs inhibit the action of CYP51 preventing the production of ergosterol.

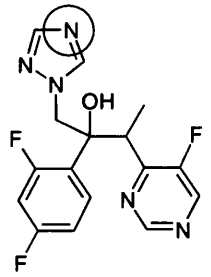
Azole antifungal drugs were first introduced in 1969 with the marketing of the topical azoles miconazole and clotrimazole (Gupta *et al.*, 1994). The combination of drug resistance and the requirement for new drugs that could be used to treat a greater range of infections and/or be administered by different routes (e.g. orally) resulted in the continued development of this class of antifungal agents (Sheehan *et al.*, 1999). There are two types of azole antifungal drugs: imidazoles (e.g. clotrimazole) and triazoles (e.g. voriconazole), which are classified according to the number of nitrogens in the azole ring (two and three respectively) (figure 6.2) (Vanden Bossche *et al.*, 1995; Lamb *et al.*, 1999). The azole attaches as the sixth ligand of the haem iron of the target enzyme either by the N-3 of the imidazole ring or by the N-4 of the triazole ring (circled in figure 6.2) (Lamb *et al.*, 1999). In binding studies with cytochrome P450, this results in the production of type II spectra due to a low-spin complex being formed. The azoles used in this study are shown in figure 6.3 and are clotrimazole, fluconazole, itraconazole, ketoconazole and voriconazole.

Imidazole



Example: Clotrimazole

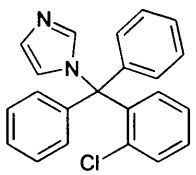
Triazole



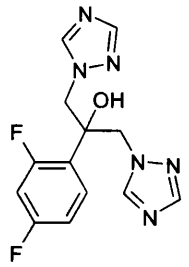
Example: Voriconazole

Figure 6.2 – Examples of imidazole and triazole drugs. Red circles indicate the nitrogen (N) atoms, which bind to the haem iron (of the cytochrome P450) as the sixth ligand.

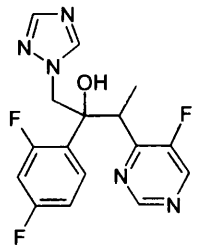
Clotrimazole



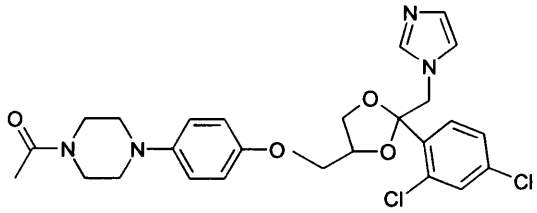
Fluconazole



Voriconazole



Ketoconazole



Itraconazole

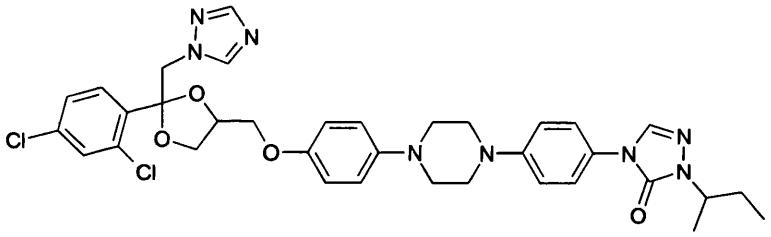


Figure 6.3 – Azole antifungal drugs used in this study. Clotrimazole and ketoconazole belong to the imidazole class of azoles. Fluconazole, voriconazole and itraconazole are triazoles.

Clotrimazole (figure 6.3) is an imidazole antifungal drug, which was first discovered in 1969 (Gupta *et al.*, 1994; Sheehan *et al.*, 1999). It is used as a topical treatment against a number of superficial infections, including mucocutaneous candidiasis (caused by *Candida* infection affecting oropharyngeal and cutaneous tissue), onychomycosis (nail infection) and dermatophytosis (fungal skin infections affecting various areas of the body, including the feet (athlete's foot/tinea pedis), facial hair (tinea barbae) and the scalp (tinea capitis)) (Sheehan *et al.*, 1999).

Ketoconazole (figure 6.3) is also an imidazole antifungal drug, however, unlike clotrimazole it can be used to treat both systemic and superficial infections as it is

available in both topical and oral preparations (Kauffman & Carver, 1997). In its topical form, ketoconazole can be used to treat onychomycosis caused by *Candida*. It can be used to treat many of the same superficial infections as clotrimazole, except it can also be used in the treatment of oesophageal and chronic mucocutaneous candidasis (Sheehan *et al.*, 1999). In its oral formulation, ketoconazole can be used to treat a number of systemic infections, including blastomycosis (rare infection caused by the inhalation of *Blastomyces dermatitidis*), histoplasmosis (caused by the inhalation of *Histoplasma capsulatum*), coccidioidomycosis (caused by *Coccidioides immitis* spores), paracoccidioidomycosis (caused by *Paracoccidioides brasiliensis*) and pseudallescheriasis (caused by *Pseudallescheria boydii*) (Sheehan *et al.*, 1999). However, the ability of the oral preparation of ketoconazole to treat these systemic and superficial infections is limited by food and gastric acidity as it is absorbed more effectively in acidic conditions (Chin *et al.*, 1995; Sheehan *et al.*, 1999).

Itraconazole (figure 6.3) is a triazole antifungal drug, which is available as an oral preparation (in tablet and liquid forms) for the treatment of superficial and systemic infections (Bartroli *et al.*, 1995). This drug can be used to treat many of the same infections as ketoconazole, except itraconazole is also used in the treatment of many other systemic infections, including aspergillosis, candidiemia, cryptococcosis and sporotrichosis (chronic skin infection caused by *Sporothrix schenckii*) (Sheehan *et al.*, 1999). As with ketoconazole, the effects of itraconazole are limited by food and gastric acidity (Sheehan *et al.*, 1999).

Fluconazole (figure 6.3) is a triazole antifungal agent, which is considered to be the most successful clinically used triazole (McLean *et al.*, 2002). It is available in both oral and intravenous preparations and can be used to treat similar superficial and systemic infections as itraconazole, however, it is not used to treat pseudallescheriasis nor infections caused by *Aspergillus* (Sheehan *et al.*, 1999). Unlike itraconazole and ketoconazole, this drug is able to readily enter the cerebral spinal fluid so it can be used in the treatment of fungal meningitis (Goa & Barradell, 1995).

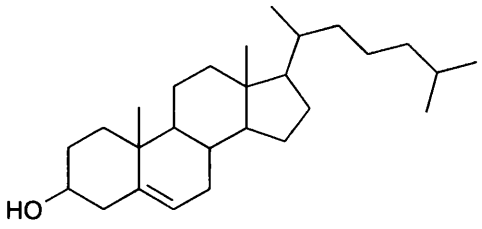
Voriconazole (figure 6.3) is a triazole antifungal drug, which is available in oral and intravenous preparations. It is a derivative of fluconazole (Ally *et al.*, 2001; Goa &

Barradell, 1995; Sheehan *et al.*, 1999). As with the other commercially available trizoles, itraconazole and fluconazole, this azole has been used to treat a wide variety of systemic fungal infections. However, it has been shown to be more potent than fluconazole against a number of infections, particularly those that are caused by *Candida* and *Aspergillus* species (Ally *et al.*, 2001; Hoffman & Rathbun, 2002; McLean *et al.*, 2002; Smith *et al.*, 2006).

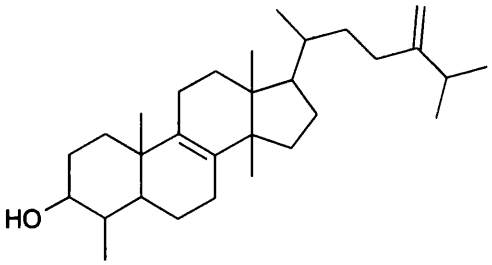
Resistance to a particular azole can occur due to the need for long-term azole treatment, particularly in patients with weakened immune systems (e.g. HIV/AIDS sufferers), thus promoting the need for alternative treatments to these infections (Hitchcock, 1993; Morio *et al.*, 2010). Interestingly, all of these azoles inhibit fungal CYP51 activity, but at therapeutic doses do not affect human CYP51s as higher concentrations are required to inhibit these enzymes (Ghannoum & Rice, 1999; Sheehan *et al.*, 1999). This is because fungal CYP51s have a greater binding affinity to the azoles developed as drugs than the human CYP51 (Ghannoum & Rice, 1999; Sheehan *et al.*, 1999).

The CYP51 family of cytochrome P450 enzymes is ubiquitous in nature. They are found not only in fungi and humans, but also in other animals, plants and some bacteria (Yoshida *et al.*, 2000). Each of these enzymes is involved in the production of membrane sterols, but the initial substrates and final products of the sterol biosynthetic pathway differ according to the organism. For example, fungal CYP51s are involved in the 14α-demethylation of lanosterol, which eventually results in the production of ergosterol, but in animals this final sterol product is cholesterol (figure 6.4). In plants, however, CYP51 is involved in the 14α-demethylation of obtusifoliol (figure 6.4), rather than lanosterol, and the sterol synthesis pathway results in the production of campesterol, stigmasterol and sitosterol (figure 6.4), instead of ergosterol or cholesterol.

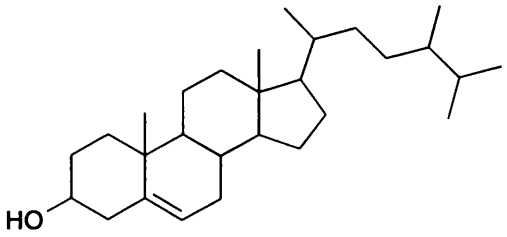
Cholesterol



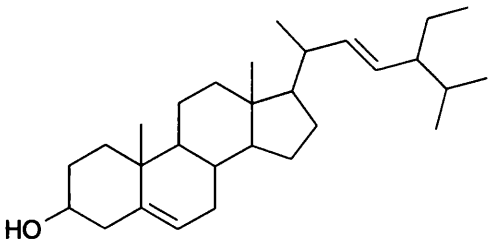
Obtusifoliol



Campesterol



Stigmasterol



Sitosterol

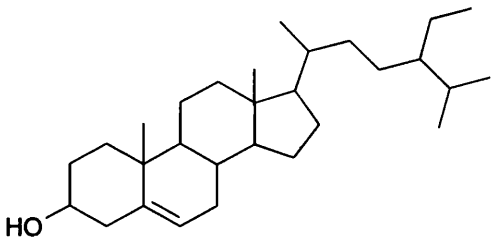


Figure 6.4 – Structures of cholesterol, obtusifoliol, campesterol, stigmasterol and sitosterol. Cholesterol is the final product of sterol synthesis. In plants, obtusifoliol is metabolised by CYP51. The sterol synthesis pathway in this organism results in the production of campesterol, stigmasterol and sitosterol.

In fungi, another cytochrome P450 enzyme, CYP61 has also been shown to be involved in the ergosterol biosynthetic pathway (Kelly *et al.*, 1995; Kelly *et al.*, 1997). It is a C22-desaturase, which catalyses the metabolism of ergosta-5,7-dienol (figure 6.5) in the formation of ergosterol (Kelly *et al.*, 1995; Kelly *et al.*, 1997). It is not the target of azole antifungal drugs (unlike CYP51), however, its activity has been shown to be inhibited by fluconazole and ketoconazole, suggesting azoles may also act on this enzyme during treatment to prevent ergosterol production (Kelly *et*

al., 1997). To date, no other fungal cytochromes P450 have been shown to bind or metabolise sterols.

Ergosta-5,7-dienol

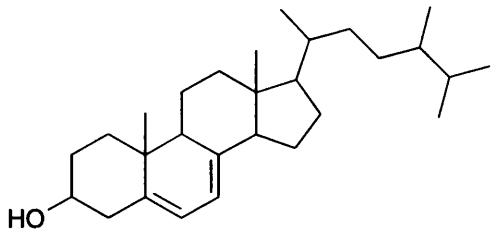


Figure 6.5 – Structure of ergosta-5,7-dienol.

CYP51 and CYP61 have been shown to be inhibited by azole antifungal drugs. To evaluate whether Alk1L456S can bind to these drugs the azoles in figure 6.3 will be used in binding studies with this protein. Although, Alk1L456S is not thought to be a target of these drugs, binding studies will allow the calculation of the $[azole]_{0.5}$ to occur, which will allow the binding affinity for each azole to Alk1L456S to be determined. This may indicate whether this cytochrome P450 has a potential role in the inhibition of cell growth, such as on lipid or alkane substrate, or in the resistance of *Candida albicans* to azole antifungal drugs. Although sterols have been shown to be the substrates of fungal CYP51 and CYP61 in this chapter, in other species, such as *Mycobacterium tuberculosis*, cytochrome P450 enzymes other than CYP51 (or other enzymes involved in sterol synthesis) have been shown to bind and metabolise sterols (e.g. CYP125) (Capyk *et al.*, 2009; Ouellet *et al.*, 2010). Therefore, these substrates will be used in binding studies with Alk1L456S to determine whether they can bind to this protein. Sterols that exhibit type I binding, which is an indicator of a potential substrate will be used in reconstitution assays to assess whether AlkL456S is involved in the metabolism of sterol.

6.2 Materials and Methods

6.2.1 Azoles – Stock Solutions for Azole Binding to Alk1L456S

Azole antifungal drugs were dissolved in dimethyl sulfoxide (DMSO) to a final concentration of 0.05mg/ml. Stored at -20°C.

6.2.2 Sterols – Stock Solutions for Sterol Binding to Alk1L456S

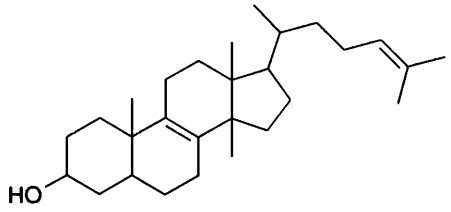
The sterols used for binding were dissolved in a solution of 200µl chloroform, 2ml acetone and 5% (v/v) Tween 80. The solution was evaporated under nitrogen gas until all the solvent was removed. The residue was resuspended in water to produce a stock sterol solution of 0.5mg/ml. Stored at 4°C.

Figure 6.6 shows the structural overview of the sterols used in this study.

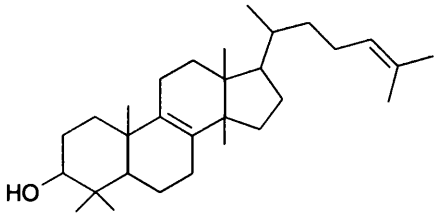
6.2.3 Sterol Standards

The sterols: cholesterol, 14-methyl-zymosterol and 32-cyclopropylidenemethyl-dihydrolansterol were made up in chloroform to a final concentration of 2mg/ml. 5µl of each standard was evaporated to complete dryness in the SpeedVac, silylated (section 2.2.12) and run on the GCMS (section 2.2.13.2). For 32-cyclopropylidenemethyl-dihydrolanosterol the final temperature (280°C) was held for 40 minutes.

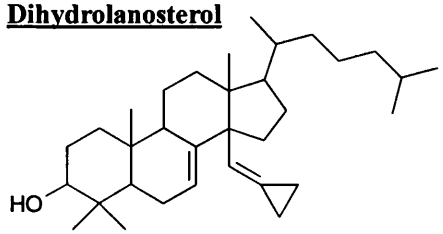
14-Methyl-Zymosterol



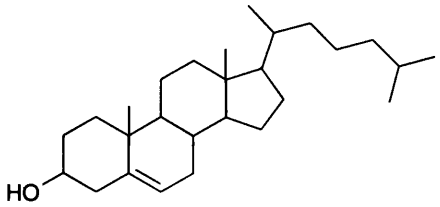
Lanosterol



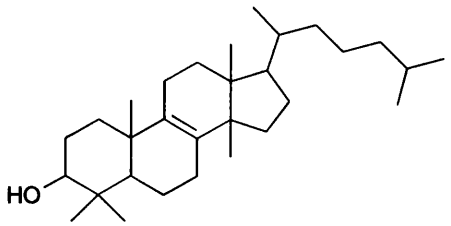
32-Cyclopropylidenemethyl-Dihydrolanosterol



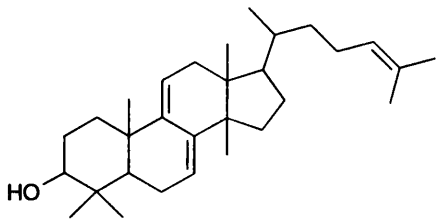
Cholesterol



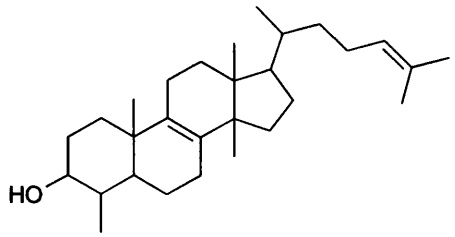
Dihydrolanosterol



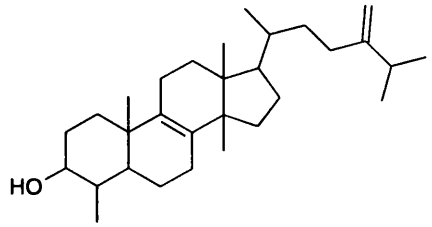
Argosterol



31-Nor-Lanosterol



Obtusifoliol



Δ^7-Dihydrolanosterol

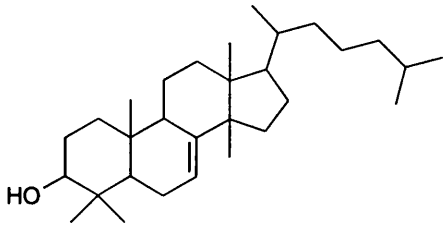


Figure 6.6 – Sterols used in this study were kindly provided by Prof. David Nes, Texas Tech University.

6.2.4 Methods

6.2.4.1 Substrate Binding

All binding studies were undertaken using a Hitachi U-3310 UV-Vis scanning spectrophotometer.

6.2.4.1.1 Azole Binding to Alk1L456S

Azoles were individually titrated against 1ml of Ni^{2+}-NTA agarose purified Alk1L456S protein (at a final concentration of 1nmole/ml of protein) in a split cuvette. A reference sample was also set-up in a second split cuvette with DMSO being added to the cytochrome P450. Buffer containing 0.1M Tris HCl, pH8.1 and 25% (w/v) glycerol was used for all spectral determinations. Spectra were measured between 350 and 550nm and until saturation with azole drug was reached.

6.2.4.1.2 Sterol Binding to Alk1L456S

The sterols in section 6.2.1.2 were titrated against 2μM of purified Alk1L456S protein. 5% (v/v) Tween 80 was added to the same concentration of protein in the reference cuvette. 0.1M Tris HCl, pH8.1 and 25% (w/v) glycerol buffer was used for all spectral determinations. Spectra were measured between 350 and 550nm until saturation with the sterol had been reached.

CYP51 protein from *C. albicans* was used as a control in these binding studies. CYP51 protein was a gift from Dr. Andrew Warrilow, Swansea University.

6.2.4.2 Reconstitution Assays

Micelles were created according to the method in section 2.2.11.1 using 10μl of 2mg/ml of lanosterol, 14-methyl zymosterol, 32-cyclopropylidenemethyl-dihydrolanosterol or cholesterol dissolved in chloroform.

All assays were set-up according to the method described in section 2.2.11.2 with the addition of 2μM of purified Alk1L456S or CYP51 protein and either 10μM of ScΔ33:CPR or 3.8μM CaCPR. All samples were saponified (section 2.2.11.3), silylated (section 2.2.12) and analysed on the GCMS (section 2.2.13.2).

Reconstitution assays were also undertaken using CYP51 as a positive control. Negative controls were set-up with either the purified cytochrome P450 or CPR protein omitted.

6.3 Results

6.3.1 Azole Binding

The binding of azole antifungal drugs to Alk1L456S resulted in the production of type II binding spectra (figure 6.7) implying that these azoles may be active in the inhibition of Alk1L456S activity.

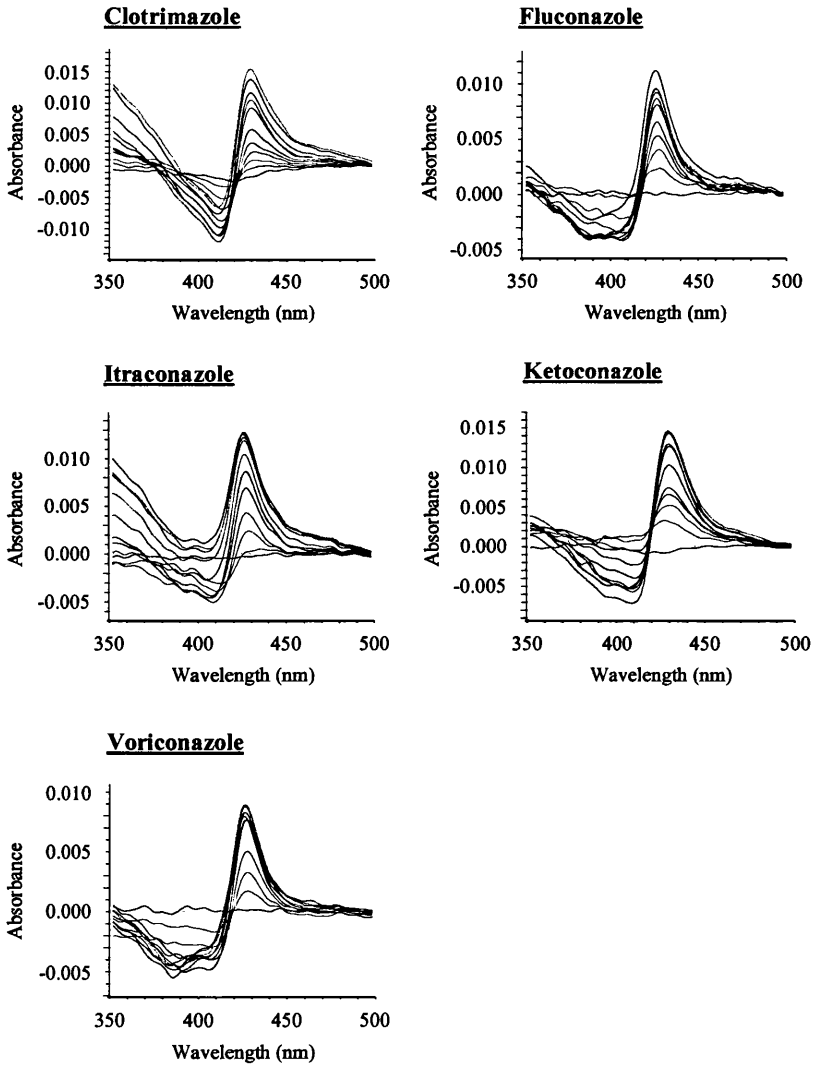


Figure 6.7 – Spectra produced in binding studies with Alk1L456S and azole antifungal drugs. All the azoles used in this study produced type II binding spectra when titrated against Alk1L456S.

To determine if tight 1:1 binding occurred the $[azole]_{0.5}$ value for each azole was calculated. $[Azole]_{0.5}$ is the azole concentration at which half δAmax for a known protein concentration is generated. A proportional increase in this value to the cytochrome P450 protein concentration indicates tight binding. At 1μM Alk1L456S the $[clotrimazole]_{0.5}$ value was 0.6μM. Increasing the protein concentration five fold to 5μM resulted in a $[clotrimazole]_{0.5}$ value of 2.5μM. This increase in $[clotrimazole]_{0.5}$ was proportional to the increase in Alk1L456S concentration suggesting tight binding between Alk1L456S and the azoles used in this study.

$[Azole]_{0.5}$ values for each azole are shown in table 6.1. This data can be used to determine the binding affinity of these azoles to Alk1L456S where the more tightly bound the azole is to the enzyme, the smaller the binding constant. Using the data shown in table 6.1, it is possible to rank the azole binding affinity for Alk1L456S from the most tightly bound to the least:

Ketoconazole > Itraconazole > Clotrimazole > Voriconazole > Fluconazole

Azole	**$[Azole]_{0.5}$ (μM)**
Ketoconazole	0.37±0.06
Itraconzole	0.40±0.10
Clotrimazole	0.58±0.19
Voriconazole	0.72±0.08
Fluconazole	1.04±0.33

Table 6.1 – $[Azole]_{0.5}$ values of the azole-Alk1L456S complex.

6.3.2 Sterol Binding

Figure 6.8 shows the spectra produced on adding sterol to Alk1L456S and CYP51.

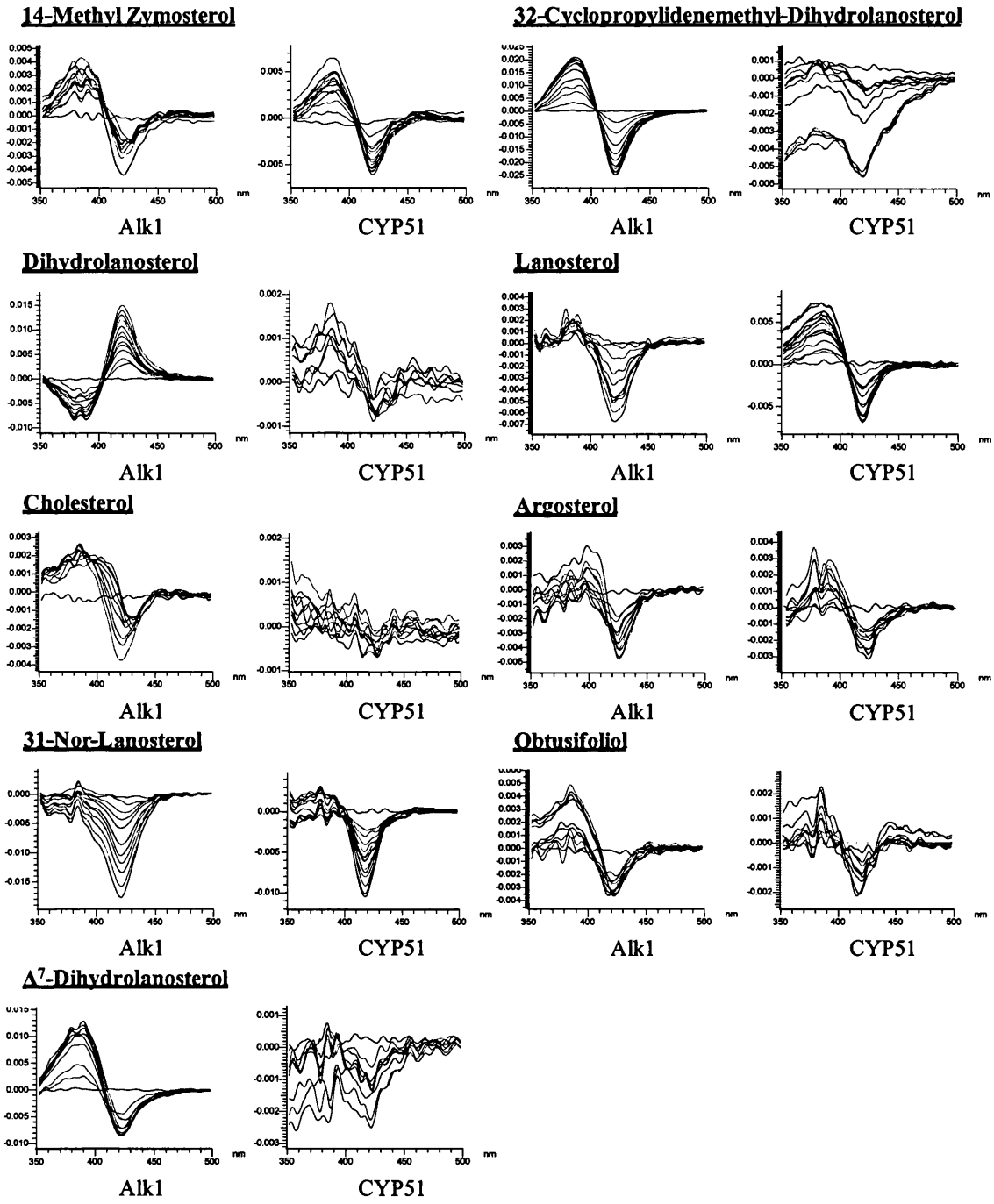


Figure 6.8 – Spectra produced when different sterols were titrated against CYP51 (control) and Alk1L456S. CYP51 showed type I binding with 14-methyl zymosterol, dihydrolanosterol, 31-nor-lanosterol, lanosterol, argosterol and obtusifoliol. Cholesterol, Δ^7-dihydrolanosterol and 32-cyclopropylidenemethyl-dihydrolanosterol showed no binding. Alk1L456S showed type I binding with all the sterols used in this study, except dihydrolanosterol, which produced reverse type I spectra.

The sterols used in this study showed type I binding with Alk1L456S, except for dihydrolanosterol, which showed reverse type I (also known as modified type II) (figure 6.8). This spectra is a mirror image of type I spectrum with a peak at 419nm and trough at 379nm (figure 6.8). This data suggests that the sterols used in this study are potential substrates of Alk1L456S as this shift is very often predictive of cytochrome P450 substrates.

Interestingly, not all the sterols used in this study were shown to bind to CYP51. As expected lanosterol binding to CYP51 showed type I binding spectra as did 14-methyl-zymosterol, argosterol, 31-nor-lanosterol, obtusifoliol and dihydrolanosterol (albeit weakly) (figure 6.8). However, cholesterol, Δ^7-dihydrolanosterol and 32-cyclopropylidenemethyl-dihydrolanosterol did not produce any binding spectra when titrated against CYP51.

The Michaelis-Menten equation was used to calculate the substrate binding constant (K_s) for each sterol. This data is shown in table 6.2.

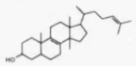
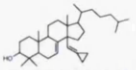
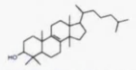
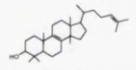
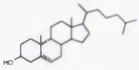
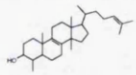
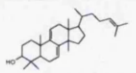
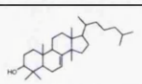
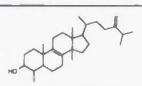
Sterol	Structure	Alk1		CYP51	
		Binding	Ks (μM)	Binding	Ks (μM)
14-Methyl-Zymosterol		Type I	2.87±0.86	Type I	3.84±0.28
32-Cyclopropylidenemethyl-dihydrolanosterol		Type I	4.77±0.60	No Binding	
Dihydrolanosterol		Reverse Type I	5.40±0.51	Weak Type I	1.76±0.67
Lanosterol		Type I	5.21±0.53	Type I	5.28±0.21
Cholesterol		Type I	4.72±1.55	No Binding	
31-Nor-Lanosterol		Type I	9.22±1.42	Type I	9.61±0.86
Argosterol		Type I	5.78±2.18	Type I	4.85±0.68
Delta-7-Dihydrolanosterol		Type I	2.63±0.71	No Binding	
Obtusifoliol		Type I	2.80±0.478	Weak Type I	3.12±1.20

Table 6.2 – Substrate binding constants (K_s) of the sterol-Alk1L456S and sterol-CYP51 complexes.

Table 6.2 shows that Alk1L456S and CYP51 share similar binding affinities for lanosterol and 31-nor-lanosterol. However, CYP51 showed greater affinity for dihydrolanosterol and argosterol, whereas Alk1L456S showed greater affinity for obtusifoliol and 14-methyl-zymosterol (table 6.2). It is important to note the weak nature of the binding between CYP51 and the sterols dihydrolanosterol and obtusifoliol interfered with the accurate determination of K_s.

Using data from table 6.2 it is possible to rank the binding affinities of each sterol for Alk1L456S. The sterols are ranked from the sterol with the strongest binding affinity to the weakest:

Δ^7-dihydrolanosterol > obtusifoliol > 14-methyl-zymosterol > cholesterol > 32-cyclopropylidenemethyl-dihydrolanosterol > dihydrolanosterol > argosterol > 31-nor-lanosterol

Lanosterol and 14-methyl-zymosterol were used for reconstitution assays (despite Δ^7-dihydrolanosterol having the greatest binding affinity to Alk1L456S) as both these sterols showed similar affinities for Alk1L456S and CYP51. Cholesterol was chosen for reconstitution assays because it has been shown to be the substrate of other cytochromes P450, such as CYP125 from *M. tuberculosis,* which also metabolise fatty acids (Capyk *et al.*, 2009; Ouellet *et al.*, 2010). 32-cyclopropylidenemethyl-dihydrolanosterol was used in reconstitution assays with Alk1L456S due to its unique structure (compared to other sterols used in this study).

6.3.3 Detection of Sterols on the GCMS

The sterol standards were run on the GCMS to ensure they could be detected for analysis.

6.3.3.1 Cholesterol

Cholesterol has a molecular weight of 386, which when silylated becomes a mass ion of 458. The retention time is 25.444 minutes (figure 6.9).

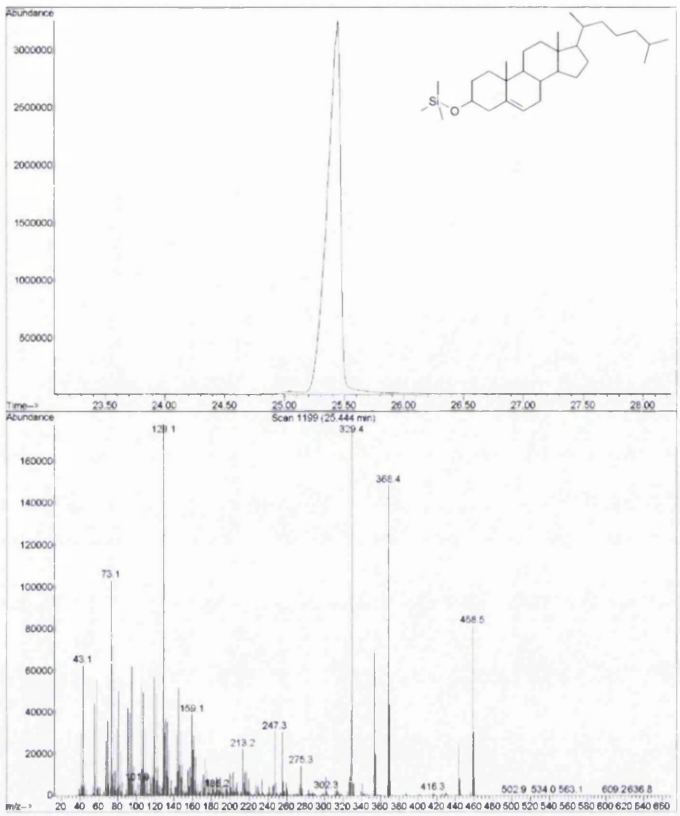


Figure 6.9 – TMS-derivatised cholesterol. GCMS results showing the GC trace/chromatograph (top) and mass spectrum (bottom) for cholesterol. Inset (top) is the TMS-derivatised structure of the standard.

6.3.3.2 14-Methyl Zymosterol

Silylated 14-methyl-zymosterol has a mass ion of 470 and a retention time of 27.059 minutes (figure 6.10).

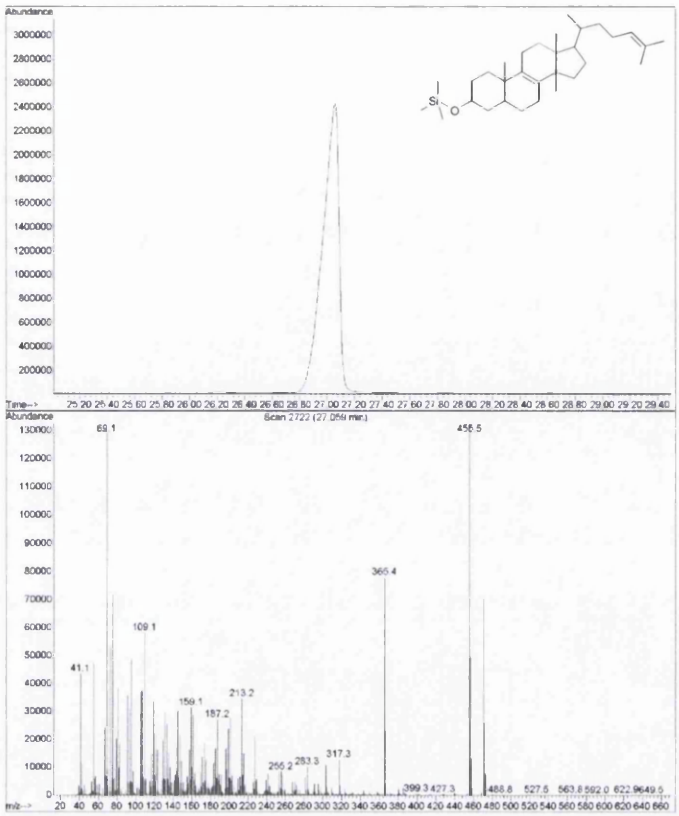


Figure 6.10 – TMS-derivatised 14-methyl zymosterol. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) of TMS-derivatised 14-methyl zymosterol. Inset (top) is the structure of this standard.

6.3.3.3 32-Cyclopropylidenemethyl-Dihydrolanosterol

The GCMS oven program was increased to ensure 32-cyclopropylidenemethyl-dihydrolanosterol could be detected as it has a larger molecular weight (MW = 466) compared to the other sterols. The silylated version of this sterol has a mass ion of 538 and a retention time of 42.038 minutes (figure 6.11).

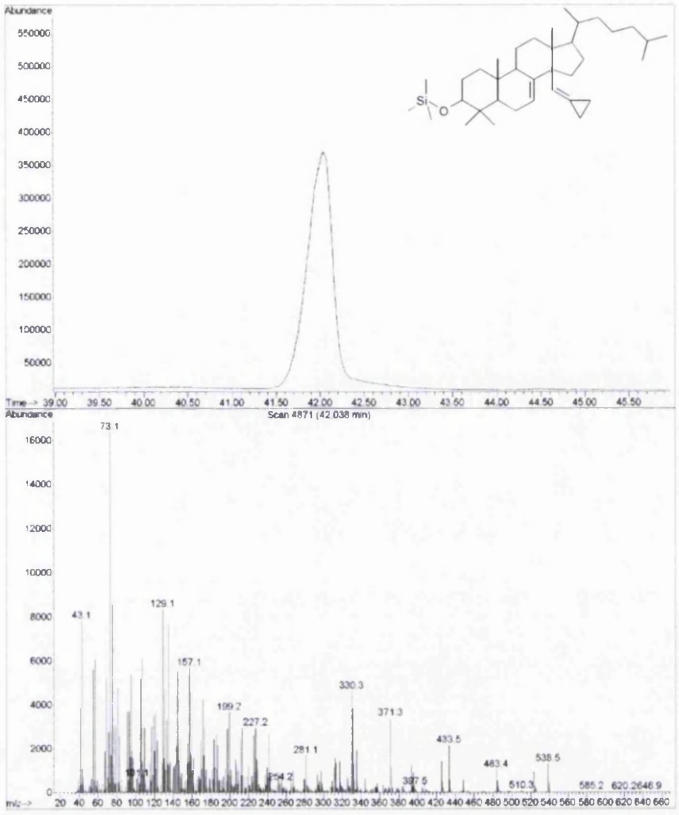


Figure 6.11 – TMS-derivatised 32-cyclopropylidenemethyl-dihydolanosterol. GCMS results showing the GC chromatograph (top) and mass spectrum (bottom) of this standard. Inset (top) is the TMS-derivatised structure.

6.3.4 Reconstitution Assays

When Alk1L456S was incubated with lanosterol no product formation was seen in the presence of ScΔ33CPR and CaCPR (figure 6.12). In contrast, CYP51 was able to demethylate lanosterol under the same assay conditions (figure 6.12). The use of CaCPR resulted in a 2.5 fold increase in product formation from 8% (using ScΔ33:CPR) to 20%.

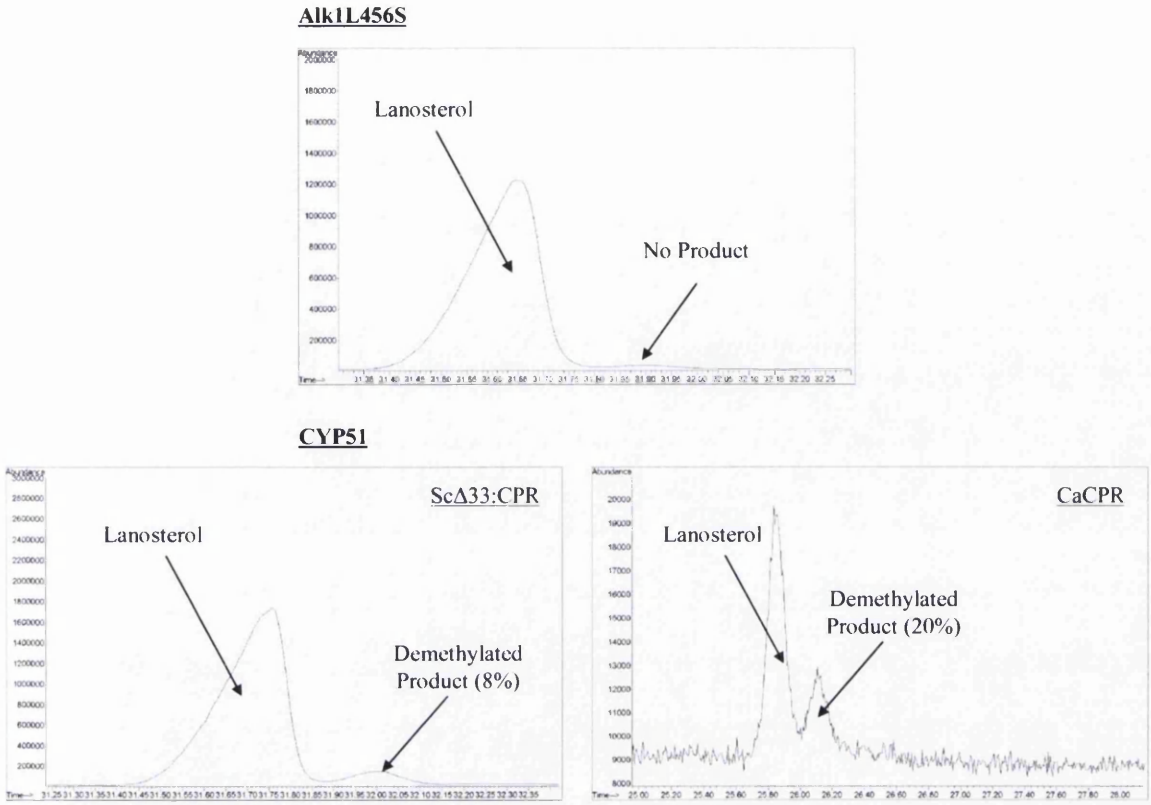
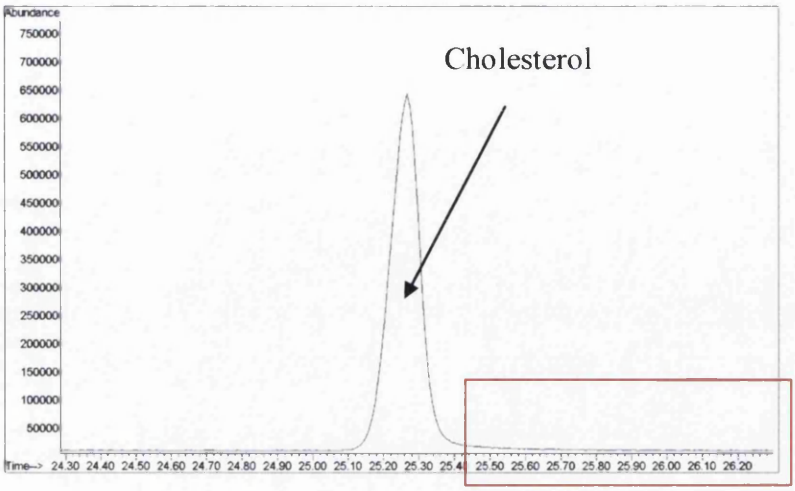


Figure 6.12 – GC chromatographs for reconstitution assays using lanosterol as the substrate. Reconstitution assays involving CYP51 show the presence of a peak representing the demethylated version of lanosterol. No product peak was seen when Alk1L456S was used to catalyse the reaction.

Neither Alk1L456S nor CYP51 (in the presence of ScΔ33:CPR or CaCPR) were shown to produce demethylated or hydroxylated products of cholesterol (figure 6.13), 14-methyl-zymosterol (figure 6.14) or 32-cyclopropylidenemethyl-dihydrolanosterol (figure 6.15) when these sterols were used as substrates in reconstitution assays.

Alk1L456S



CYP51

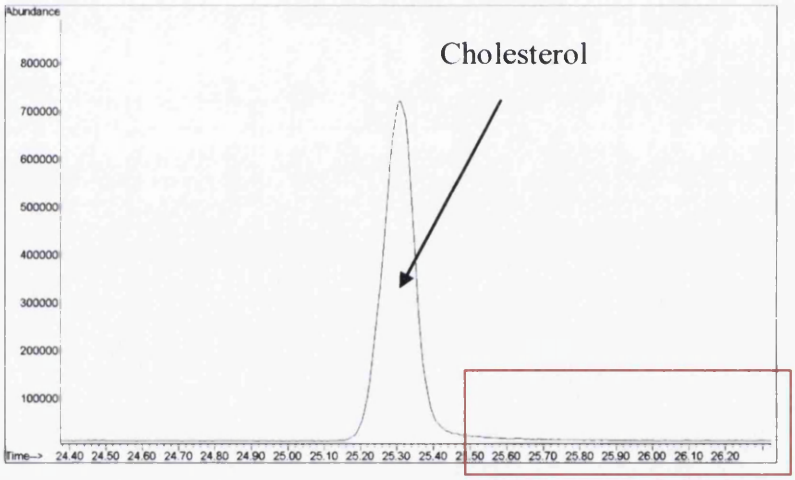
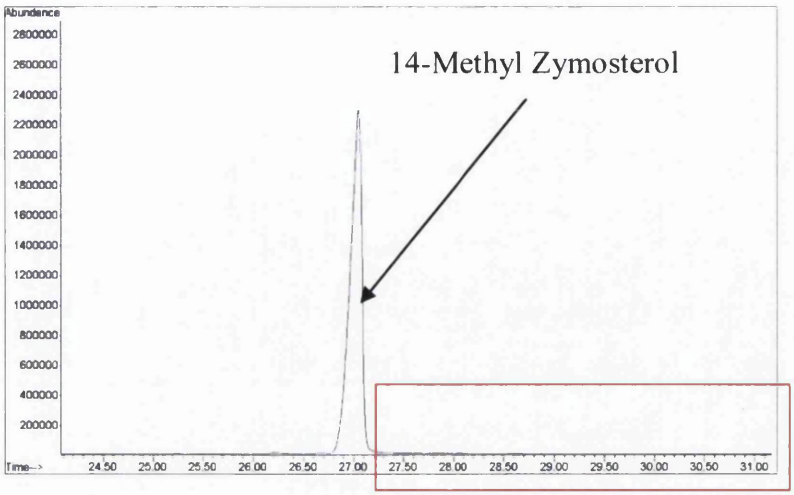


Figure 6.13 – GC chromatographs for reconstitution assays using cholesterol as the substrate. No product seen. The red square shows the potential location of peaks representing either demethylated or hydroxylated cholesterol.

Alk1L456S



CYP51

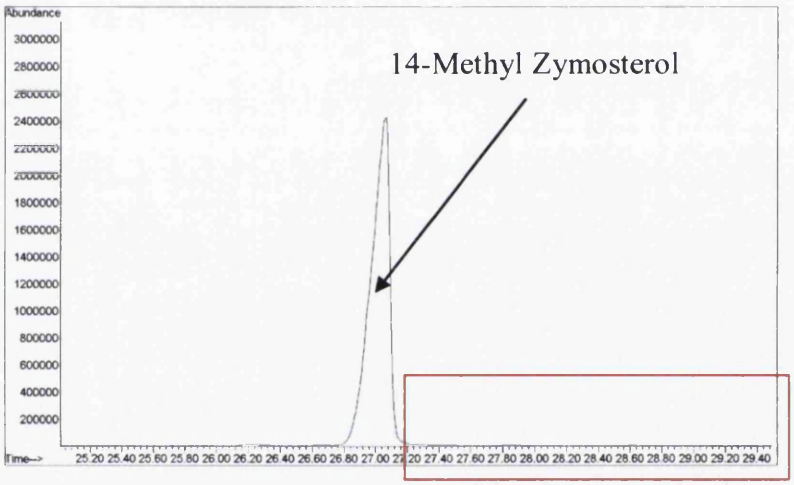
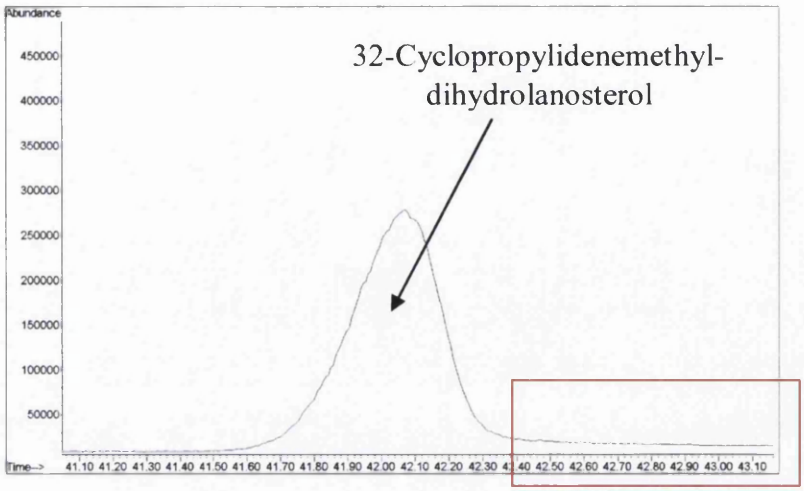


Figure 6.14 – GC chromatographs for reconstitution assays using 14-methyl zymosterol as the substrate. No product produced. The red square shows the potential location of peaks representing either demethylated or hydroxylated 14-methyl zymosterol.

Alk1L456S



CYP51

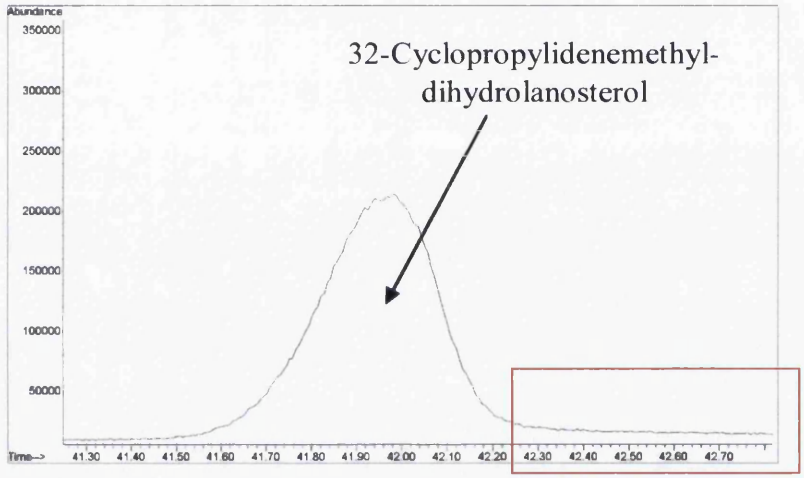


Figure 6.15 – GC chromatographs for reconstitution assays using 32-cyclopropylidenemethyl-dihydrolanosterol as the substrate. No product formation. The red square shows the potential location of peaks representing either demethylated or hydroxylated 32-cyclopropylidenemethyl-dihydrolanosterol.

6.4 Discussion

6.4.1 Azole Binding

Azole antifungal drugs are known to inhibit CYP51 activity, preventing ergosterol production and resulting in fungistasis (Strushkevich *et al.*, 2010). They bind tightly with these enzymes producing type II spectra binding data, which is indicative of inhibition. Binding experiments in this study with these drugs and Alk1L456S also showed similar results, i.e. tight type II binding, suggesting azoles are potential inhibitors of Alk1L456S. Table 6.3 shows the $[azole]_{0.5}$ data for azole binding to

Alk1L456S compared to those for CYP51 from *C. albicans* and *Phanerochaete chrysosporium*.

Enzyme	Species	$\|Azole\|_{0.5}$ (μM)					Reference
		$\|Clot.\|_{0.5}$	$\|Fluc.\|_{0.5}$	$\|Itra.\|_{0.5}$	$\|Keto.\|_{0.5}$	$\|Vori.\|_{0.5}$	
Alk1L456S	*C. albicans*	0.58	1.04	0.40	0.37	0.72	This Study
CYP51	*C. albicans*		0.11		0.13		Kelly *et al.*, 1999; Lamb *et al.*, 1999; Warrilow *et al.*, 2008
CYP51	*P. chrysosporium*	0.54	0.56	0.51	0.56	0.49	Warrilow *et al.*, 2008

Table 6.3 – Comparison of $[Azole]_{0.5}$ values for Alk1L456S, CYP51 from *C. albicans* and CYP51 from *P. chrysosporium*.

The available $[azole]_{0.5}$ data shown in table 6.3 shows that the azoles fluconazole and ketoconazole have a greater binding affinity for CYP51 from *C. albicans* than Alk1L456S. This suggests that despite the ability of azoles to bind to AlkL456S, CYP51 is still the main target of these drugs in *C. albicans*. Interestingly, Alk1L456S shows smaller $[Azole]_{0.5}$ values for itracontrazole and ketoconazole than CYP51 from *P. chrysosporium* suggesting a potential role for Alk1 in antifungal treatment. For instance, if Alk1 was expressed under certain conditions (e.g. growth in a lipid-rich environment) azoles may inhibit the activity of this enzyme. This could prevent growth of the organism on hydrocarbons or pathogenic processes that require lipid metabolism.

In chapter 5, reconstitution assays failed to show functional Alk1L456S protein activity so IC_{50} experiments could not be undertaken with this protein to determine if these azoles do inhibit Alk1L456S. However, Kim and co-workers (2007) showed that Alk8 (CYP52A21) from *C. albicans* was not inhibited by fluconazole, but weakly inhibited by ketaconazole (IC_{50} = 104μM). This suggests that CYP52 enzymes may not be a target of azole antifungal drugs (Kim *et al.*, 2007). The weak inhibition of Alk8 by azoles was also shown by Panwar and his colleagues (2001) who found a 100 fold excess of itraconazole was required to inhibit lauric acid oxidation in a *C. albicans* strain overexpressing Alk8. No azole binding data is

available for this enzyme and Alk1L456S has not been used in IC_{50} experiments so no direct comparison can be made regarding the interaction of azole drugs between these enzymes. As the affinity for different alkanes and fatty acids differs between CYP52 members from the same organism, the affinity for azoles may also differ between these same enzymes. Furthermore, no other published data have shown interactions between other CYP52 enzymes and azole antifungal drugs, suggesting more work needs to be done in this area as the CYP52 family of cytochromes P450 may influence azole antifungal treatment and resistance.

Yeast and fungi, such as *C. albicans*, can become resistant to azole antifungal drugs. The mechanisms involved in this resistance include alterations in the CYP51 sequence (whereby the enzyme remains active, but resistant to azoles), CYP51 overexpression and an increase in efflux (ABC and major facilitator superfamily) transporter activity (Ghannoum & Rice, 1999). There are also considered to be other mechanisms by which antifungal resistance is influenced (including alterations in other enzymes of the sterol biosynthetic pathway), but these mechanisms are largely unknown. Panwar and co-workers (2001) suggest that Alk8 may act as one of these mechanisms as they hypothesise CYP52 enzymes may be involved in the modification of azole antifungal drugs into potentially harmless substrates causing resistance of the cell to these drugs. Although, this has yet to be proven, another cytochrome P450 enzyme has been shown to have an ability to modify these drugs in a fashion similar to that proposed in this hypothesis.

CYP3A4 from liver extract has been shown to produce type II binding spectra in studies with azole drugs. Fluconazole binding to CYP3A4 was shown to produce a K_d value of 9.8±0.5μM (9.4 fold higher than Alk1L456S), but CYP3A4 had a similar binding affinity for itraconazole as Alk1L456S (0.456±0.008μM) (Locuson *et al.*, 2007). Interestingly, CYP3A4 was also shown to be able to metabolise itraconazole, but not fluconazole (Locuson *et al.*, 2007). This suggests that the hypothesis of Panwar and co-workers (2001) may be correct as at least one cytochrome P450 enzyme has been shown to metabolise an azole antifungal drug. Alternatively, CYP52 enzyme may not be involved in azole metabolism, but rather bind to these drugs, thus removing them from the immediate environment. This would mean an increased dosage of the azole may be required to inhibit CYP51 activity in the cell.

However, more research needs to be undertaken to understand how CYP52 enzymes and azole antifungal drugs interact before assigning a role for CYP52s in azole resistance.

6.4.2 Sterol Binding and Reconstitution Assays

Sterols are involved in the production of membrane lipids, which are required for membrane permeability and fluidity (Ghannoum & Rice, 1999; Strushkevich *et al.*, 2010). In yeast and fungi, CYP51 and CYP61 have been shown to be the only cytochromes P450 involved in the sterol biosynthetic pathway. No other fungal cytochromes P450 to date have been shown to either bind or metabolise sterols. Binding studies in this chapter have shown that Alk1L456S is able to bind to a variety of sterols producing type I or reverse type I spectra. This is the first time a member of the CYP52 family has been shown to bind these substrates. Interestingly, CYP51 did not bind to all of the sterols used, suggesting that Alk1L456S and CYP51 do not bind to the substrate in the same way and proving that there is a structural basis for sterol binding to cytochromes P450.

Figure 6.16 shows the similar/dissimilar areas of the sterol structures (used in this study) that produce the different types of binding spectra seen with CYP51.

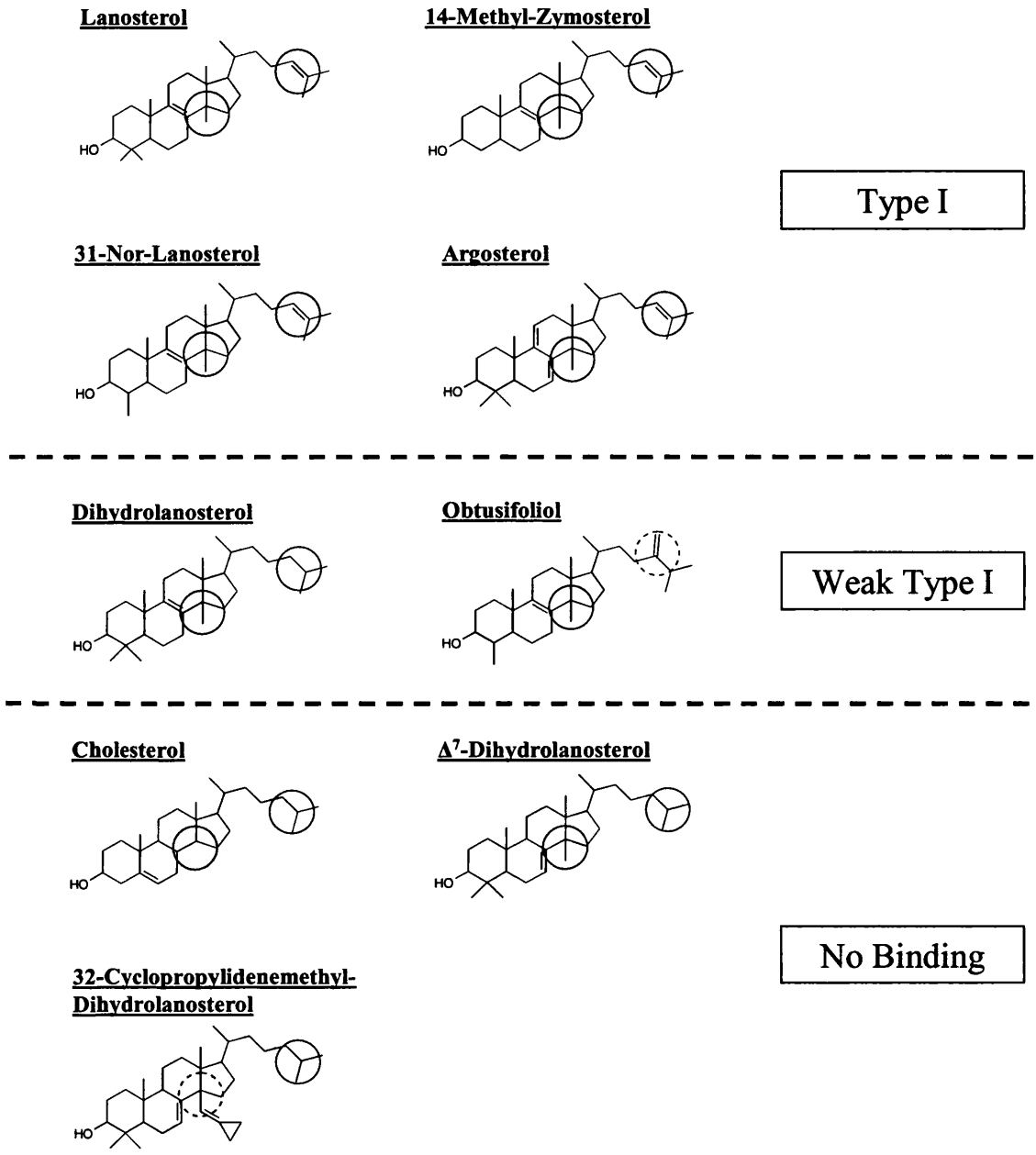


Figure 6.16 – Binding spectra produced by each sterol titrated against CYP51. Sterols are grouped by the binding spectra produced. Red circles show areas of similarity in the alkyl chains of sterols that produce type I binding spectra. Pink circles show alkyl chain structures that do not match those highlighted by the red circles, but are similar in structure. The dashed pink circle shows an alkyl side chain that does not fit into either group. The blue circle indicates the 14-methyl group in these sterol structures. Green circles highlight sterols that do not contain this structure. The dashed green circle shows sterols that do not fit into either category.

CYP51 enzymes metabolise sterols that contain a methyl group (-CH_3) at carbon 14 (shown as a blue circle in figure 6.16), e.g. lanosterol. It would be expected that all sterols with a methyl group at this position would produce type I spectra where used in binding studies with CYP51 as this side group interacts with the haem. However, Δ^7-dihydrolanosterol produced no binding spectra despite containing this methyl group at carbon 14 suggesting other areas of the sterol structure are important for the production of type I binding spectra. Interestingly, the four sterols, which produced type I spectra (lanosterol, 14-methyl-zymosterol, 31-nor-lanosterol and argosterol) all contain a double bond between carbons 24 and 25 (red circle), which is missing from the structures of those sterols (cholesterol, Δ^7-dihydrolanosterol and 32-cyclopropylidenemethyl-dihydrolanosterol) that do not bind to CYP51 (pink circle) suggesting this could be the part of the sterol structure that interacts with the substrate binding site. However, the two sterols, which produce weak type I binding spectra, dihydrolanosterol and obtusifoliol, do not contain this double bond suggesting other interactions may be taking place. Dihydrolanosterol and Δ^7-dihydrolanosterol are similar in structure. Both contain a double bond between carbons 24 and 25 (pink circle) and a methyl group at carbon 14 (blue circle). Yet dihydrolanosterol binds to CYP51 with weak type I spectra and Δ^7-dihydrolanosterol does not produce any discernable spectra. The difference between these two structures is the position of the double bond in the second ring (also known as ring B). In dihydrolanosterol this double bond is situated between carbons 8 and 9, but in Δ^7-dihydrolanosterol it is adjacent to the methyl group between carbons 7 and 8. This suggests the sole hydrogen atom projecting from carbon 7 in Δ^7-dihydrolanosterol is at an orientation where it can interfere with the coordination of the methyl group to the haem, thus preventing the production of type I binding spectra. Obtusifoliol differs from the other sterols used in this study as the double bond is at carbon 24, but projects to another carbon (carbon 28), which is not part of the other sterol structures (dashed pink circle). Obtusifoliol has a similar structure to eburicol, a substrate of CYP51 (not used in this study) (figure 6.17) as both structures contain a double bond between carbons 24 and 28 (red circle). Although, a methyl group at carbon 4 is omitted from the obtusifoliol structure, thus making this sterol a C4-mono-methylated sterol (opposed to eburicol, which is a C4-double-methylated sterol). This suggests double bonds are required to be present in the alkyl side chain of the sterol to ensure the substrate is at the correct orientation to

coordinate the methyl group at carbon 14 with the haem, thus producing type I binding spectra.

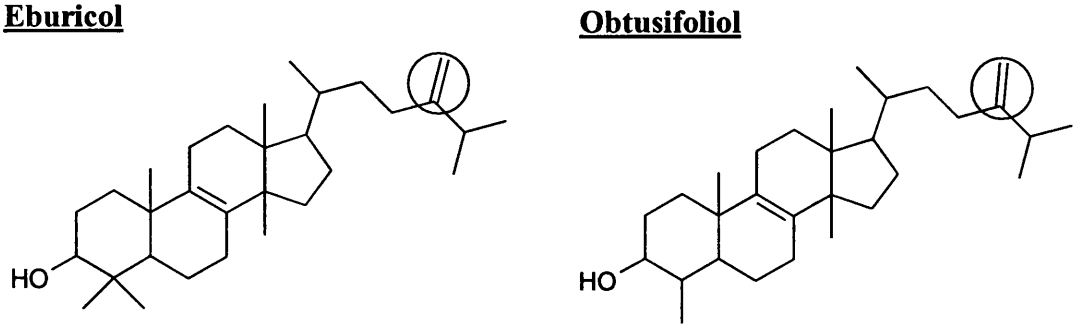


Figure 6.17 – Comparison of eburicol and obtusifoliol structures. The red circle indicates the similar alkyl chain structures in both molecules.

Interestingly, binding studies with CYP51 from *Leishmania infantum* (a human pathogen, which grows in the intestine of the sand fly and is transmitted by its bite) showed that this enzyme favours obtusifoliol and norlanosterol over eburicol and lanosterol, which is in contrast to CYP51 from *C. albicans* (Hargrove *et al.*, 2011). CYP51 from *L. infantum* appears to bind to sterols with different affinities depending on the number of methylated side-chains at carbon 4. This enzyme binds more tightly to sterols with one methyl group at this carbon (C4-mono-methylated, i.e. obtusifoliol and norlanosterol) and weaker to C4-double-methylated sterols (i.e. lanosterol, dihydrolanosterol and eburicol) (Hargrove *et al.*, 2011). It is also able to bind to sterols that do not contain a methyl group at this position (C4-des-methylated, i.e. 14-methyl-zymosterol) (Hargrove *et al.*, 2011). (Type I spectra was produced in all cases.) This is in contrast to *C. albicans* CYP51, which seems to discriminate between sterol structures depending on the composition of the alkyl side-chain and the presence of a double bond at carbon 7 rather than the number of methylated side-chains at carbon 4.

Unlike CYP51 from *C. albicans*, Alk1L456S was shown to bind to all the sterols used for binding in this study, producing either type I or reverse type I binding spectra (see figure 6.18).

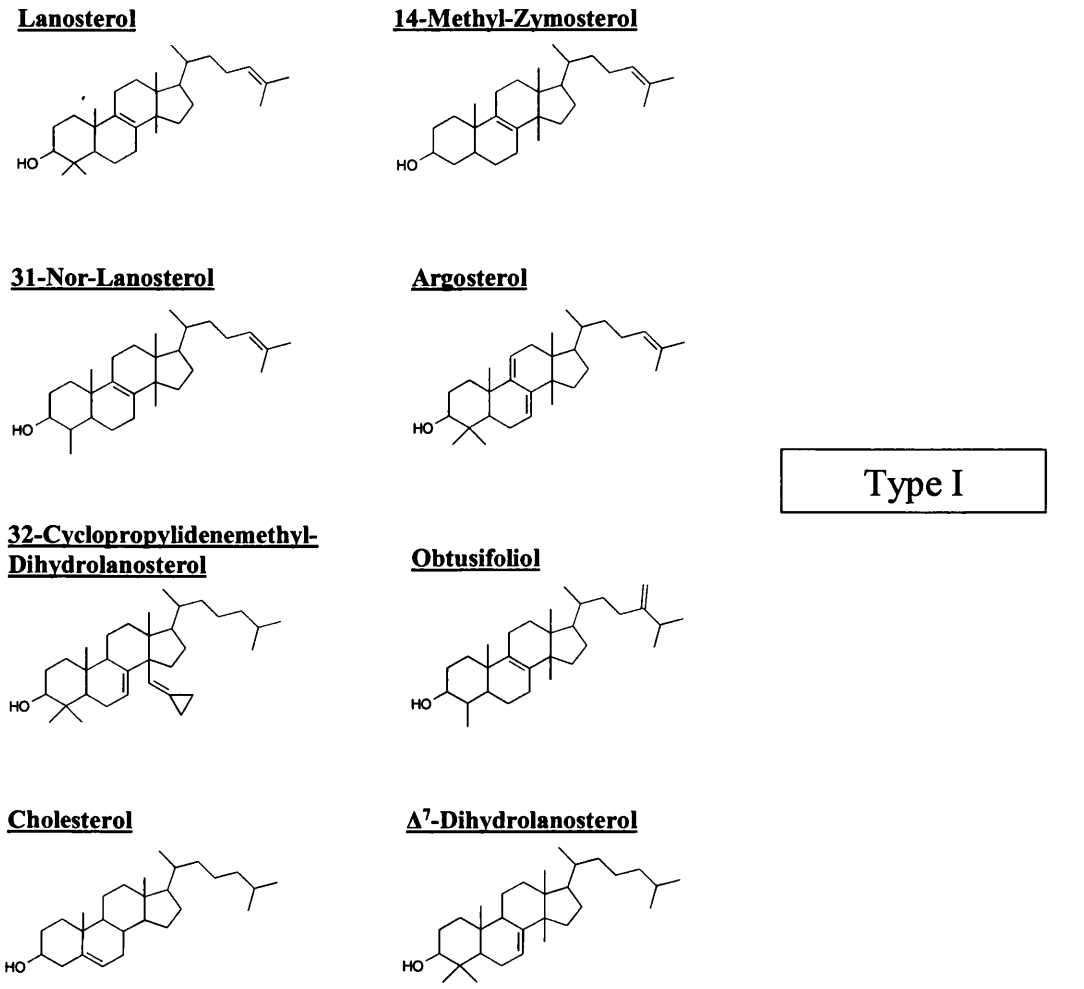


Figure 6.18 – Binding spectra produced titrating sterols against Alk1L456S. Sterols are separated into groups according to the binding spectra produced.

Unlike CYP51, Alk1L456S was able to bind sterols that do not contain a methyl group at carbon 14 in the structure, like cholesterol and 32-cyclopropylidenemethyl-dihydrolanosterol (which contains a cycloalkyl group at this position). This indicated this side chain is not a requirement for type I binding with Alk1L456S. The sterols that show type I binding (figure 6.18) with Alk1L456S have different alkyl side chains. Lanosterol, 14-methyl-zymosterol, 31-nor-lanosterol and

argosterol contain a double bond between carbons 24 and 25. 32-cyclopropylidenemethyl-dihydrolanosterol, cholesterol and Δ^7-dihydrolanosterol have no double bonds in their alkyl side chain. Obtusifoliol has a double bond between carbon 24 and 27 (and is C4-mono-methylated). This suggests that unlike CYP51 the double bond composition of the alkyl side chain and the presence of a double bond at carbon 7 are not integral to type I binding with Alk1L456S.

All the sterols used in this study have methyl groups at carbon 18 and 19, which may play a role in binding to Alk1L456S. However, how the sterols bind to Alk1L456S is difficult to ascertain as the majority of the sterols used in this study bind producing a type I spectra. Dihydrolanosterol is the only sterol not to produce "classical" type I spectra, but rather reverse type I. These spectra can either be produced by the hydroxyl group (OH-) at carbon 3 (on ring A) interacting with the haem or the substrate interacting with the enzyme at a different site to the other substrates used in this study (Schenkman *et al.*, 1972). Interestingly, while dihydrolanosterol produces reverse type I binding spectra, Δ^7-dihydrolanosterol (which has a similar structure) produces type I. However, until these substrates have been shown to be metabolised in reactions catalysed by Alk1L456S it will be difficult to ascertain how these sterols bind to Alk1L456S.

Lanosterol was used in reconstitution assays with Alk1L456S and CYP51 as this sterol was shown to have similar binding affinities for both of these enzymes. As expected, CYP51 was shown to demethylate lanosterol. However, no activity was seen with Alk1L456S, suggesting this enzyme is not involved in lanosterol demethylation, although, an active reconstitution assay for this protein remains a future aim. 14-methyl-zymosterol was also used for reconstitution assays as it also shares similar binding affinities for Alk1L456S and CYP51. Interestingly, no product was detected by GCMS with either enzyme despite 14-methyl-zymosterol sharing structural similarities with lanosterol. This means either CYP51 and Alk1L456S proteins can bind to this sterol, but not metabolise it or these enzymes did catalyse the metabolism of this sterol, but the turnover was so low it could not be detected on the GCMS. Cholesterol and 32-cyclopropylidenemethyl-dihydrolanosterol were also used in reconstitution assays (despite the inability of these sterols to bind to CYP51) because cholesterol has been shown to be

metabolised by other cytochromes P450 from other organisms and 32-cyclopropylidenemethyl-dihydrolanosterol has a unique sterol structure. As expected no product was identified in reactions catalysed by CYP51. Alk1L456S also failed to produce any products in reconstitution assays with these sterols.

The reconstitution assays undertaken in this chapter seem to suggest that Alk1L456S is able to bind to lanosterol, 14-methyl-zymosterol, 32-cyclopropylidenemethyl-dihydrolanosterol and cholesterol, but not metabolise them. Reconstitution assays using Alk1L456S to catalyse the reaction may require an increased concentration of CaCPR to produce detectable product. In this study 3.8μM of CaCPR was used as higher concentrations of protein could not be obtained in the volume of the reconstitution system used (i.e. 1ml). To rectify this problem either the volume of the reconstitution system used needs to be increased so that greater volumes of CaCPR can be used (thereby increasing the concentration), the concentration of Alk1L456S protein needs to be reduced to increase the ratio of CaCPR to Alk1L456S or a greater yield of CaCPR needs to be produced during protein expression. However, 3.8μM of CaCPR was sufficient to produce 20% demethylated product in reconstitution assays using CYP51 to catalyse the reaction and lanosterol as the substrate, compared to 8% when 10μM of ScΔ33:CPR was used. This difference in the product yield could be due to CaCPR being the homologous CPR of the CYP51 used in this study or it could be due to CaCPR being a full-length reductase whereas ScΔ33:CPR is a truncated version of CPR from *S. cerevisiae.* Full-length CPR contains a membrane anchor region in its N-terminus, which allows the protein to bind/anchor itself to the membrane allowing it to be involved in electron transfer to the haem iron of the (membrane-bound) cytochrome P450 (Paine *et al.*, 2005). A truncated CPR protein, such as ScΔ33:CPR, may not be able to fully interact with the cytochrome P450 as it is not anchored to the membrane and this binding is thought to be essential for the correct spatial orientation of this redox partner (Paine *et al.*, 2005). However, ScΔ33:CPR has been shown to achieve similar results as full-length *S. cerevisiae* CPR in kinetic studies (Lamb *et al.*, 2001b), suggesting the lack of N-terminus may not be responsible for the decreased 14-demethylated product formation seen in this study.

Alternatively, Alk1L456S may catalyse the metabolism of these sterols, but the turnover may be too low to be detected on the GCMS. As with the fatty acid reconstitution assays undertaken in chapter 5, other chromatographic methods, such as TLC and HPLC, may be required in conjunction with radioactive substrate to detect products of Alk1L456S sterol metabolism. These detection methods may also be required as Alk1L456S may be involved in the hydroxylation of these sterols (rather than the demethylation of lanosterol and possibly 14-methyl-zymosterol) and these products may not be easily detected using the current method of detection.

Other cytochromes P450 from other organisms have been shown to be involved in the hydroxylation of cholesterol. Table 6.4 and figure 6.19 show the cytochromes P450, which catalyse this reaction and the products produced.

Enzyme	Species	Product	Ks (μM)	Reference
CYP46A1	Brain	24-Hydroxycholesterol	8.58±1.51	Mast *et al.*, 2004
CYP125	*M. tuberculosis*	26-Hydroxycholesterol	0.20±0.02	Capyk *et al.*, 2009
CYP125A1	*M. tuberculosis*	26-Hydroxycholesterol	0.107±0.064	Ouellet *et al.*, 2010
CYP125	*Rhodococcus joshii* RHA1	26-Hydroxycholesterol	0.20±0.08	Rosłoniec *et al.*, 2009
CYP7A1	Liver	7α-Hydroxycholesterol	4.3±0.01	Mast *et al.*, 2005
Alk1L456S	*C. albicans*	?	4.7±1.55	This study

Table 6.4 – Comparison of cholesterol binding and metabolising cytochromes P450. All the cytochromes P450 shown here (except Alk1L456S) are involved in the hydroxylation of cholesterol. Alk1L456S and CYP7A1 share a similar binding affinity for this substrate, but it binds tighter to CYP125 from *M. tuberculosis* and *R. joshii* RHA1. CYP46A1 binds to cholesterol the weakest of all the enzymes shown in this table.

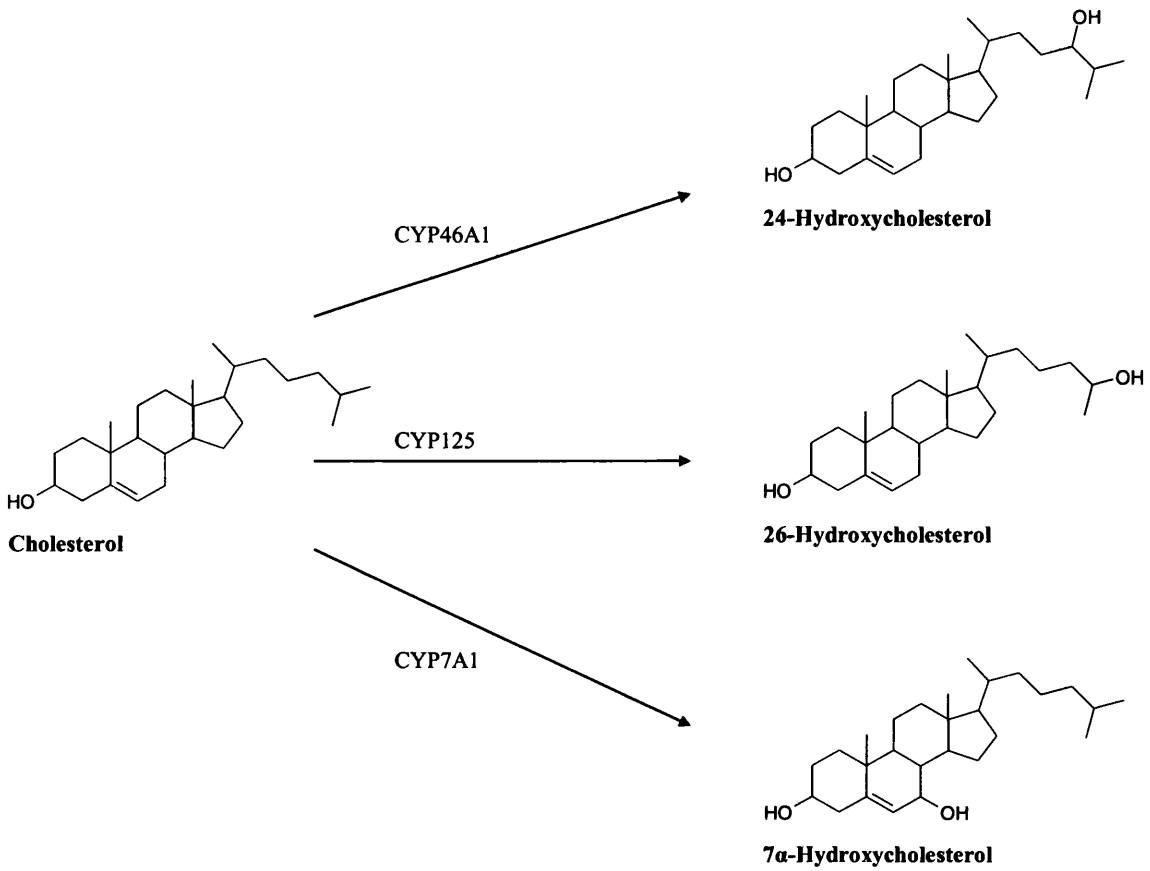


Figure 6.19 – Products of cholesterol hydroxylation catalysed by cytochrome P450 enzymes. CYP46A1 catalyses the hydroxylation of cholesterol at carbon 24. CYP125 is involved in the hydroxylation of this molecule at carbon 26. CYP7A1 catalyses the hydroxylation of the cholesterol at carbon 7 (ring B).

CYP46A1 from brain extracts (Mast *et al.*, 2004), CYP7A1 from liver extracts (Mast *et al.*, 2005) and CYP125 from *M. tuberculosis* and *Rhodococcus joshii* RHA1 (Capyk *et al.*, 2009; Rosłoniec *et al.*, 2009; Ouellet *et al.*, 2010) have all been shown to hydroxylate cholesterol at different carbons within the structure (table 6.4 and figure 6.19). CYP46A1 has been implicated in Alzheimer's disease. It is believed to reduce the concentration of cholesterol in the brain of healthy individuals by converting it to its hydroxy alternative (figure 6.19), which can be eliminated in the cerebral spinal fluid (Garcia *et al.*, 2009). However, when this enzyme is impaired or inhibited, cholesterol is thought to be accumulated in the brain leading to the formation of β-amyloid plaques, which in turn are involved in the pathology of Alzheimer's disease (Garcia *et al.*, 2009). CYP7A1 is also involved in the breakdown of cholesterol. It catalyses the first step of bile salt production whereby

cholesterol is converted to 7α-hydroxylcholesterol (figure 6.19) (Mast *et al.*, 2005). These bile salts can then be used to eliminate cholesterol from the body preventing the accumulation of cholesterol in the arteries (Mast *et al.*, 2005). Unlike CYP46A1 and CYP7A1, which are involved in the elimination of cholesterol, CYP125 (in some bacteria) is involved in the metabolism of this sterol for use in growth (Capyk *et al.*, 2009). CYP125 catalyses the conversion of cholesterol to 26-hydroxycholesterol (figure 6.19), which can be broken down further by other enzymes to form pyruvate. This molecule can then be used by the organism for growth (Capyk *et al.*, 2009). The role of CYP125 in cholesterol metabolism may suggest a potential role for Alk1L456S in the growth of *C. albicans* on cholesterol and, in turn, the pathogenicity of the organism in humans if Alk1L456S can be shown to hydroxylate cholesterol in metabolism studies.

In conclusion, Alk1L456S has been shown to bind tightly with azole antifungal drugs producing type II spectra in binding studies, thus indicating a potential role for the enzyme in the resistance of *C. albicans* to azole drugs. Alk1L456S has also been shown to produce type I binding spectra with sterols in this study. This is the first time a fungal cytochrome P450 enzyme other than CYP51 or CYP61 has been shown to interact with these substrates. Although, reconstitution assays with Alk1L456S and the sterols: lanosterol, 14-methyl-zymosterol, 32-cyclopropylidenemethyl-dihydrolanosterol and cholesterol failed to show product formation in this study, Alk1L456S may play a role in the conversion of cholesterol for the use in growth in *C. albicans*.

Chapter 7: Discussion

Cytochromes P450 are a superfamily of haem-containing proteins, which are involved in the metabolism of a wide variety of substrates including xenobiotics and sterols. Many cytochromes P450 have been shown to hydroxylate alkanes and/or fatty acids and the products formed (particularly diacids) have been shown to be important platform chemicals in the production of many industrial products, including hot-melting adhesives, plastics, lubricants, fragrances and antibiotics (Eschenfeldt *et al.*, 2003; Liu *et al.*, 2003).

CYP52s are a family of cytochrome P450 proteins, which are only found in yeast and fungi and are involved in the hydroxylation of alkanes and fatty acids. To date, 52 CYP52s have been identified in a number of yeast and fungal species, including *Candida tropicalis*, *C. maltosa*, *Yarrowia lipolytica* and *Aspergillus nidulans* (http://drnelson.uthsc.edu/CytochromeP450.html). However, only 8 of these proteins have been shown to be catalytically active at the protein level, including CYP52A3 and CYP52A4 from *C. maltosa* and CYP52A13 and CYP52A17 from *C. tropicalis*, which have been shown to be involved in the production of diacids from alkanes and fatty acids (Zimmer *et al.*, 1995; Zimmer *et al.*, 1996; Scheller *et al.*, 1998; Eschenfeldt *et al.*, 2003). The ability of these enzymes to produce such fatty acid derivatives indicates a potential role for CYP52 proteins in biotechnology. These fatty acid derivatives could be used as platform chemicals to form the basis of a number of products including plastics, lubricants, fragrances and antibiotics (Eschenfeldt *et al.*, 2003; Liu *et al.*, 2003).

To understand fully the role of this family of cytochromes P450 in the physiology of yeast and fungi and their potential use in biotechnology, the enzymatic activity of each protein needs to be determined. Therefore, in this study the potential CYP52 from *C. albicans* with the greatest homology to the known alkane-assimilating CYP52 from *C. maltosa*, CYP52A3, was expressed at the protein level and studied. Using a bioinformatic approach, 10 putative cytochrome P450 genes were identified in *C. albicans*. These were *ALK1*, *ALK2*, *ALK3*, *ALK6*, *ALK8*, *TRI4*, *DIT2*, *PAH1*, *ERG5* and *ERG11*. Of these genes, five were identified as putative alkane-assimilating CYP52s: *ALK1*, *ALK2*, *ALK3*, *ALK6* and *ALK8*. *ALK1* was shown to

have the greatest homology (79% identity) to *CYP52A3* and was used in subsequent studies.

ALK1 was cloned into *E. coli* for the expression of recombinant protein. However, protein expression experiments resulted in the formation of protein with a reduced carbon monoxide peak of 420nm, rather than a peak at 450nm, which is indicative of cytochrome P450 (Omura & Sato, 1964). The inability to express active protein was due to an alteration in the universal genetic code, which occurs in *C. albicans*. In this species, the triplet CUG encodes for the amino acid serine rather than the expected leucine (Santos & Tuite, 1995). As serine is a polar, hydrophilic amino acid and is non-polar and hydrophobic, such a difference in the properties of the two amino acids can have a detrimental effect on the folding of protein leading to inactive protein, as in the case of Alk1. In the *ALK1* nucleotide sequence, the CUG codon was altered to TCT, which encodes for serine, resulting in the expression of protein producing a reduced carbon peak at 447nm, which is indicative of active and correctly folded cytochrome P450. The resultant Alk1L456S protein was purified and used for binding studies and reconstitution assays.

As Alk1L456S is a putative member of the CYP52 family of cytochromes P450, alkanes were initially used in substrate binding studies. Hexadecane (C16:0 alkane) was used as the control CYP52 in this study, CYP52A3 had previously been shown to bind to this alkane and catalyse its hydroxylation (Scheller *et al.*, 1996; Scheller *et al.*, 1998). Binding studies with Alk1L456S and this alkane, as well as dodecane (C12:0 alkane), failed to produce any discernible binding spectra. However, binding studies with the fatty acids lauric acid (C12:0), myrisitic acid (C14:0) and palmitic acid (C16:0) resulted in type I binding spectra suggesting these hydrocarbons are substrates of Alk1L456S. Alk8, another CYP52 enzyme from *C. albicans*, was shown to be able to bind to (and metabolise) fatty acids, but not alkanes (Kim *et al.*, 2007), indicating some CYP52 proteins from *C. albicans* are unable to interact with these hydrocarbons. Interestingly, Alk8 has a greater binding affinity for lauric acid than Alk1L456S has for the same fatty acid and preferentially catalyses the hydroxylation of lauric acid over myrisitic acid and palmitic acid in reconstitution assays (Kim *et al.*, 2007), indicating that CYP52 proteins from *C. albicans* are

involved in the preferential hydroxylation of different hydrocarbons, but are also able to hydroxylate/bind to multiple fatty acid substrate.

The ability for a cytochrome P450 enzyme to be involved in the hydroxylation of multiple fatty acids is not exclusive to CYP52s as other cytochromes P450 have been shown to have similar enzymatic activities. CYP105D5 from *Streptomyces coelicolor* A3(2) is able to hydroxylate both lauric acid and oleic acid (C18:1) (Chun *et al.*, 2007), wherease CYP124 from *Mycobacterium tuberculosis* is able to terminally hydroxylate a number of methyl-branched fatty acids, including 15-methyl palmitic acid, phytanic acid and isoprenoids (Johnston *et al.*, 2009). The plant cytochrome P450, CYP77A4 is not only able to hydroxylate lauric acid and palmitic acid, but is also involved in the epoxidation of oleic acid (Sauveplane *et al.*, 2009). Other cytochromes P450 have also been shown to be involved in the metabolism of a broad range of substrates, most notably the liver cytochrome P450, CYP3A4. This enzyme has been shown to be involved in the metabolism of a number of therapeutic compounds, including antibiotics (e.g. erythromycin (Watkins *et al.*, 1985)), cancer therapeutic agents (e.g. tamoxifen (Jacolot *et al.*, 1991)), azole antifungals (e.g. itraconazole (Locuson *et al.*, 2007) and cholesterol lowering drugs (e.g. lovastatin (Wang *et al.*, 1991)).

As palmitic acid was shown to have the greatest binding affinity for Alk1L456S, it was used in reconstitution assays with this protein. Reconstitution assays with this substrate and protein failed to show the production of metabolites when analysed on the GCMS. This was either due to the inefficacy of the method used to detect any product formed at low levels or the inability of Alk1L456S to catalyse the reaction.

To determine substrate specificity *in vivo* experiments using *S. cerevisiae* AH22 cells transformed with the YEp51:Alk1L456S_yeast plasmid grown on hexadecane showed that Alk1L456S was able to catalyse the hydroxylation of this alkane, despite hexadecane producing no discernable binding spectra with Alk1L456S. Hexadecane is the substrate of choice for CYP52A3 and *S. cerevisiae* is unable to use this hydrocarbon for growth, although, it has been shown to grow on a range of fatty acids (Rehm & Reiff, 1981). Therefore, growth could be attributed to Alk1L456S expression and function. This was unexpected as binding studies with alkanes (as

previously mentioned) failed to show any discernible binding suggesting these hydrocarbons were not substrates of Alk1L456S. The lack of binding spectra in this study may have been due to either the conditions of the methods used not favouring alkane binding to Alk1L456S, or the binding of the substrate to the protein was too weak to produce any discernible binding spectra. Kim and co-workers (2007) showed that Alk8 was unable to bind and metabolise alkanes in reconstitution assays, suggesting other proteins/enzymes maybe required for the stability and/or activity of CYP52s to occur. However, other CYP52s, including CYP52A13 and CYP52A17 from *C. tropicalis*, have been implicated in the hydroxylation of fatty acids, but not alkanes (Craft *et al.*, 2003; Eschenfeldt *et al.*, 2003) despite *C. tropicalis* being able to use alkanes for growth (Picataggio *et al.*, 1992). This suggests more *in vivo* experiments need to be undertaken with CYP52s to investigate whether these enzymes (as with Alk1L456S) are also involved in the hydroxylation of alkanes, despite no evidence *in vitro*. Alternatively, the ability for some CYP52s (e.g. Alk8, CYP52A13 and CYP52A17) to be involved in the hydroxylation of fatty acids rather than alkanes may be required for diacid formation. These enzymes may work in conjunction with alkane-hydroxylating CYP52s and other enzymes to produce diacids, which can be used by the cell for growth. Diacids can also be formed in environments rich in fatty acids to prevent their accumulation as high fatty acid concentrations can be harmful to the cell (Vamecq *et al.*, 1985; Tehlivets *et al.*, 2007). High concentrations of the fatty acids, lauric acid and capric acid (C10:0) have been shown to result in the inhibition of yeast growth and are thought to have potential *in vitro* uses in long-term antibiotic therapy (Kabara *et al.*, 1972; Bergsson *et al.*, 2001).

In yeast genera, such as *Candida*, which can utilise alkanes and fatty acids, CYP52 enzymes are used to catalyse the initial, rate-limiting step of the α-oxidation pathway in the endoplasmic reticulum; the hydroxylation of the alkane to its respective alcohol (figure 7.1) (Scheller *et al.*, 1998; Fickers *et al.*, 2005). As described in chapter 1, following this step the alcohol is further converted to its respective aldehyde and fatty acid (figure 7.1). These conversions are either catalysed by CYP52 alone or by the enzymes alcohol dehydrogenase and aldehyde dehydrogenase (figure 7.1) (Scheller *et al.*, 1998; Eschenfeldt *et al.*, 2003; Fickers *et al.*, 2005). The fatty acid produced can either be transported from the endoplasmic reticulum to the

peroxisome to be subjected to β-oxidation or alternatively it can be converted further in the endoplasmic reticulum, by the ω-oxidation pathway, to produce diacid (figure 7.1) (Porter & Coon, 1991; Jiao *et al.*, 2000; Eschenfeldt *et al.*, 2003). If this occurs then the fatty acid is initially converted, by CYP52, to its respective fatty alcohol (hydroxyl fatty acid), which is further converted to the fatty aldehyde and diacid (figure 7.1). As with the α-oxidation pathway, these latter conversions are either catalysed by only CYP52 enzymes or by alcohol dehydrogenase and aldehyde dehydrogenase (figure 7.1) (Scheller *et al.*, 1998; Eschenfeldt *et al.*, 2003; Fickers *et al.*, 2005). The diacid produced, like the fatty acid, can be transported to the peroxisome, where it is subjected to β-oxidation and used for growth (figure 7.1) (Porter & Coon, 1991; Jiao *et al.*, 2000; Eschenfeldt *et al.*, 2003). As described, CYP52 enzymes can generate diacids alone (which has been shown to occur in *in vitro* assays with hexadecane and CYP52A3 (Scheller *et al.*, 1996)), which is an advantage in the biotechnological applications and production of this hydrocarbon. However, *in vivo*, this would cause a reduction in NADPH, which is required for CYP52 activity, reducing the cells ability to hydroxylate alkanes/produce fatty acids and diacids. Therefore, CYP52 enzymes may act in conjunction with alcohol dehydrogenase and aldehyde dehydrogenase to increase the efficiency of the conversion/hydroxylation of the alkane substrate (Scheller *et al.*, 1998; Eschenfeldt *et al.*, 2003). However, if other enzymes are used by the cell to convert alcohols to fatty acids/diacids then the cell may also use separate CYP52s to hydroxylate alkanes and fatty acids to ensure maximal alkane hydroxylation for growth. This may explain the presence of multiple enzymes belonging to the CYP52 family of enzymes in the cell and the ability of these enzymes to catalyse the hydroxylation of multiple hydrocarbons (despite having a preferred substrate).

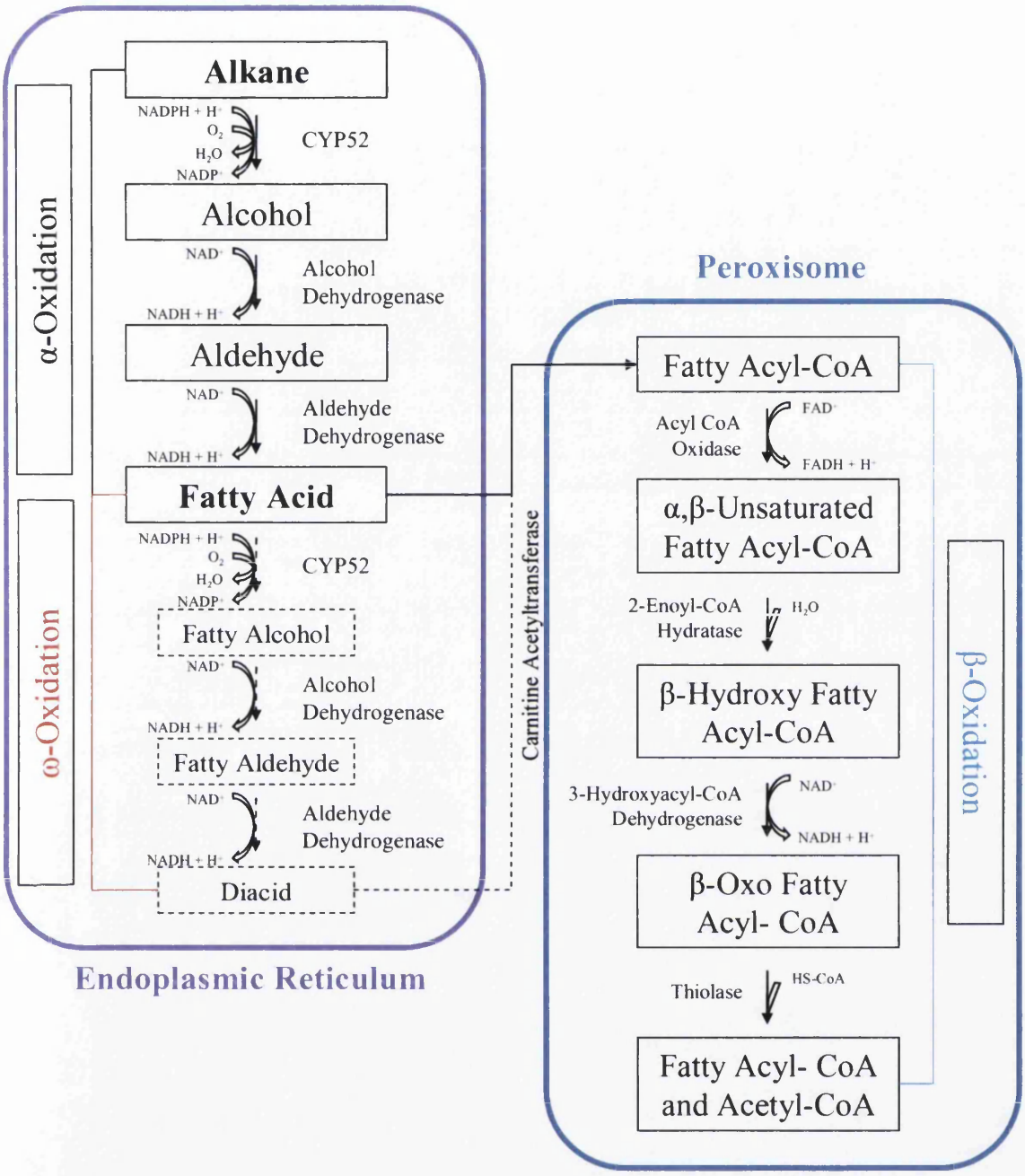


Figure 7.1 – Overview of the α-, ω- and β-oxidation pathways identifying the enzymes involved and the products produced. The α- and ω-oxidation pathways both take place in the endoplasmic reticulum, whereas the β-oxidation pathway takes place in the peroxisome. The products of the major pathways in alkane metabolism are in solid boxes. The products of the minor diacid-producing pathway are in dashed boxes. * Alcohol dehydrogenase is substituted for alcohol oxidase in some papers.

The *in vivo* experiments undertaken with *S. cerevisiae* AH22 cells transformed with the plasmid YEp51:Alk1L456S_yeast is the first time Alk1L456S has been shown to be involved in the hydroxylation of alkanes. However, the activity of this enzyme was inferred from the ability of the transformed cells to grow on alkanes (unlike the negative control) and the similarities of the lipid profile produced between these cells and the positive control (AH22 cells containing YEp51:CYP52A3_yeast). Peaks representing hexadecanol, 16-hydroxypalmitic acid and thapsic acid were not identified on the GC chromatograph, although this does not mean Alk1L456S was unable to produce these metabolites as the positive control also failed to show peaks representing these hydrocarbons despite CYP52A3 having been shown previously to catalyse the conversion of hexadecane to thapsic acid (Scheller *et al.*, 1998). This experiment shows that other enzymes such as those that make up the β-oxidation pathway (figure 7.1) were acting on the products of Alk1L456S (and CYP52A3) as the metabolites were required for cell growth.

The β-oxidation pathway (figure 7.1) is ubiquitous in nature and is required for the break down of fatty acids (and diacids) for use in growth (Wanders *et al.*, 2007). This pathway is not only used by yeast, but it is also used by a number of other organisms, including animals, which utilise hydrocarbons for growth. Interestingly, in these organisms β-oxidation is carried out in both the mitochondrion and peroxisome (to a lesser extent), but in yeast β-oxidation only occurs within the peroxisome as it does not have enzymes for β-oxidation located within the mitochrondria (Kunau *et al.*, 1988; Hettema & Tabak, 2000; Kragt *et al.*, 2006). Picataggio and co-workers (1992) showed that it is possible to produce diacids from dodecane and myrisitic acid in an industrial strain of *C. tropicalis* by disrupting the genes that encode for acyl CoA oxidase (POX4 and POX5), the first enzyme in the β-oxidation pathway (figure 7.1). The media used contained glucose as co-substrate for growth, which aided diacid production. Diacid yield was further increased by amplifying the diacid-producing CYP52 enzymes and CPR in the POX knockout strain (Picataggio *et al.*, 1992). This method could be used to aid diacid production in *S. cerevisiae* cells containing YEp51:Alk1L456S_yeast. As with the above method an additional growth substrate such as glucose, would be required to maintain cell growth while diacid is being formed. To aid with this experiment it may be beneficial to replace the inducible YEp51 vector (Rose & Broach, 1991) with

a constitutive expression vector, such as pMA91 (Mellor *et al.*, 1983). This vector contains a strong PGK promoter, which does not require the addition of galactose for the transcription of the gene (as required by YEp51) (Mellor *et al.*, 1983). However, using this type of vector may cause toxicity so an inducible vector or a weaker promoter may be required for this experiment to be viable.

However, other studies investigating the disruption of the POX genes in the β-oxidation pathway of other *C. tropicalis* strains failed to produce similar results to Picataggio and co-workers (1992), suggesting the ability to produce diacids may be strain specific as other genes could have also been disrupted in the industrial strain used (Hara *et al.*, 2001). Therefore, another possible way of producing diacids from alkanes other than POX gene knockout is the impairment of carnitine acetyltransferase (Cao *et al.*, 2006).

Prior to the breakdown of fatty acids (and diacids) by β-oxidation, they require activation by coenzyme A (CoA) (Carman *et al.*, 2008; van Roermund *et al.*, 2008; van Roermund *et al.*, 2011). CoA, a thiol (organosulphur compound), interacts with the hydroxyl (-OH) group of the fatty acid to produce the thiolester, fatty acyl-CoA (figure 7.2). This process is catalysed by acyl-CoA synthetase and the activation of the fatty acids allows these normally hydrophobic hydrocarbons to move from the cytoplasm to the organelles and/or from one organelle to another (Kanayama *et al.*, 1998). However, when diacids are present instead of fatty acids, they are activated by dicarboxylyl-CoA synthetase rather than acyl-CoA synthetase (Vamecq *et al.*, 1985).

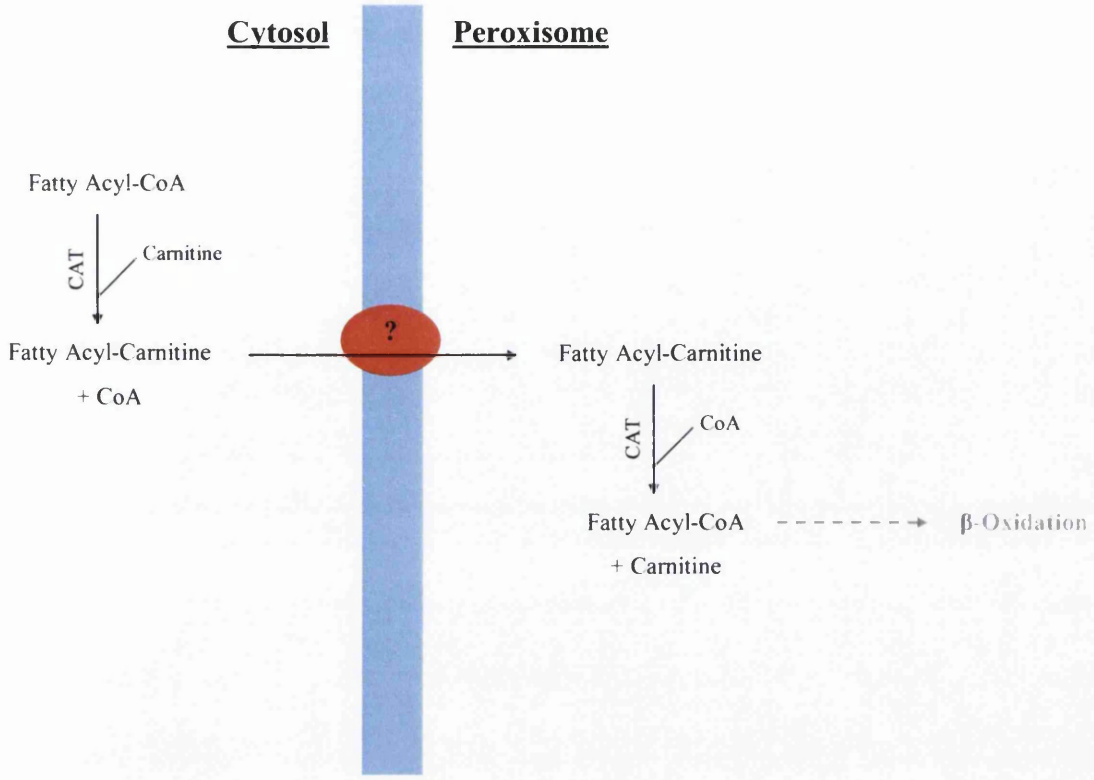


Figure 7.2 – Transport of fatty acyl-CoA into the peroxisome. The fatty acyl-CoA molecule is converted to fatty acyl-carnitine in the presence of carnitine acetyltranseferase (CAT) and carnitine. This molecule is transported across the peroxisomal membrane, by an unknown mechanism (possibly involving an ABC transporter (Hettema *et al.*, 1996; van der Klei & Veenhuis, 1997)), into the peroxisome where it is then converted back into fatty acyl-CoA in the presence of CAT and coenzyme A (CoA). This molecule can then enter the β-oxidation pathway.

The fatty acyl-CoA works in tandem with carnitine and carnitine transferase to allow the fatty acid to cross the peroxisomal membrane into the organelle. The fatty acyl-CoA binds to free carnitine producing fatty-acyl carnitine and free CoA at the cytosolic side of the peroxisomal membrane, in a reaction catalysed by carnitine acetyltransferase (CAT) (figure 7.2) (Bieber, 1988; van Roermund *et al.*, 1995; Prigneau *et al.*, 2003; Cao *et al.*, 2006; Carman *et al.*, 2008; Strijbis *et al.*, 2008; Zhou & Lorenz, 2008).

The fatty acyl-carnitine is transported across the membrane into the peroxisomal matrix. Here, the fatty acyl-carnitine is converted back into fatty acyl-CoA for use in β-oxidation and this reaction is catalysed by another CAT enzyme located within the

peroxisome (figure 7.2) (Bieber, 1988; van Roermund *et al.*, 1995; Prigneau *et al.*, 2003; Cao *et al.*, 2006; Carman *et al.*, 2008; Strijbis *et al.*, 2008; Zhou & Lorenz, 2008). The dissociated carnitine then moves out of the organelle into the cytosol, possibly across a diffusion gradient produced in conjunction with the free CoA in the cytosol, which moves into the peroxisome.

Therefore, disrupting the CAT gene would prevent the transport of fatty acids into the peroxisome, thus favouring diacid production as fatty acid accumulation would occur (Cao *et al.*, 2006) thus disrupting this gene in *S. cerevisiae* cells containing YEp51:Alk1L456S_yeast could aid diacid production. This would allow the strain to be used as a biotechnological tool in the formation of platform chemicals.

Although, diacids are the preferential products of CYP52 hydroxylation for their use as platform chemicals in the production of industrial products, whole-cell experiments have also been undertaken to produce large yields of fatty alcohols. Fatty alcohols can be used as the chemical basis for the production of bioplastics and are a product of CYP52 hydroxylation of fatty acids in yeast and fungi (Lu *et al.*, 2010). Lu and co-workers (2010) constructed a *C. tropicalis* knockout mutant, which was able to produce fatty alcohols using a strain where the β-oxidation pathway was known to be blocked. They knocked out the genes encoding CYP52A13 and CYP52A17 (and their allelic pairs CYP52A14 and CYP52A18 respectively) to prevent the formation of diacids, which are known to be formed in reactions catalysed by these enzymes (Lu *et al.*, 2010). This meant only the CYP52s involved in the production of fatty alcohols from fatty acids remained active. These disruptions were insufficient to prevent the further metabolism of the fatty alcohols produced to diacids as other enzymes are involved in this pathway. The additional disruption of alcohol oxidase failed to prevent diacid formation (in this experiment) (Lu *et al.*, 2010). However, alcohol dehydrogenase is involved in the conversion of fatty alcohols in the endoplasmic reticulum not alcohol oxidase (which is found in the peroxisome) (figure 7.1) (Fickers *et al.*, 2005). Therefore, disruption of the genes encoding alcohol dehydrogenase rather than alcohol oxidase or the disruption of both genes would have been required to prevent diacid production. However, the formation of this product was prevented when the gene for aldehyde dehydrogenase was disrupted (Lu *et al.*, 2010), thus resulting in the production of 12-hydroxylauric

acid (C12:0 fatty alcohol), 14-hydroxymyrisitic acid (C14:0 fatty alcohol) and 16-hydroxypalmitic acid (C16:0 fatty alcohol), when the *C. tropicalis* mutant was grown on lauric acid, myristic acid and palmitic acid respectively (Lu *et al.*, 2010). This shows that whole pathways need to be considered in the production of diacids and other fatty acid derivatives when using whole-cell techniques in biotechnology as enzymes other than CYP52 are involved in the metabolism of fatty acids.

Other techniques have also been employed to increase metabolite production using recombinant cytochrome P450. In this study, recombinant assays using Alk1L456S protein to catalyse the reaction failed to result in the identification of hydroxylated product of palmitic acid. However, further *in vitro* experiments using this protein (e.g. using labelled substrate or chromatographic techniques, such as TLC and HPLC, for analysis) might have been expected to result in the formation and detection of hydroxylated products.

Directed evolution and site-specific mutagenesis strategies have been employed to improve enzymatic activity and alter the substrate specificity of cytochrome P450 proteins. These are important techniques used in biotechnology for the metabolism of novel substrates and have been extensively used with $P450_{BM-3}$ (CYP102A1) from *Bacillus megaterium*. This is a self-sufficient cytochrome P450-CPR fusion protein, which is involved in the hydroxylation of fatty acids (Miura & Fulco, 1975; Boddupalli *et al.*, 1992; Schneider *et al.*, 1998; Gustafsson *et al.*, 2004; Warman *et al.*, 2005; Lamb *et al.*, 2010). Cytochromes P450 has been subjected to direct evolution (such as random mutagenesis) and site-specific mutagenesis techniques to produce a protein which is able to catalyse the metabolism of substrates other than fatty acids. It has been adapted to hydroxylate small alkanes, such as propane (C3:0 alkane), which it is unable to metabolise in its wild-type form (Fasan *et al.*, 2007; Bloom & Arnold, 2009). $P450_{BM-3}$ has also been altered to show enhanced enzymatic activity towards the polycyclic aromatic hydrocarbons (PAHs): phenanthrene, fluoranthene and pyrene (figure 7.3) (Carmichael & Wong, 2001). Other experiments involving the substitution of the amino acid phenylalanine with alanine at residue 87 in the wild-type $P450_{BM-3}$ protein has been shown to result in the formation of a mutant, which is able to catalyse fatty acids at the ω-carbon (Cirino & Arnold, 2002). The wild-type version of this protein is unable to

hydroxylate fatty acids at this carbon as it favours the ω1-, ω2- and ω3-carbons of these substrates (Miura & Fulco, 1975; Boddupalli *et al.*, 1992; Schneider *et al.*, 1998; Gustafsson *et al.*, 2004; Lamb *et al.*, 2010).

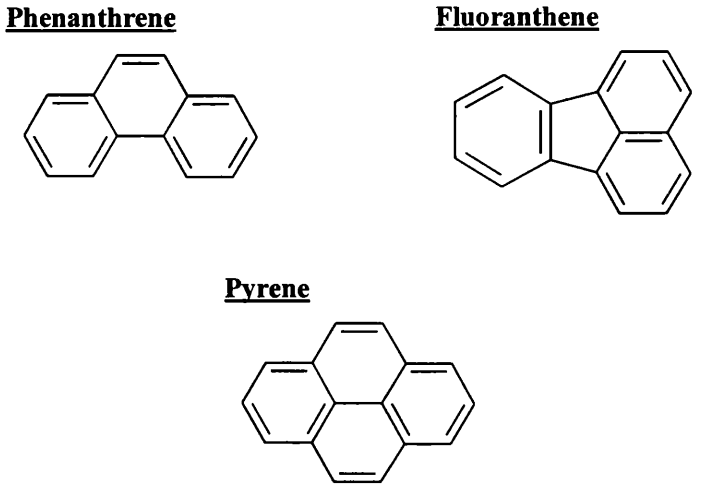
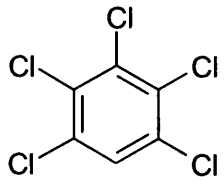


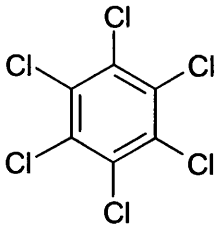
Figure 7.3 – Polycyclic aromatic hydrocarbons (PAHs): phenanthrene, fluroranthene and pyrene.

Other cytochromes P450 have also been subjected to directed evolution and site-specific mutagenesis, including $P450_{cam}$ (CYP101) from *Pseudomonas putida*, which is involved in the hydroxylation of camphor, a terpenoid commonly found in wood. Mutants of this cytochrome P450 have been produced, which show enhanced catalytic activity towards the heavily chlorinated benzenes: pentachlorobenzene and hexachlorobenzene to produce pentachlorophenol (figure 7.4) (Jones *et al.*, 2001). The wild-type version of this protein is unable to catalyse the hydroxylation of these substrates, but is involved in the metabolism of dichloro- and trichlorobenzene (figure 7.4) (Jones *et al.*, 2001).

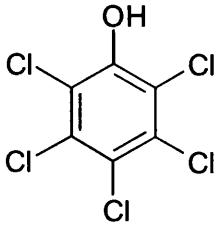
Pentachlorobenzene



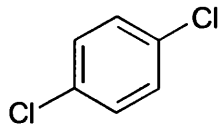
Hexachlorobenzene



Pentachlorophenol

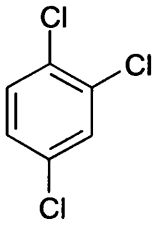


Dichlorobenzene



e.g. 1,4-Dichlorobenzene

Trichlorobenzene



e.g. 1,2,4-Trichlorobenzene

Figure 7.4 – Pentachlorobenzene, hexachlorobenzene, pentachlorophenol, dichlorobenzene and trichlorobenzene.

These techniques have not only been used to adapt the enzymatic function and specificity of cytochrome P450 protein, they have also been undertaken on other related proteins too, including NADPH-cytochrome P450 reductases. A mutated form of human CPR has been produced, which is dependent on NADH for electrons rather than NADPH (Döhr *et al.*, 2001). The amino acid tryptophan was substituted for alanine at residue 676 in the FAD binding domain of this protein allowing for this alteration in cofactor preference (Döhr *et al.*, 2001). The ability of a CPR to utilise NADH instead of NADPH is an advantage in biotechnology as NADPH is an expensive compound to purchase, but NADH is much cheaper thus reducing the overall cost of any *in vitro* assays performed using CPR as the redox partner (Urlacher & Eiben, 2006).

Alternatively, electrodes can be used to deliver electrons to cytochrome P450 proteins via CPR eradicating the need for NADPH (or NADH) in the reaction. Purified proteins are immobilised on an electrode to allow continuous electron transfer to the haem centre of the cytochrome P450 to drive catalytic activity (Krishnan *et al.*, 2011). In this technique it is advantageous to use cytochrome P450-CPR fusion proteins as electrons can be passed from the electrode to the CPR and then directly to the cytochrome P450 without needing to ensure that the two separate proteins can interact with each other and the electrode (Krishnan *et al.*, 2011). However, CPR is not necessarily required as electrons can be transferred directly to the haem iron of the cytochrome P450 as hydrogen peroxide can be used instead of oxygen to produce the ferryloxy species of the haem required for the hydroxylation of the substrate to occur (Krishnan *et al.*, 2011). Although, utilising a fusion protein would mimic electron transfer *in vivo* allowing for greater understanding of the activity of cytochrome P450 in the cell. This would require the production of an artificial cytochrome P450-CPR fusion protein as few have been discovered in nature. Many artificial cytochrome P450-CPR fusion proteins have been produced including CYP4A1-CPR (Alterman *et al.*, 1995; Chaurasia *et al.*, 1995), human CYP1A1-rat CPR (Chun *et al.*, 1996), CYP2D6-CPR (Deeni *et al.*, 2001), CYP51-CPR (Kitahama *et al.*, 2009) and CYP71B1-CPR (Lamb *et al.*, 1998). These fusion proteins have been used in reconstitution assays and are ideal for enhancing the understanding of how cytochromes P450 and CPR interact with each other and the substrate to drive enzymatic activity. Together with electrode immobilisation, these proteins can become an important tool in biotechnology.

The *in vivo* and *in vitro* experiments described above could be used to enhance Alk1L456S catalytic activity for the use of this protein in the production of platform chemicals from alkane and fatty acids. However, the role of this protein and other CYP52s in yeast and fungi is not fully understood.

This study has shown that Alk1L456S is able to bind to sterol substrates producing type I binding spectra. This is the first time a cytochrome P450 from yeast/fungi, other than CYP51 and CYP61, has been shown to interact with these substrates thus suggesting a greater role for these enzymes in the physiology of the cell. Interestingly, other cytochromes P450 from other species have also been shown to

bind to (and metabolise) sterols, including CYP125 from *M. tuberculosis*, which has been shown to be involved in the hydroxylation of cholesterol (Capyk *et al.*, 2009; Ouellet *et al.*, 2010). *M. tuberculosis* is able to utilise local extracellular stores of cholesterol as the sole carbon source for growth (Pandey & Sassetti, 2008). The cholesterol is transported into the cell, where it is subjected to hydroxylation by CYP125 to form 26-hydroxycholesterol, which is further broken down to form pyruvate. This molecule can then be subjected to the Krebs cycle and used for growth (Pandey & Sassetti, 2008; Capyk *et al.*, 2009). The role of CYP125 in the metabolism of cholesterol in *M. tuberculosis* may indicate a potential role for Alk1L456S in the utilisation of cholesterol for growth in *C. albicans*. This, in turn, may link the pathogenicity of the organism in humans to cholesterol metabolism by Alk1L456S, particularly if the protein can be shown to hydroxylate the substrate in further metabolism studies.

Alk1L456S has also been shown in this study to bind tightly to azole antifungal drugs producing type II binding spectra. These drugs are known to target CYP51 enzymes, inhibiting the production of ergosterol in yeast, resulting in fungistasis (Strushkevich *et al.*, 2010). Comparison of the $[azole]_{0.5}$ values for Alk1L456S to those for CYP51s from *C. albicans* and *Phanerochaete chrysosporium* showed that Alk1L456S has a greater binding affinity for ketoconazole and itraconazole than CYP51 from *P. chrysosporium* suggesting a potential role for Alk1L456S in antifungal treatment and possibly the resistance of *C. albicans* to these drugs. This may also mean that these drugs may interfere with the growth of *C. albicans* on hydrocarbons or may inhibit pathogenicity if Alk1L456S is shown to hydroxylate cholesterol for growth in a similar manner to CYP125.

In summary, Alk1 was shown to have the greatest homology to the archetypal alkane-assimilating cytochrome P450, CYP52A3, of the five putative CYP52s identified in *C. albicans* using a bioinformatic approach. Although, the native *ALK1* gene sequence was unable to be used for expression to an active cytochrome P450 form, substitution of the CUG codon resulted in the expression of protein with a reduced carbon peak at 447nm, which is indicative of cytochrome P450 protein. The resultant Alk1L456S protein was purified to homogeneity and was shown to bind to lauric acid, myristic acid and palmitic acid, producing a type I binding spectra.

Although, reconstitution assays utilising Alk1L456S and palmitic acid failed to show product formation, *in vivo* experiments with *S. cerevisiae* AH22 cells transformed with YEp51:Alk1L456S_yeast plasmid were shown to utilise hexadecane for growth. This showed that Alk1L456S was involved in the metabolism of hexadecane as *S. cerevisiae* is unable to utilise alkanes for growth and the transformed cells showed a similar lipid composition to the positive control (YEp51:CYP52A3_yeast transformed *S. cerevisiae* cells), which occurred as a result of palmitic acid accumulation. This study was the first time Alk1L456S has been shown to be involved in alkane metabolism, suggesting a potential role for this protein in biotechnology, particularly if the product of this enzyme can be identified in further reconstitution assays and whole cell experiments. Alk1L456S was also shown to be the first yeast cytochrome P450, other than the biosynthetic CYP51 and CYP61, to bind to sterol substrates. Further metabolism studies will show if and how this protein metabolises these substrates, possibly giving rise to a greater understanding of the role of Alk1L456S in *C. albicans* growth and possibly pathogenesis. Alk1L456S was also shown to bind tightly with azole antifungal drugs producing type II binding spectra, thus indicating a potential role for this enzyme in the resistance of *C. albicans* to these drugs and possible inhibition of growth on hydrocarbons. Therefore, this study has shown a potential role for Alk1L456S in biotechnology and has indicated a possible physiological role for CYP52s in sterol metabolism, which was previously unknown.

7.1 Summary of Research

- Using a bioinformatic approach, 10 potential cytochromes P450 were identified in *C. albicans*. Of these, 5 were shown to be putative CYP52s. This is in contrast to Kim *et al.* (2007) and Park *et al.* (2011) who suggest there is only one CYP52 enzyme in *C. albicans*.
- Alk1 was shown to have the greatest homology to CYP52A3. Although, reconstitution assays failed to show Alk1 enzymatic activity in the presence of alkanes or fatty acids, *in vivo* experiments showed Alk1 is involved in the hydroxylation of the long-chain alkane, hexadecane. This suggests a role for Alk1 in alkane metabolism, membrane lipid formation and growth.

- Alk1 was shown to bind to sterols producing type I spectra. This is the first time a fungal cytochrome P450, other than CYP51 and CYP61, has been shown to interact with these substrates. The ability of Alk1 to bind to sterols may indicate a potential role for this enzyme in the virulence of *C. albicans* as other cytochromes P450, e.g. CYP125 from *M. tuberculosis*, are able to hydroxylate cholesterol to produce a precursor for growth (Pandey & Sassetti, 2008; Capyk *et al.*, 2009).
- Alk1 was also shown to bind tightly with azole antifungal drugs producing type II binding spectra. This suggests a potential role for Alk1 in azole resistance as the enzyme may bind to or modify these drugs preventing them from inhibiting CYP51 activity. Also, it may result in the possible inhibition of *C. albicans* growth on hydrocarbons.

7.2 Future Research

Alk1 has a potential use in biotechnology and further experiments could be undertaken to enhance its uses.

- Further experiments could be undertaken to show Alk1 activity at the *in vitro* level. Radiolabelled alkanes, fatty acids and sterols could be used as the substrates in reconstitution assays to help identify any products produced by Alk1 catalysed reactions.
- IC_{50} experiments could be undertaken to determine if the azole antifungal drugs that bind tightly to Alk1 do inhibit this enzyme. This would also allow a direct comparison to be made between Alk1 and Alk8 from *C. albicans* (as well as other cytochromes P450) regarding the interaction of azole drugs between these enzymes.
- The *S. cerevisiae* cells containing the plasmid YEp51:Alk1L456S_yeast could be used for further whole-cell experiments to produce large yields of fatty alcohols and/or diacids. For instance, disruption of the genes encoding for acyl CoA oxidase (which is involved in the β-oxidation pathway) may result in increased diacid production. Fatty alcohol production could also be enhanced in this strain by disrupting alcohol oxidase/dehydrogenase and aldehyde dehydrogenase activity preventing any possible conversion of the hydrocarbon to its respective aldehyde and diacid. The products of these

experiments could then potentially be used as platform chemicals in the production of bioplastics and other industrial products.

- An Alk1-CPR fusion protein could be constructed, which would be used to understand how CPR and Alk1 interact with each other and the substrate to drive enzymatic activity in the cell.

Chapter 8: References

Abrams, E.S., Murdaugh, S.E. and Lerman, L.S. (1990) Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. *Genomics*; 7(4), 463-475.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) *Molecular Biology of the Cell.* Fourth Edition, Garland Science, New York.

Ally, R., Schürmann, D., Kreisel, W., Carosi, G., Aguirrebengoa, K., Dupont, B., Hodges, M., Troke, P., Romero, A.J. and the Esophageal Candidiasis Study Group (2001) A randomized, double-blind, double-dummy, multicenter trial of voriconazole and fluconazole in the treatment of esophageal candidiasis in immunocompromised patients. *Clinical Infectious Diseases*; 33(9), 1447-1454.

Alterman, M.A., Chaurasia, C.S., Lu, P., Hardwick, J.P. and Hanzlik, R.P. (1995) Fatty acid discrimination and ω-hydroxylation by cytochrome P450 4A1 and a cytochrome P4504A1/NADPH-P450 reductase fusion protein. *Archives of Biochemistry and Biophysics*; 320(2), 289-296.

Altschul, S.F. (1991) Amino acid substitution matrices from an information theoretic perspective. *Journal of Molecular Biology*; 219(3), 555-565.

Altschul, S.F. and Gish, W. (1996) Local alignment statistics. *Methods in Enzymology*; 266, 460-480.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*; 25(17), 3389-3402.

Amann, E., Brosius, J. and Ptashne, M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli. Gene*; 25(2-3), 167-178.

Amet, Y., Berthou, F., Baird, S., Dreano, Y., Bail, J.P. and Menez, J.F. (1995) Validation of the (ω-1)-hydroxylation of lauric acid as an *in vitro* substrate probe for human liver CYP2E1. *Biochemical Pharmacology*; 50(11), 1775-1782.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J. and Zdobnov, E.M. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research*; 29(1), 37-40.

Apweiler, R., Bairoch, A. and Wu, C.H. (2004) Protein sequence databases. *Current Opinion in Chemical Biology*; 8, 76-80.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. and Yeh, L.S. (2004b) UniProt: the univeral protein knowledgebase. *Nucleic Acids Research*; 32, D115-D119.

Arase, M., Waterman, M.R. and Kagawa, N. (2006) Purification and characterization of bovine steriod 21-hydroxylase (P450c21) efficiently expressed in *Escherichia coli. Biochemical and Biophysical Research Communications*; 344 (1), 400-405.

Arié, J.-P., Miot, M., Sasson, N. and Betton, J.-M. (2006) Formation of active inclusion bodies in the periplasm of *Escherichia coli*. *Molecular Microbiology*; 62(2), 427-437.

Ashburner, M. And Goodman, N. (1997) Informatics – genome and genetic databases. *Current Opinion in Genetics & Development*; 7(6), 750-756.

Attwood, T.K., Croning, M.D., Flower, D.R., Lewis, A.P., Mabey, J.E., Scordis, P., Selley, J.N. and Wright, W. (2000) PRINTS-S: the database formally known as PRINTS. *Nucleic Acids Research*; 28(1), 225-227.

Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A. and Zygouri, C. (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research*; 31(1), 400-402.

Ball, S.E., Maurer, G., Zollinger, M., Ladona, M. and Vickers, A.E. (1992) Characterization of the cytochrome P-450 gene family responsible for the N-dealkylation of the ergot alkaloid CQA 206-291 in humans. *Drug Metabolism and Disposition*; 20(1), 56-63.

Barnes, H.J., Arlotto, M.P. and Waterman, M.R. (1991) Expression and enzymatic activity of recombinant cytochrome P450 17α-hydroxylase in *Escherichia coli*. *PNAS*; 88(13), 5597-5601.

Barnes, H.J. (1996) Maximizing expression of eukaryotic cytochrome P450s in *Escherichia coli*. *Methods in Enzymology*; 272, 3-14.

Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) A different genetic code in human mitochondria. *Nature*; 282(5735), 189-194.

Bartroli, J., Turmo, E., Algueró, M., Boncompte, E., Vericat, M.L., García-Rafanell, J. and Forn, J. (1995) Synthesis and antifungal activity of new azole derivatives containing an N-acylmorpholine ring. *Journal of Medicinal Chemistry*; 38(20), 3918-3932.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L. (2000) The Pfam protein families database. *Nucleic Acids Research*; 28(1), 263-266.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L.L. (2002) The Pfam protein families database. *Nucleic Acids Research*; 30(1), 276-280.

Basconcillo, L.S. and McCarry, B.E. (2008) Comparison of three GC/MS methodologies for the analysis of fatty acids in *Sinorhizobium meliloti*: development of a micro-scale, one-vial method. *Journal of Chromatography B*; 871(1), 22-31.

Beggs, J.D. (1978) Transformation of yeast by replicating hybrid plasmid. *Nature*, 275(5676), 104-109.

Bell, S.G., Dale, A., Rees, N.H. and Wong, L.L. (2010) A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*. *Applied Microbiology and Biotechnology*; 86(1), 163-175.

Benveniste, I., Tijet, N., Adas, F., Phillips, G., Salaün, J.-P. and Durst, F. (1998) CYP86A1 from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid ω-hydroxylase. *Biochemical and Biophysical Research Communications*; 243(3), 688-693.

Bergsson, G., Arnfinnsson, J., Steingrímsson, Ó. And Thormar, H. (2001) *In vitro* killing of *Candida albicans* by fatty acids and monoglycerides. *Antimicrobial Agents and Chemotherapy*; 45(11), 3209-3212.

Berman, J. and Sudbery, P.E. (2002) *Candida albicans*: a molecular revolution built on lessons from budding yeast. *Nature Reviews*; 3(12), 918-930.

Bieber, L.L. (1988) Carnitine. *Annual Review of Biochemistry*; 57, 261-283.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*; 7(6), 1513-1523.

Black, S.M., Harikrishna, J.A., Szklarz, G.D. and Miller, W.L. (1994) The mitochondrial environment is required for activity of the cholesterol side-chain cleavage enzyme, cytochrome P450scc. *PNAS*; 91(15), 7247-7251.

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V. Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. And Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*; 277(5331), 1453-1462.

Bligh, E.G. and Dyer, W.J. (1959) A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*; 37(8), 911-917.

Bloom, J.D. and Arnold, F.H. (2009) In the light of directed evolution: pathways of adaptive protein evolution. *PNAS*; 106, 9995-10000.

Boddupalli, S.S., Oster, T., Estabrook, R.W. and Peterson, J.A. (1992) Reconstitution of the fatty acid hydroxylation function of cytochrome P-450$_{BM-3}$ utilizing its individual recombinant hemo- and flavoprotein domains. *The Journal of Biological Chemistry*; 267(15), 10375-10380.

Bork, P. and Koonin, E.V. (1996) Protein sequence motifs. *Current Opinion in Structural Biology*; 6(3), 366-376.

Briza, P., Eckerstorfer, M. and Breitenbach, M. (1994) The sporulation-specific enzymes encoded by the *DIT1* and *DIT2* genes catalyze a two-step reaction leading to a soluble LL-dityrosyne-containing precursor of the yeast spore wall. *PNAS*; 91(10), 4524-4528.

Briza, P., Kalchhauser, H., Pittenauer, E., Allmaier, G. and Breitenbach, M. (1996) *N,N'*-Bisformyl dityrosine is an *in vivo* precursor of the yeast ascospore wall. *European Journal of Biochemistry*; 239(1), 124-131.

Broach, J.R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K. and Hicks, J.B. (1983) Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harbor Symposia on Quantitative Biology*; 47(2), 1165-1173.

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. *The Journal of Biological Chemistry*; 284(18), 11792-11805.

Brown, A.C., Eberl, M., Crick, D.C., Jomaa, H. and Parish, T. (2010) The nonmevalonate pathway of isoprenoid biosynthesis in *Mycobacterium tuberculosis* is essential and transcriptionally regulated by Dxs. *Journal of Bacteriology*; 192(9), 2424-2433.

Brown, A.J.P. and Gow, N.A.R. (1999) Regulatory networks controlling *Candida albicans* morphogenesis. *TRENDS in Microbiology*; 7(8), 333-338.

Brown, T.A. (1995) *Gene Cloning: An Introduction*. Third Edition, Nelson Thornes Ltd.

Buters, J.T.M., Korzekwa, K.R., Kunze, K.L., Omata, Y., Hardwick, J.P. and Gonzalez, F.J. (1994) cDNA-directed expression of human cytochrome P450

CYP3A4 using baculovirus. *Drug Metabolism and Disposition*; 22(5), 688-692.

Cabello-Hurtado, F., Batard, Y., Salaün, J.-P., Durst, F., Pinot, F. and Werck-Reichhart, D. (1998) Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids. *The Journal of Biological Chemistry*; 273(13), 7260-7267.

Cahoon, E.B., Ripp, K.G., Hall, S.E. and McGonigle, B. (2002) Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme from *Euphorbia lagascae* seed. *Plant Physiology*; 128(2), 615-624.

Candida Genome Database www.candidagenome.org

Cao, Z., Gao, H., Liu, M. and Jiao, P. (2006) Engineering the acetyl-CoA transportation system of *Candida tropicalis* enhances the production of dicarboxylic acid. *Biotechnology Journal*; 1(1), 68-74.

Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W.R. Jr., Eltis, L.D. and Mohn, W.W. (2009) Mycobacterial cytochrome P450 125 (CYP125) catalyzes the terminal hydroxylation of C27 steroids. *The Journal of Biological Chemistry*; 284(51), 35534-35542.

Carman, A.J., Vylkova, S. and Lorenz, M.C. (2008) Role of acetyl coenzyme A synthesis and breakdown in alternative carbon source utilization in *Candida albicans*. *Eukaryotic Cell*; 7(10), 1733-1741.

Carmichael, A.B. and Wong, L.-L. (2001) Protein engineering of *Bacillus megaterium* CYP102. *European Journal of Biochemistry*; 268(10), 3117-3125.

Carrapiso, A.I. and García, C. (2000) Development in lipid analysis: some new extraction techniques and in situ transesterification. *Lipids*; 35(11), 1167-177.

Casas, J.A. and García-Ochoa, F. (1999) Sophorolipid production by *Candida bombicola*: medium composition and culture methods. *Journal of Bioscience and Bioengineering*; 88(5), 488-494.

Chaurasia, C.S., Alterman, M.A., Lu, P. and Hanzlik, R.P. (1995) Biochemical characterization of lauric acid ω-hydroxylation by a CYP4A1/NADPH-cytochrome P450 reductase fusion protein. *Archives of Biochemistry and Biophysics*; 317(1), 161-169.

Christie, W.W. (1989) *Gas Chromatography and Lipids: A Practical Guide.* The Oily Press, Ayr.

Chin, T.W.F., Loeb, M. and Fong, I.W. (1995) Effects of an acidic beverage (Coca-Cola) on absorption of ketoconazole. *Antimicrobial Agents and Chemotherapy*; 39(8), 1671-1675.

Chun, Y.-J., Shimada, T. and Guengerich, F.P. (1996) Construction of a human cytochrome P450 1A1: rat NADPH-cytochrome P450 reductase fusion protein cDNA and expression in *Escherichia coli*, purification, and catalytic properties of the enzyme in bacterial cells and after purification. *Archives of Biochemistry and Biophysics*; 330(1), 48-58.

Chun, Y.J., Shimada, T., Sanchez-Ponce, R., Martin, M.V., Lei, L., Zhao, B., Kelly, S.L., Waterman, M.R., Lamb, D.C. and Guengerich, F.P. (2007) Electron transport pathway for a Streptomyces cytochrome P450: cytochrome P450 105D5-catalyzed fatty acid hydroxylation in *Streptomyces coelicolor* A3(2). *The Journal of Biological Chemistry*; 282(24), 17486-17500.

Cirino, P.C. and Arnold, F.H. (2002) Protein engineering of oxygenases for biocatalysis. *Current Opinion in Chemical Biology*; 6(2), 130-135.

Clark, B.J. and Waterman, M.R. (1991) The hydrophobic amino-terminal sequence of bovine 17-α hydroxylase is required for the expression of a functional hemoprotein in COS 1 cells. *The Journal of Biological Chemistry*; 266(9), 5898-5904.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. (2003) A vision for the future of genomics research. *Nature*; 422(6934), 835-847.

Corpet, F., Servant, F., Gouzy, J. and Kahn, D. (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Research*; 28(1), 267-269.

Cosme, J. and Johnson, E.F. (2000) Engineering microsomal cytochrome P450 2C5 to be a soluble monomeric enzyme. *The Journal of Biological Chemistry*; 275(4), 2545-2553.

Craft, D.L., Madduri, K.M., Eshoo, M. and Wilson, C.R. (2003) Identification and characterization of the CYP52 family of *Candida tropicalis* ATCC 20336, important for the conversion of fatty acids and alkanes to α,ω-dicarboxylic acids. *Applied and Environmental Mircobiology*; 69(10), 5983-5991.

Crick, F.H. (1968) The origin of the genetic code. *Journal of Molecular Biology*; 38(3), 367-379.

Cryle, M.J., Matovic, N.J. and De Voss, J.J. (2003) Products of cytochrome $P450_{BioI}$ (CYP107H1)-catalysed oxidation of fatty acids. *Organic Letters*; 5(18), 3341-3344.

Cryle, M.J. and Schlichting, I. (2008) Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *PNAS*; 105(41), 15696-15701.

Cytochrome P450 Homepage http://drnelson.uthsc.edu/CytochromeP450.html

Daiber, A., Shoun, H. and Ullrich, V. (2005) Nitric oxide reductase (P450nor) from *Fusarium oxysporum*. *Journal of Inorganic Biochemistry*; 99(1), 185-193.

de Boer, H.A., Comstock, L.J. and Vasser, M. (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. *PNAS*; 80(1), 21-25.

Deeni, Y.Y., Paine, M.J.I., Ayrton, A.D., Clarke, S.E., Chenery, R. and Wolf, C.R. (2001) Expression, purification, and biochemical characterization of a human cytochrome P450 CYP2D6-NADPH cytochrome P450 reductase fusion protein. *Archives of Biochemistry and Biophysics*; 396(1), 16-24.

Dickinson, F.M. and Wadforth, C. (1992) Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. *The Biochemical Journal*; 282(2), 325-331.

Dietrich, M., Eiben, S., Asta, C., Do, T.A., Pleiss, J. and Urlacher, V.B. (2008) Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monoxygenase from *Bacillus licheniformis*. *Applied Microbiology and Biotechnology*; 79, 931-940.

Dhar, M., Sepkovic, D.W., Hirani, V., Magnusson, R.P. and Lasker, J.M. (2008) ω-oxidation of 3-hydroxy fatty acids by the human CYP4F gene subfamily enzyme CYP4F11. *Journal of Lipid Research*; 49(3), 612-624.

DNA Data Bank of Japan (DDBJ) www.ddbj.nig.ac.jp

Dobritsa, A.A., Shrestha, J., Morant, M., Pinot, F., Matsuno, M., Swanson, R., Møller, B.L. and Preuss, D. (2009) CYP704B1 is a long-chain fatty acid ω-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis. *Plant Physiology*; 151(2), 574-589.

Döhr, O., Paine, M.J.I., Friedberg, T., Roberts, G.C.K. and Wolf, C.R. (2001) Engineering of a functional human NADH-dependent cytochrome P450 system. *PNAS*; 98(1), 81-86.

Dong, M.-S., Yamazaki, H., Guo, Z. and Guengerich, F.P. (1996) Recombinant human cytochrome P450 1A2 and an N-terminal-truncated form: construction, purification, aggregation properties, and interactions with flavodoxin, feredoxin, and NADPH-cytochrome P450 reductase. *Archives of Biochemistry and Biophysics*; 327(1), 11-19.

Duppel, W., Lebeault, J.-M. and Coon, M.J. (1973) Properties of a yeast cytochrome P-450-containing enzyme system which catalyzes the hydroxylation of fatty acids, alkanes, and drugs. *European Journal of Biochemistry*; 36(2), 583-592.

Eberhart, D.C. and Parkinson, A. (1991) Cytochrome P450 IIIA1 (P450p) requires cytochrome b5 and phospholipid with unsaturated fatty acids. *Archives of Biochemistry and Biophysics*; 291(2), 231-240.

Edlund, A.F., Swanson, R. and Preuss, D. (2004) Pollen and stigma structure and function: the role of diversity in pollination. *The Plant Cell*; 16, S84-S97.

Eschenfeldt, W.H., Zhang, Y., Samaha, H., Stols, L., Eirich, L.D., Wilson, C.R. and Donnelly, M.I. (2003) Transformation of fatty acids catalyzed by cytochrome P450 monooxygenase enzymes of *Candida tropicalis*. *Applied and Environmental Microbiology*; 69(10), 5992-5999.

European Bioinformatics Institute http://www.ebi.ac.uk

Fasan, R., Chen, M.M., Crook, N.C. and Arnold, F.H. (2007) Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angewandte Chemie International Edition*; 46(44), 8414-8418.

Fer, M., Corcos, L., Dréano, Y., Plée-Gautier, E., Salaün, J.-P., Berthou, F. and Amet, Y. (2008) Cytochromes P450 from family 4 are the main ω-hydroxylating enzymes in humans: CYP4F3B is the prominent player in PUFA metabolism. *Journal of Lipid Research*; 49(11), 2379-2389.

Fickers, P., Benetti, P.-H., Waché, Y., Marty, A., Mauersberger, S., Smit, M.S. and Nicaud, J.-M. (2005) Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Research*; 5(6-7), 527-543.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. And Ysebaert, M. (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*; 260(5551), 500-507.

Fisher, C.W., Caudle, D.L., Martin-Wixtrom, C., Quattrochi, L.C., Tukey, R.H., Waterman, M.R. and Estabrook, R.W. (1992) High-level expression of functional human cytochrome P450 1A2 in *Escherichia coli*. *The FASEB Journal*; 6(2), 759-764.

Fisher, M.B., Zheng, Y.-M. and Rettie, A.E. (1998) Positional specificity of rabbit CYP4B1 for ω-hydroxylation of short-medium chain fatty acids and hydrocarbons. *Biochemical and Biophysical Research Communications*; 248(2), 352-355.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A. Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F. Dougherty, B.A., Merrick, J.M. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenza* Rd. *Science*; 269(5223), 496-512.

Froehlich, J.E., Itoh, A. and Howe, G.A. (2001) Tomato allene oxide synthase and fatty acid hydroperoxide lyase, two cytochrome P450s involved in oxylipin metabolism, are targeted to different membranes of chloroplast envelope. *Plant Physiology*; 125(1), 306-317.

Fuchs, B., Süβ, R., Teuber, K., Eibisch, M. and Schiller, J. (2011) Lipid analysis by thin-layer chromatography – a review of the current state. *Journal of Chromatography A*; 1218(19), 2754-2774.

Fukuda, T., Imai, Y., Komori, M., Nakamura, M., Kusunose, E., Satouchi, K and Kusunose, M. (1994) Different mechanisms of regioselection of fatty acid hydroxylation by laurate (ω-1)-hydroxylation P450s, P450 2C2 and P450 2E1. *Journal of Biochemistry*; 115(2), 338-344.

Fungal Cytochrome P450 Database http://p450.riceblast.snu.ac.kr

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

Funhoff, E.G., Salzmann, J., Bauer, U., Witholt, B. and van Beilen, J.B. (2007) Hydroxylation and epoxidation reactions catalysed by CYP153 enzymes. *Enzyme and Microbial Technology*; 40(4), 806-812.

Garcia, A.N., Muniz, M.T., Souza e Silva, H.R., da Silva, H.A. and Athayde-Junior, L. (2009) CYP46 polymorphisms in Alzheimer's disease: a review. *Journal of Molecular Neuroscience*; 39(3), 342-345.

Garfinkel, D. (1958) Studies on pig liver microsomes. I. enzymic and pigment composition of different microsomal fractions. *Archives of Biochemistry and Biophysics*; 77(2), 493-509.

GenBank (NCBI) www.ncbi.nlm.nih.gov/Genbank

Gene3D http://gene3d.biochem.ucl.ac.uk/Gene3D/

Ghannoum, M.A. and Rice, L.B. (1999) Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clinical Microbiology Reviews*; 12(4), 501-517.

Gigon, P.L., Gram, T.E., and Gillette, J.R. (1969) Studies on the rate of reduction of hepatic microsomal cytochrome P-450 by reduced nicotinamide adenine dinucleotide phosphate: effect of drug substrates. *Molecular Pharmacology*; 5(2), 109-122.

Gillam, E.M.J., Guo, Z., Martin, M.V., Jenkins, C.M. and Guengerich, F.P. (1995) Expression of cytochrome P450 2D6 in *Escherichia coli,* purification, and spectral and catalytic characterization. *Archives of Biochemistry and Biophysics*; 319(2), 540-550.

Girhard, M., Schuster, S., Dietrich, M., Dürre, P. and Urlacher, V.B. (2007) Cytochrome P450 monooxygenase from *Clostridium acetobutylicum*: a new α-fatty acid hydroxylase. *Biochemical and Biophysical Research Communications*; 362(1), 114-119.

Girvan, H.M., Waltham, T.N., Neeli, R., Collins, H.F., McLean, K.J., Scrutton, N.S., Leys, D. and Munro, A.W. (2006) Flavocytochrome P450 BM3 and the origin of CYP102 fusion species. *Biochemical Society Transactions*; 34(6), 1173-1177.

Gish, W. (1996-2000) http://blast.wustl.edu

Gluzman, Y. (1981) SV40-transformed simian cells support the replication of early SV40 mutants. *Cell*; 23(1), 175-182.

Goa, K.L. and Barradell, L.B. (1995) Fluconazole. An update of its pharmacodynamic and pharmacokinetic properties and therapeutic use in

major superficial and systemic mycoses in immunocompromised patients. *Drugs*; 50(4), 658-690.

Golding, G.B. (2003) DNA and the revolutions of molecular evolution, computational biology, and bioinformatics. *Genome*; 46(6), 930-935.

Gonzalez, F.J., Kimura, S., Tamura, S. and Gelboin, H.V. (1991) Expression of mammalian cytochrome P450 using baculovirus. *Methods in Enzymology*; 206, 93- 99.

Gonzalez, F.J. and Korzekwa, K.R. (1995) Cytochrome P450 expression systems. *Annual Review of Pharmacology and Toxicology*; 35, 369-390.

Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *The Journal of Biological Chemistry*; 267(1), 83-90.

Gow, N.A.R. (1994) Growth and guidance of the fungal hypha. *Microbiology*; 140(12), 3193-3205.

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), 24-29.

Greilhuber, J., Borsch, T., Müller, K., Worberg, A., Porembski, S. and Barthlott, W. (2006) Smallest angiosperm genomes found in *Lentibulariaceae*, with chromosomes of bacterial size. *Plant Biology (Stuttgart)*; 8(6), 770-777.

Groner, A. (1986) Specificity and safety of baculoviruses in *The Biology of Baculoviruses.* Granados, R.R. and Federici, B.A. (eds)., CRC Press, Boca Raton, Florida.

Groves, J. T. (2005) Models and Mechanisms of Cytochrome P450 Action. *Cytochrome P450: Structure, Mechanism and Biochemistry.* Ortiz De Montellano, P. R. (ed.), Third Edition, Kluwer Academic/Plenum Publishers, New York.

Guengerich, F. P. and Macdonald, T. L. (1984) Chemical mechanisms of catalysis by cytochrome P-450: a unified view. *Accounts of Chemical Research*; 17(1), 9-16.

Guengerich, F. P. (1991) Reactions and significance of cytochrome P-450 enzymes. *The Journal of Biological Chemistry*; 266(16), 10019-10022.

Guengerich, F.P., Brian, W.R., Sari, M.-A. and Ross, J.T. (1991) Expression of mammalian cytochrome P450 enzymes using yeast-based vectors. *Methods in Enzymology*; 206, 130-145.

Guengerich, F. P. (2007) Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *Journal of Biochemical and Molecular Toxicology*; 21(4), 163-168.

Gupta, A.K., Sauder, D.N. and Shear, N.H. (1994) Antifungal agents: an overview. Part I. *Journal of the American Academy of Dermatology*; 30(5) 677-698.

Gurvitz, A., Hamilton, B., Ruis, H. and Hartig, A. (2001) Peroxisomal degradation of *trans*-unsaturated fatty acids in the yeast *Saccharomyces cerevisiae. The Journal of Biological Chemistry*; 276(2), 895-903.

Gustafsson, M.C.U., Roitel, O., Marshall, K.R., Noble, M.A., Chapman, S.K., Pessegueiro, A., Fulco, A.J., Cheesman, M.R., von Wachenfeldt, C. and Munro, A.W. (2004) Expression, purification, and characterization of *Bacillus subtilis* cytochromes P450 CYP102A2 and CYP102A3: flavocytochrome homologues of P450 BM3 from *Bacillus megaterium. Biochemistry*; 43(18), 5474-5487.

Haft, D.H., Selengut, J.D. and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Research*; 31(1), 371-373.

Hall, B.G. (2001) *Phylogenetic Trees Made Easy: A How-To Manual for Molecular Biologists*. Sinauer Associates Inc.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*; 41, 95-98.

Han, J., Clement, J.M., Li, J., King, A., Ng, S. and Jaworski, J.G. (2010) The cytochrome P450 CYP86A22 is a fatty acyl-CoA ω-hydroxylase essential for Estolide synthesis in the stigma of *Petunia hybrida. The Journal of Biological Chemistry*; 285(6), 3986-3996.

Hanahan, D. (1985) Techniques for transformation of *E. coli. DNA Cloning. Volume 1. A Practical Approach*. Glover, D.M. (ed.), IRL Press Limited, Oxford, England, 109-135.

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

Hara, A., Ueda, M., Matsui, T., Arie, M., Saeki, H., Matsuda, H., Furuhashi, K., Kanai, T. and Tanaka, A. (2001) Repression of fatty-acyl-CoA oxidase-encoding gene expression is not necessarily a determinant of high-level production of dicarboxylic acids in industrial dicarboxylic-acid-producing *Candida tropicalis. Applied Microbiology and Biotechnology*; 56(3-4), 478-485.

Hargrove, T.Y., Wawrzak, Z., Liu, J., Nes, W.D., Waterman, M.R. and Lepesheva, G.I. (2011) Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14α-demethylase (CYP51) from *Leishmania infantum. The Journal of Biological Chemistry*; 286(30), 26838-26848.

Harmon, S.D., Fang, X., Kaduce, T.L., Hu, S., Gopal, V.R., Falck, J.R. and Spector, A.A. (2006) Oxygenation of ω-3 fatty acids by human cytochrome P450 4F3B: effect on 20-hydroxyeicosatetraenoic acid production. *Prostaglandins, Leukotrienes and Essential Fatty Acids*; 75(3), 169-177.

Hawkes, D.B., Adams, G.W., Burlingame, A.L., Ortiz de Montellano, P.R. and De Voss. J.J. (2002) Cytochrome P450(cin) (CYP176A), isolation, expression, and characterization. *The Journal of Biological Chemistry*; 277(31), 27725-27732.

He, F. and Chen, Y.T. (2005) Cloning and heterologous expression of the NADPH cytochrome P450 oxidoreductase genes from an industrial dicarboxylic acid-producing *Candida tropicalis. Yeast,* 22(6), 481-491.

He, X., Cryle, M.J., De Voss, J.J. and Ortiz de Montellano, P.R. (2005) Calibration of the channel that determines the ω-hydroxylation regiospecificity of cytochrome P4504A1. *The Journal of Biological Chemistry*; 280(24), 22697-22705.

Hettema, E.H., van Roermund, C.W.T., Distel, B., van den Berg, M., Vilela, C., Rodrigues-Pousada, C., Wanders, R.J.A. and Tabak, H.F. (1996) The ABC transporter proteins Pat1 and Pat2 are required for import of long-chain fatty acids into peroxisomes of *Saccharomyces cerevisiae. The EMBO Journal*; 15(15), 3813-3822.

Hettema, E.H. and Tabak, H.F. (2000) Transport of fatty acids and metabolites across the peroxisomal membrane. *Biochimica et Biophysica Acta*; 1486(1), 18-27.

Hitchcock, C.A. (1993) Resistance of *Candida albicans* to azole antifungal agents. *Biochemical Society Transactions*; 21(4), 1039-1047.

HMMTOP - http://www.enzim.hu/hmmtop/

Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L. and Franke, R. (2008) The *Arabidopsis* cytochrome P450 CYP86A1 encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental Botany*; 59(9), 2347-2360.

Hoffman, H.L. and Rathbun, R.C. (2002) Review of the safety and efficiacy of voriconazole. *Expert Opinion on Investigational Drugs*; 11(3), 409-429.

Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Research*; 27(1), 215-219.

Hunter, D.J., Roberts, G.A., Ost, T.W., White, J.H., Muller, S., Turner, N.J., Flitsch, S.L. and Chapman, S.K. (2005) Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Letters*; 579(10), 2215-2220.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A.F., Selengut, J.D., Sigrist, C.J.A., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H. and Yeats, C. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Research*; 37, D211-D215.

Imai, Y., Matsunaga, I., Kusunose, E. and Ichihara, K. (2000) Unique heme environment at the putative distal region of hydrogen peroxide-dependent fatty acid α-hydroxylase from *Sphingomonas paucimobilis* (peroxygenase P450(SPα). *Journal of Biochemistry*; 128(2), 189-194.

Imaishi, H., Matsuo, S., Swai, E. and Ohkawa, H. (2000) CYP78A1 preferentially expressed in developing inflorescences of *Zea mays* encoded a cytochrome P450-dependent lauric acid 12-monooxygenase. *Bioscience, Biotechnology and Biochemistry*; 64(8), 1696-1701.

Imaishi, H. and Ohkawa, H. (2002) Cytochrome P450 species specifically expressed in flower buds metabolize fatty acids. *International Congress Series*; 1233, 115-120.

Imaishi, H. and Petkova-Andonova, M. (2007) Molecular cloning of CYP76B9, a cytochrome P450 from *Petunia hybrida*, catalyzing the ω-hydroxylation of capric acid and lauric acid. *Bioscience, Biotechnology and Biochemistry*; 71(1), 104-113.

Imaoka, S., Ogawa, H., Kimura, S. and Gonzalez, F.J. (1993) Complete cDNA sequence and cDNA-directed expression of CYP4A11, a fatty acid ω-hydroxylase expressed in human kidney. *DNA and Cell Biology*; 12(10), 893-899.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*; 409, 860-921.

Itoh, A. and Howe, G.A. (2001) Molecular cloning of a divinyl ether synthase. Identification as a CYP74 cytochrome P-450. *The Journal of Biological Chemistry*; 276(5), 3620-3627.

Iverson, S.J., Lang, S.L. and Cooper, M.H. (2001) Comparison of the Bligh and Dyer and Folch methods for total lipid determination in a broad range of marine tissue. *Lipids*; 36(11), 1283-1287.

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. *The Journal of Biological Chemistry*; 277(49), 46959-46965.

Jackson, R.G., Rylott, E.L., Fournier, D., Hawari, J. and Bruce, N.C. (2007) Exploring the biochemical properties and remediation applications of the unusual explosive-degrading P450 system XplA/B. *PNAS*; 104(43), 16822-16827.

Jacolot, F., Simon, I., Dreano, Y., Beaune, P., Riche, C. and Berthou, F. (1991) Identification of the cytochrome P450IIIA family as the enzymes involved in the N-demethylation of tamoxifen in human liver microsomes. *Biochemical Pharmacology*; 41(12), 1911-1919.

Jiao, P., Ma, S., Hua, Y., Huang, Y. and Cao, Z. (2000) Isolation and enzyme determination of *Candida tropicalis* mutants for DCA production. *The Journal of General and Applied Microbiology*; 46(5), 245-249.

Johnston, J.B., Kells, P.M., Podust, L.M. and Ortiz de Montellano, P.R. (2009) Biochemical and structural characterization of CYP124: a methyl-branched lipid ω-hydroxylase from *Mycobacterium tuberculosis*. *PNAS*; 106(49), 20687-20692.

Jones, J.P., O'Hare, E.J. and Wong, L.-L. (2001) Oxidation of polychlorinated benzenes by genetically engineered CYP101 (cytochrome $P450_{cam}$). *European Journal of Biochemistry*; 268(5), 1460-1467.

Juárez, M., Polvillo, O., Contò, M., Ficco, A., Ballico, S. and Failla, S. (2008) Comparison of four extraction/methylation analytical methods to measure fatty acid composition by gas chromatography in meat. *Journal of Chromatography A*; 1190(1-2), 327-332.

Jukes, T.H. and Osawa, S. (1996) CUG codons in Candida spp. *Journal of Molecular Evolution*; 42(2), 321-322.

Kabara, J.J., Swieczkowski, D.M., Conley, A.J. and Truant, J.P. (1972) Fatty acids and derivatives as antimicrobial agents. *Antimicrobial Agents and Chemotherapy*; 2(1), 23-28.

Kahn, R.A., Le Bouquin, R., Pinot, F., Benveniste, I. and Durst, F. (2001) A conservative amino acid substitution alters the regiospecificity of CYP94A2, a fatty acid hydroxylase from the plant *Vicia sativa*. *Archives of Biochemistry and Biophysics*; 391(2), 180-187.

Kalsotra, A. and Strobel, H.W. (2006) Cytochrome P450 4F subfamily: at the crossroads of eicosanoid and drug metabolism. *Pharmacology and Therapeutics*; 112(3), 589-611.

Kanayama, N., Ueda, M., Atomi, H. and Tanaka, A. (1998) Genetic evaluation of physiological functions of thiolase isozymes in the *n*-alkane-assimilating yeast *Candida tropicalis*. *Journal of Bacteriology*; 180(3), 690-698.

Kandel, S., Morant, M., Benveniste, I., Blée, E., Werck-Reichhart, D. and Pinot, F. (2005) Cloning, functional expression, and characterization of CYP709C1, the first sub-terminal hydroxylase of long chain fatty acid in plants. Induction by chemicals and methyl jasmonate. *The Journal of Biological Chemistry*; 280(43), 35881-35889.

Kandel, S., Sauveplane, V., Compagnon, V., Franke, R., Millet, Y., Schreiber, L., Werck-Reichhart, D. and Pinot, F. (2007) Characterization of a methyl jasmonate and wounding-responsive cytochrome P450 of *Arabidopsis thaliana* catalyzing dicarboxylic fatty acid formation *in vitro*. *The FEBS Journal*; 274(19), 5116-5127.

Karam, W.G. and Chiang, J.Y. (1994) Expression and purification of human cholesterol 7α-hydroxylase in *Escherichia coli*. *Journal of Lipid Research*; 35(7), 1222-1231.

Karlin, S. and Altschul, S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *PNAS*; 87(6), 2264-2268.

Kato, T., Yamaguchi, Y., Namai, T. and Hirukawa, T. (1993) Oxygenated fatty acids with anti-rice blast fungus activity in rice plants. *Bioscience, Biotechnology and Biochemistry*; 57(2), 283-287.

Kauffman, C.A. and Carver, P.L. (1997) Antifungal agents in the 1990s. Current status and future developments. *Drugs*; 53(4), 539-549.

Kawaguchi, Y., Honda, H., Taniguchui-Morimura, J. and Iwasaki, S. (1989) The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. *Nature*; 341, 164-166.

Kaya, M., Matsumura, K., Higashida, K., Hata, Y., Kawato, A., Abe, Y., Akita, O., Takaya, N. and Shoun, H. (2004) Cloning and enhanced expression of the cytochrome P450nor gene (*nicA*; *CYP55A5*) encoding nitric oxide reductase from *Aspergillus oryzae*. *Bioscience, Biotechnology and Biochemistry*; 68(10), 2040-2049.

Kelly, S.L., Lamb, D.C., Corran, A.J., Baldwin B.C., Parks, L.W. and Kelly, D.E. (1995) Purification and reconstitution of activity of *Saccharomyces cerevisiae* P450 61, a sterol Δ^{22}-desaturase. *FEBS Letters*; 377(2), 217-220.

Kelly, S.L., Lamb, D.C., Baldwin, B.C., Corran, A.J. and Kelly, D.E. (1997) Characterization of *Sacchromyces cerevisiae* CYP61, sterol Δ^{22}-desaturase, and inhibition by azole antifungal agents. *The Journal of Biological Chemistry*; 272(15), 9986-9988.

Kelly, S.L., Lamb, D.C., Cannieux, M., Greetham, D., Jackson, C.J., Marczylo, T., Ugochukwu, C. and Kelly D.E. (2001) An old activity in the cytochrome P450 superfamily (CYP51) and a new story of drugs and resistance. *Biochemical Society Transactions*; 29(2), 122-128.

Kelly, S.L., Lamb, D.C., Jackson, C.J., Warrilow, A.G.S. and Kelly, D.E. (2003) The biodiversity of microbial cytochromes P450. *Advances in Microbial Physiology*; 47, 131-186.

Kelly, S.L., Kelly, D.E., Jackson, C.J., Warrilow, A.G.S. and Lamb, D.C. (2005) Cytochrome P450: structure, mechanism, and biochemistry. *Cytochrome P450: Structure, Mechanism and Biochemistry.* Ortiz De Montellano, P. R. (ed.), Third Edition, Kluwer Academic/Plenum Publishers, New York.

Kikuta, Y., Kusunose, E., Ito, M. and Kusunose, M. (1999) Purification and characterization of recombinant rat hepatic CYP4F1. *Archives of Biochemistry and Biophysics*; 369(2), 193-196.

Kim, D., Cryle, M.J., De Voss, J.J. and Ortiz de Montellano, P.R. (2007) Functional expression and characterization of cytochrome P450 52A21 from *Candida albicans*. *Archives of Biochemistry and Biophysics*; 464(2), 213-220.

Kitahama, Y., Nakamura, M., Yoshida, Y. and Aoyama, Y. (2009) The construction and characterization of self-sufficient lanosterol 14-demethylase fusion

proteins consisting of yeast CYP51 and its reductase. *Biological and Pharmaceutical Bulletin*; 32(4), 558-563.

Klingenberg, M. (1958) Pigments of rat liver microsomes. *Archives of Biochemistry and Biophysics*; 75(2), 376-386.

Kogure, T., Horiuchi, H., Matsuda, H., Arie, M., Takagi, M. and Ohta, A. (2007) Enhanced induction of cytochrome P450alk that oxidize methyl-ends of n-alkanes and fatty acids in the long-chain dicarboxylic acid-hyperproducing mutant of *Candida maltosa*. *FEMS Microbiology Letters*; 271(1), 106-111.

Konishi, M., Fukuoka, T., Morita, T., Imura, T. and Kitamoto, D. (2008) Production of new types of sophorolipids by *Candida batistae*. *Journal of Oleo Science*; 57(6), 359-369.

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

Kothavade, R.J., Kura, M.M., Valand, A.G. and Panthaki, M.H. (2010) *Candida tropicolis*: its prevalence, pathogenicity and increasing resistance to fluconazole. *Journal of Medical Microbiology*; 59(8), 873-880.

Kragt, A., Benne, R. and Distel, B. (2006) *Protein degradation. Vol.3: Cell Biology of the Ubiquitin-Proteasome System.* Mayer, R.J., Ciechanover, A. and Rechsteiner M. (eds.), Wiley-VCH.

Krishnan, S., Schenkman, J.B. and Rusling, J.F. (2011) Bioelectronic delivery of electrons to cytochrome P450 enzymes. *The Journal of Physical Chemistry B*; 115(26), 8371-8380.

Kriventseva, E.V., Biswas, M. and Apweiler, R. (2001) Clustering and analysis of protein families. *Current Opinion in Structural Biology*; 11(3), 334-339.

Kudo, T., Tomura, D., Liu, D.L., Dai, X.Q. and Shoun, H. (1996) Two isozymes of P450nor of *Cylindrocarpon tonkinense*; molecular cloning of the cDNAs and genes, expressions in the yeast, and the putative NAD(P)H-binding site. *Biochimie*; 78(8-9), 792-799.

Kunau, W.H., Bühne, S., de la Garza, M., Kionka, C., Mateblowski, M., Schultz-Borchard, U. and Thieringer, R. (1988) Comparative enzymology of β-oxidation. *Biochemical Society Transactions*; 16(3), 418-420.

Kusano, K., Kagawa, N., Sakaguchi, M., Omura, T. and Waterman, M.R. (2001) Importance of a proline-rich sequence in the amino-terminal region for correct folding of mitochondrial and soluble microbial P450s. *Journal of Biochemistry*; 129(2), 271-277.

Kusano, K., Sakaguchi, M., Kagawa, N., Waterman, M.R. and Omura, T. (2001b) Microsomal P450s use specific proline-rich sequences for efficient folding, but not for maintenance of the folded structure. *Journal of Biochemistry*; 129(2), 259-269.

Lai, M.H. and Kirsch, D.R. (1989) Nucleotide sequence of cytochrome P450 L1A1 (lanosterol 14α-demethylase) from *Candida albicans*. *Nucleic Acids Research*; 17(2), 804.

Lamb, D.C., Kelly, D.E., Hanley, S.Z., Mehmood, Z. and Kelly, S.L. (1998) Glyphosate is an inhibitor of plant cytochrome P450: functional expression of *Thlaspi arvensae* cytochrome P45071B1/reductase fusion protein in *Escherichia coli*. *Biochemical and Biophysical Research Communications*; 244(1), 110-114.

Lamb, D.C., Kelly, D.E., Waterman, M.R., Stromstedt, M., Rozman, D. and Kelly, S.L. (1999) Characteristics of the heterologously expressed human lanosterol 14α-demethylase (other names: P45014DM, CYP51, P45051) and inhibition of the purified human and *Candida albicans* CYP51 with azole antifungal agents. *Yeast*; 15(9), 755-763.

Lamb, D.C., Kaderbhai, N.N., Venkateswarlu, K., Kelly, D.E., Kelly, S.L. and Kaderbhai, M.A. (2001) Human sterol 14α-demethylase activity is enhanced by the membrane-bound state of cytochrome b(5). *Archives of Biochemistry and Biophysics*; 395(1), 78-84.

Lamb, D.C., Warrilow, A.G.S., Venkateswarlu, K., Kelly, D.E. and Kelly, S.L. (2001b) Activities and kinetic mechanisms of native and soluble NADPH-cytochrome P450 reductase. *Biochemical and Biophysical Research Communications*; 286(1), 48-54.

Lamb, D.C., Lei, L., Zhao, B., Yuan, H., Jackson, C.J., Warrilow, A.G.S., Skaug, T., Dyson, P.J., Dawson, E.S., Kelly, S.L., Hachey, D.L. and Waterman, M.R. (2010) *Streptomyces coelicolor* A3(2) CYP102 protein, a novel fatty acid hydroxylase encoded as a heme domain without an N-terminal redox partner. *Applied and Environmental Microbiology*; 76(6), 1975-1980.

Larson, J.R., Coon, M.J. and Porter, T.D. (1991) Alcohol-inducible cytochrome P-450IIE1 lacking the hydrophobic NH_2-terminal segment retains catalytic activity and is membrane-bound when expressed in *Escherichia coli*. *The Journal of Biological Chemistry*; 266(2), 7321-7324.

Lasker, J.M., Chen, W.B., Wolf, I., Bloswick, B.P., Wilson, P.D. and Powell, P.K. (2000) Formation of 20-hydroxyeicosatetraenoic acid, a vasoactive and natriuretic eicosanoid, in human kidney. *The Journal of Biological Chemistry*; 275(6), 4118-4126.

Lau, S.M., Harder, P.A. and O'Keefe, D.P. (1993) Low carbon monoxide affinity allene oxide synthase is the predominant cytochrome P450 in many plant tissues. *Biochemistry*; 32, 1945-1950.

Lebeault, J.M., Lode, E.T. and Coon, M.J. (1971) Fatty acid and hydrocarbon hydroxylation in yeast: role of cytochrome P-450 in *Candida tropicalis*. *Biochemical and Biophysics Research Communications*; 42(3), 413-419.

Le Bouquin, R., Skrabs, M., Kahn, R., Benveniste, I., Salaün, J.-P., Schreiber, L., Durst, F. and Pinot, F. (2001) CYP94A5, a new cytochrome P450 from *Nicotiana tabacum* is able to catalyze the oxidation of fatty acids to the ω-alcohol and to the corresponding diacid. *European Journal of Biochemistry*; 268(10), 3083-3090.

Lee, D.-S., Yamada, A., Sugimoto, H., Matsunaga, I., Ogura, H., Ichihara, K., Adachi, S.-I., Park, S.-Y. and Shiro, Y. (2003) Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from *Bacillus subtilis*. *The Journal of Biological Chemistry*; 278(11), 9761-9767.

Lesk, A.M. (2005) *Introduction to Bioinformatics*. Second Edition, Oxford University Press.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J. and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Research*; 34, D257-D260.

Li, A., Jiao, X., Munier, F.L., Schorderet, D.F., Yao, W., Iwata, F., Hayakawa, M., Kanai, A., Chen, M.S., Lewis, R.A., Heckenlively, J., Weleber, R.G., Traboulsi, E.I., Zhang, Q., Xiao, X., Kaiser-Kupfer, M., Sergeev, Y.V. and Hejtmancik, J.F. (2004) Bietti crystalline corneoretinal dystrophy is caused

by mutations in the novel gene *CYP4V2*. *American Journal of Human Genetics*; 74(5), 817-826.

Li, H., Pinot, F., Sauveplane, V., Werck-Reichhart, D., Diehl, P., Schreiber, L., Franke, R., Zhang, P., Chen, L., Gao, Y., Liang, W. and Zhang, D. (2010) Cytochrome P450 family member CYP704B2 catalyzes the ω-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. *The Plant Cell*; 22(1), 173-190.

Li, Y.C. and Chiang, J.Y. (1991) The expression of a catalytically active cholesterol 7α-hydroxylase cytochrome P450 in *Escherichia coli*. *The Journal of Biological Chemistry*; 266(29), 19186-19191.

Li-Beisson, Y., Pollard, M., Sauveplane, V., Pinot, F., Ohlrogge, J. and Beisson F. (2009) Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *PNAS*; 106(51), 22008-22013.

Lim, Y.-R., Eun, C.-Y., Park, H.-G., Han, S., Han, J.-S., Cho, K.S., Chun, Y.J. and Kim, D. (2010) Regioselective oxidation of lauric acid by CYP119, an orphan cytochrome P450 from *Sulfolobus acidocaldarius*. *Journal of Microbiology and Biotechnology*; 20(3), 574-578.

Lima, E.S. and Abdalla, D.S.P. (2002) High-performance liquid chromatography of fatty acids in biological samples. *Analytica Chimica Acta*; 465(1-2), 81-91.

Liu, J. and Rost, B. (2003) Domains, motifs and clusters in the protein universe. *Current Opinion in Chemical Biology*; 7, 5-11.

Liu, S., Li, C., Xie, L. & Cao, Z. (2003) Intracellular pH and metabolic activity of long-chain dicarboxylic acid-producing yeast *Candida tropicalis*. *Journal of Bioscience and Bioengineering;* 96(4), 349-353.

Locuson, C.W., Hutzler, J.M. and Tracy, T.S. (2007) Visible spectra of type II cytochrome P450-drug complexes: evidence that "incomplete" heme coordination is common. *Drug Metabolism and Disposition*; 35(4), 614-622.

Looman, A.C., Bodlaender, J., Comstock, L.J., Eaton, D., Jhurani, P., de Boer, H.A. and van Knippenberg, P.H. (1987) Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. *The EMBO Journal*; 6(8), 2489-2492.

López-Martínez, R. (2010) Candidosis, a new challenge. *Clinics in Dermatology*; 28(2), 178-184.

Lottermoser, K., Schunck, W.-H. and Asperger, O. (1996) Cytochromes P450 of the sophorose lipid-producing yeast *Candida apicola*: heterogeneity and polymerase chain reaction-mediated cloning of two genes. *Yeast*; 12(6), 565-575.

Lu, W., Ness, J.E., Xie, W., Zhang, X., Minshull, J. and Gross, R.A. (2010) Biosynthesis of monomers for plastics from renewable oils. *Journal of the American Chemical Society*; 132(43), 15451-15455.

Mao, W., Berenbaum, M.R. and Schuler, M.A. (2008) Modifications in the N-terminus of an insect cytochrome P450 enhance production of catalytically active protein in baculovirus-Sf9 cell expression systems. *Insect Biochemistry and Molecular Biology*; 38(1), 66-75.

Mast, N., Andersson, U., Nakayama, K., Bjorkhem, I. and Pikuleva, I.A. (2004) Expression of human cytochrome P450 46A1 in *Escherichia coli*: effects of N- and C-terminal modifications. *Archives of Biochemistry and Biophysics*; 428(1), 99-108.

Mast, N., Graham, S.E., Andersson, U., Bjorkhem, I., Hill, C., Peterson, J. and Pikuleva, I.A. (2005) Cholesterol binding to cytochrome P450 7A1, a key enzyme in bile acid biosynthesis. *Biochemistry*; 44(9), 3259-3271.

Matsui, K., Shibutani, M., Hase, T., and Kajiwara, T. (1996) Bell pepper fruit fatty acid hydroperoxide lyase is a cytochrome P450 (CYP74B). *FEBS Letters*; 394(1), 21-24.

Matsunaga, I., Yokotani, N., Gotoh, O., Kusunose, E., Yamada, M. and Ichihara, K. (1997) Molecular cloning and expression of fatty acid α-hydroxylase from *Sphingomonas paucimobilis. The Journal of Biological Chemistry*; 272(38), 23592-23596.

McLean, K.J., Marshall, K.R., Richmond, A., Hunter, I.S., Fowler, K., Kieser, T., Gurcha, S., Besra, G.S. and Munro, A.W. (2002) Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in Mycobacteria and Streptomyces. *Microbiology*; 148(10), 2937-2949.

McManus, M.E., Burgess, W.M., Veronese, M.E., Huggett, A., Quattrochi, L.C. and Tukey, R.H. (1990) Metabolism of 2-acetyl-aminofluorene and benzo(a)pyrene and activation of food-derived heterocyclic amine mutagens by human cytochromes P450. *Cancer Research*; 50(11), 3367-3376.

Meharenna, Y.T., Li, H., Hawkes, D.B., Pearson, A.G., De Voss, J. and Poulos, T.L. (2004) Crystal structure of P450cin in a complex with its substrate, 1,8-cineole, a close structural homologue to D-camphor, the substrate for P450cam. *Biochemistry*; 43(29), 9847-9494.

Mellor, J., Dobson, M.J., Roberts, N.A., Tuite, M.F., Emtage, J.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J. and Kingsman, S.M. (1983) Efficient synthesis of enzymatically active calf chymosin in *Saccharomyces cerevisiae. Gene*; 24(1), 1-14.

Melo, N.R., Moran, G.P., Warrilow, A.G.S., Dudley, E., Smith, S.N., Sullivan, D.J., Lamb, D.C., Kelly, D.E., Coleman, D.C. and Kelly, S.L. (2008) CYP56 (Dit2p) in *Candida albicans*: characterization and investigation of its role in growth and antifungal drug susceptibility. *Antimicrobial Agents and Chemotherapy*; 52(10), 3718-3724.

Menzel, R., Kärgel, E., Vogel, F., Böttcher, C. and Schunck, W.-H. (1996) Topogenesis of a microsomal cytochrome P450 and induction of endoplasmic reticulum membrane proliferation in *Saccharomyces cerevisiae. Archives of Biochemistry and Biophysics*; 330(1), 97-109.

Mi, H., Guo, N., Kejariwal, A. and Thomas, P.D. (2007) PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Research*; 35, D247-D252.

Miller, J.H., Ippen, K., Scaife, J.G. and Beckwith, J.R. (1968) The promoter-operator region of the lac operon of *Escherichia coli. Journal of Molecular Biology*; 38(3), 413-420.

Miura, Y. and Fulco, A.J. (1974) (ω-2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium. The Journal of Biological Chemistry*; 249, 1880-1888.

Miura, Y. and Fulco, A.J. (1975) ω-1, ω-2 and ω-3 hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from *Bacillus megaterium. Biochimica et Biophysica Acta*; 388(3), 305-317.

Miura, Y. (1981) ω- and (ω-1)-hydroxylation of 1-dodecanol by frog liver microsomes. *Lipids*; 16(10), 721-725.

Morant, M., Jørgensen, K., Schaller, H., Pinto, F., Møller, B.L., Werck-Reichhart, D. and Bak, S. (2007) CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *The Plant Cell*; 19(5), 1473-1487.

Morimoto, M., Hagbjork, A.L., Nanji, A.A., Ingelman-Sundberg, M., Lindros, K.O., Fu, P.C., Albano, E. and French, S.W. (1993) Role of cytochrome P4502E1 in alcoholic liver disease pathogenesis. *Alcohol*; 10(6), 459-464.

Morio, F., Loge, C., Besse, B., Hennequin, C. and Le Pape, P. (2010) Screening for amino acid substitutions in the *Candida albicans* Erg11 protein of azole-susceptible and azole-resistant clinical isolates: new substitutions and a review of the literature. *Diagnostic Microbiology and Infectious Disease*; 66(4), 373-384.

Moura, G.R., Paredes, J.A. and Santos, M.A.S. (2010) Development of the genetic code: insights from a fungal codon reassignment. *FEBS Letters*; 584(2), 334-341.

Müller-Hill, B. (1996) *The* lac *Operon: A short history of a genetic paradigm.* Walter de Gruyter & Co.

Murrieta, C.M., Hess, B.W. and Rule, D.C. (2003) Comparison of acidic and alkaline catalysts for preparation of fatty acid methyl esters from ovine muscle with emphasis on conjugated linoleic acid. *Meat Science*; 65(1), 523-529.

Nakahara, K. and Shoun, H. (1996) N-terminal processing and amino acid sequence of two isoforms of nitric oxide reductase cytochrome P450nor from *Fusarium oxysporum. Journal of Biochemistry (Tokyo)*; 120(6), 1082-1087.

Nakamura, M., Lin, J., Nishiguchi, K., Kondo, M., Sugita, J. and Miyake, Y. (2006) Bietti crystalline corneoretinal dystrophy associated with CYP4V2 gene mutations. *Advances in Experimental Medicine and Biology*; 572, 49-53.

Nakano, M., Kelly, E.J. and Rettie, A.E. (2009) Expression and characterization of CYP4V2 as a fatty acid ω-hydroxylase. *Drug Metabolism and Disposition*; 37(11), 2119-2122.

Nakayama, N., Takemae, A. and Shoun, H. (1996) Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum. Journal of Biochemistry (Tokyo)*; 119(3), 435-440.

National Sequence Databse (EMBL) www.ebi.ac.uk/embl

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

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. and Kemper, B. (1989) The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *DNA*; 8(1), 1-13.

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. *et al.* (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), 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., Gunsalus, I.C. and Nebert, D.W. (1996) P450 superfamily: update on new sequences,

gene mapping, accession numbers and nomenclature. *Pharmacogenetics*; 6(1), 1-42.

Nelson, D.R. (1998) Cytochrome P450 nomenclature. *Methods in Molecular Biology*; 107, 15-24.

Nelson, D.R. (2006) Cytochrome P450 nomenclature, 2004. *Methods in Molecular Biology*; 320, 1-10.

Nhamburo, P.T., Gonzalez, F.J., McBride, O.W., Gelboin, H.V. and Kimura, S. (1989) Identification of a new P450 expressed in human lung: complete cDNA sequence, cDNA-directed expression, and chromosome mapping. *Biochemistry*; 28(20), 8060-8066.

Nikolskaya, A.N., Arighi, C.N., Huang, H., Barker, W.C. and Wu, C.H. (2007) PIRSF family classification system for protein functional and evolutionary analysis. *Evolutionary Bioinformatics Online*; 2, 197-209.

Noble, M.A., Miles, C.S., Chapman, S.K., Lysek, D.A., MacKay, A.C., Reid, G.A., Hanzlik, R.P. and Munro, A.W. (1999) Roles of key active-site residues in flavocytochrome P450 BM3. *The Biochemical Journal*; 339(2), 371-379.

Nuñez, A., Ashby, R., Foglia, T.A. and Solaiman, D.K.Y. (2001) Analysis and characterization of sophorolipids by liquid chromatography with atmospheric pressure chemical ionization. *Chromatographia*; 53(11-12), 673-677.

Ohama, T., Suzuki, T., Mori, M., Osawa, S., Ueda, T., Watanabe, K. and Nakase, T. (1993) Non-universal decoding of the leucine codon CUG in several Candida species. *Nucleic Acids Research*; 21(17), 4039-4045.

Ohkuma, M., Hikiji, T., Tanimoto, T., Schunck, W.-H., Müller, H.G., Yano, K. and Takagi, M. (1991) Evidence that more than one gene encodes n-alkane-inducible cytochrome P-450s in *Candida maltosa*, found by two-step gene disruption. *Agricultural Biological Chemistry*; 55(7), 1757-1764.

Ohkuma, M., Zimmer, T., Iida, T., Schunck, W.-H., Ohta, A. and Takagi, M. (1998) Isozyme function of *n*-alkane-inducible cytochromes P450 in *Candida maltosa* revealed by sequential gene disruption. *The Journal of Biological Chemistry*; 273(7), 3948-3953.

Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *The Journal of Biological Chemistry*; 239, 2370-8.

Osawa, S. and Jukes, T.H. (1989) Codon reassignment (codon capture) in evolution. *Journal of Molecular Evolution*; 28(4), 271-278.

Ouellet, H., Guan, S., Johnston, J.B., Chow, E.D., Kells, P.M., Burlingame, A.L., Cox, J.S., Podust, L.M. and Ortiz de Montellano, P.R. (2010) *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Molecular Microbiology*; 77(3), 730-742.

Page, R.D.M. (2001) http://taxonomy.zoology.gla.ac.uk/rod/rod.html

Paine, M.J.I., Scrutton, N.S., Munro, A.W., Gutierrez, A., Roberts, G.C.K. and Wolf, C.R. (2005) Electron transfer partners of cytochrome P450. *Cytochrome P450: Structure, Mechanism and Biochemistry.* Ortiz De Montellano, P. R. (ed.), Third Edition, Kluwer Academic/Plenum Publishers, New York.

Pandey, A.K. and Sassetti, C.M. (2008) Mycobacterial persistence requires the utilization of host cholesterol. *PNAS*; 105(11), 4376-4380.

PANTHER http://www.pantherdb.org/

Panwar, S. L., Krishnamurthy, S., Gupta, V., Alarco, A.M., Raymond, M., Sanglard, D and Prasad, R. (2001) CaALK8, an alkane assimilating cytochrome P450,

confers multidrug resistance when expressed in a hypersensitive strain of *Candida albicans*. *Yeast*; 18(12), 1117-1129.

Park, H.G., Lim, Y.R., Eun, C.Y., Han, S., Han, J.S., Cho, K.S., Chun, Y.J. and Kim, D. (2010) *Candida albicans* NADPH-P450 reductase: expression, purification, and characterization of recombinant protein. *Biochemical and Biophysical Research Communications*; 396(2), 534-538.

Pernecky, S.J., Larson, J.R., Philpot, R.M. and Coon, M.J. (1993) Expression of truncated forms of liver microsomal P450 cytochromes 2B4 and 2E1 in *Escherichia coli*: influence of NH_2-terminal region on localization in cytosol and membranes. *PNAS*; 90(7), 2651-2655.

Pesole, G., Lotti, M., Alberghina, L. and Saccone, C. (1995) Evolutionary origin of nonuniversal CUG^{Ser} codon in some Candida species as inferred from a molecular phylogeny. *Genetics*; 141(3), 903-907.

Petkova-Andonova, M., Imaishi, H. and Ohkawa, H. (2002) CYP92B1, a cytochrome P450, expressed in petunia flower buds, that catalyzes monooxidation of long-chain fatty acids. *Bioscience, Biotechnology and Biochemistry*; 66(9), 1819-1828.

Peterson, J. A. and Graham, S. E. (1998) A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure*; 6(9), 1079-1085.

Pfam http://pfam.sanger.ac.uk/

Picataggio, S., Rohrer, T., Deanda, K., Lanning, D. Reynolds, R., Mielenz, J. and Eirich, L.D. (1992) Metabolic engineering of *Candida tropicalis* for the production of long-chain dicarboxylic acids. *Bio/Technology*; 10(8), 894-898.

Pierce, A.E. (1968) *Silylation of Organic Compounds: A Technology for Gas Phase Anaylsis.* Pierce Chemical Co., Rockford, Illinois.

Pinto, F. and Beisson, F. (2011) Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *The FEBS Journal*; 278(2), 195-205.

PIRFS http://pir.georgetown.edu/pirwww/dbinfo/pirsf.shtml

Pollard, M., Beisson, F., Li, Y. and Ohlrogge, J.B. (2008) Building lipid barriers: biosynthesis of cutin and suberin. *TRENDS in Plant Science*; 13(5), 236-246.

Pompon, D., Louerat, B., Bronine, A. and Urban, P. (1996) Yeast expression of animal and plant P450s in optimized redox environments. *Methods in Enzymology*; 272, 51-64.

Porter, T.D. and Coon, M.J. (1991) Cytochrome P450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *The Journal of Biological Chemistry*; 266(21), 13469-13472.

Poulos, T. L. and Johnson, E. F. (2005) Structures of cytochrome P450 enzymes. *Cytochrome P450: Structure, Mechanism and Biochemistry.* Ortiz De Montellano, P. R. (ed.), Third Edition, Kluwer Academic/Plenum Publishers, New York.

Prigneau, O., Porta, A., Poudrier, J.A., Colonna-Romano, S., Noel, T. and Maresca, B. (2003) Genes involved in β-oxidation, energy metabolism and glyoxylate cycle are induced by *Candida albicans* during macrophage infection. *Yeast*; 20(8), 723-730.

PRINTS http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/

ProDom http://www.toulouse.inra.fr/prodom.html

PROSITE www.expasy.ch/prosite/

Protein Information Resource http://pir.georgetown.edu/

Puchkaev, A.V. and Ortiz de Montellano, P.R. (2005) The *Sulfolobus solfataricus* electron donor partners of thermophilic CYP119: an unusual non-NAD(P)H-dependent cytochrome P450 system. *Archives of Biochemistry and Biophysics*; 434(1), 169-177.

Ratledge, C. (1984) Microbial conversions of alkane and fatty acids. *Journal of the American Oil Chemists' Society*; 61(2), 447-453.

Reed, J.R., Kelley, R.W. and Backes, W.L. (2006) An evaluation of methods for the reconstitution of cytochromes P450 and NADPH P450 reductase into lipid vesicles. *Drug Metabolism and Disposition*; 34(4), 660-666.

Rehm, B.H.A. (2001) Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. *Applied Microbiology and Biotechnology*; 57(5-6), 579-592.

Rehm, H.J. and Reiff, I. (1981) Mechanisms and occurrence of microbial oxidation of long-chain alkanes. *Reactors and Reactions*; 19, 175-215.

Richardson, T.H., Hsu, M.-H., Kronbach, T., Barnes, H.J., Chan, G., Waterman, M.R., Kemper, B. and Johnson, E.F. (1993) Purification and characterization of recombinant-expressed cytochrome P450 2C3 from *Escherichia coli*: 2C3 encodes the 6β-hydroxylase deficient form of P450 3b. *Archives of Biochemistry and Biophysics*; 300(1), 510-516.

Richardson, T.H., Jung, F., Griffin, K.J., Wester, M., Raucy, J.L., Kemper, B., Bornheim, L.M., Hassett, C., Omiecinski, C.J. and Johnson, E.F. (1995) A universal approach to the expression of human and rabbit cytochrome P450s of the 2C subfamily in *Escherichia coli*. *Archives of Biochemistry and Biophysics*; 323(1), 87-96.

Roberts, G.A., Grogan, G., Greter, A., Flitsch, S.L. and Turner, N.J. (2002) Identification of a new class of cytochrome P450 from Rhodococcus sp. *Journal of Bacteriology*; 184(14), 3898-3908.

Roberts, G.A., Celik, A., Hunter, D.J., Ost, T.W., White, J.H., Chapman, S.K., Turner, N.J. and Flitsch, S.L. (2003) A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *The Journal of Biological Chemistry*; 278(49), 48914-48920.

Romanos, M.A., Scorer, C.A. and Clare, J.J. (1992) Foreign gene expression in yeast: a review. *Yeast*; 8(6), 423-488.

Rose, M.D. and Broach, J.R. (1991) Cloning genes by complementation in yeast. *Methods in Enzymology*; 194, 195-230.

Rosłoniec, K.Z., Wilbrink, M.H., Capyk, J.K., Mohn, W.W., Ostendorf, M., van der Geize, R., Dijkhuizen, L. and Eltis, L.D. (2009) Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus joshii* RHA1. *Molecular Microbiology*; 74(5), 1031-1043.

Sacchromyces Genome Database http://yeastgenome.org

Sagara, Y., Barnes, H.J. and Waterman, M.R. (1993) Expression in *Escherichia coli* of functional cytochrome $P450_{c17}$ lacking its hydrophobic amino-terminal signal anchor. *Archives of Biochemistry and Biophysics*; 304(1), 272-278.

Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual.* Third Edition, Cold Spring Harbor Laboratory Press, New York.

Sandhu, P., Guo, Z., Baba, T., Martin, M.V., Tukey, R.H. and Guengerich, F.P. (1994) Expression of modified human cytochrome P450 1A2 in *Escherichia coli*: stabilization, purification, spectral characterization, and catalytic

activities of the enzyme. *Archives of Biochemistry and Biophysics*; 309(1), 168-177.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchinson, C.A., Slocombe, P.M. and Smith, M. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*; 265(5596), 687-695.

Sanglard, D., Chen, C. and Loper, J.C. (1987) Isolation of the alkane inducible cytochrome P450 (P450alk) gene from the yeast *Candida tropicalis*. *Biochemical and Biophysical Research Communications*; 144(1), 251-257.

Sanglard, D. and Loper, J.C. (1989) Characterization of the alkane-inducible cytochrome P450 (P450alk) gene from the yeast *Candida tropicalis*: identification of a new P450 gene family. *Gene*; 76(1), 121-136.

Santos, M.A.S. and Tuite, M.F. (1995) The CUG codon is decoded *in vivo* as serine and not leucine in *Candida albicans*. *Nucleic Acids Research*; 23(9), 1481-1486.

Sauveplane, V., Kandel, S., Kastner, P.-E., Ehlting, J., Compagnon, V., Werck-Reichhart, D. and Pinot, F. (2009) *Arabidopsis thaliana* CYP77A4 is the first cytochrome P450 able to catalyze the epoxidation of free fatty acids in plants. *The FEBS Journal*; 276(3), 719-735.

Scheller, U., Kraft, R., Schröder, K.-L. and Schunck, W.-H. (1994) Generation of the soluble and functional cytosolic domain of microsomal cytochrome P450 52A3. *The Journal of Biological Chemistry*; 269(17), 12779-12783.

Scheller, U., Zimmer, T., Kärgel, E. and Schunck, W.-H. (1996) Characterization of the *n*-alkane and fatty acid hydroxylating cytochrome P450 forms 52A3 and 52A4. *Archives of Biochemistry and Biophysics*; 328(2), 245-254.

Scheller, U., Zimmer, T., Becher, D., Scahuer, F. and Schunck, W.-H. (1998) Oxygenation cascade in conversion of n-alkanes to α,ω-dioic acids catalyzed by cytochrome P450 52A3. *The Journal of Biological Chemistry*; 273(49), 32528-32534.

Schenkman, J.B. (1970) Studies on the nature of the type I and type II spectral changes in liver microsomes. *Biochemistry*; 9(10), 2081-2091.

Schenkman, J.B., Ciniti, D.L., Orrenius, S., Moldeus, P. and Kraschnitz, R. (1972) The nature of the reverse type I (modified type II) spectral changes in liver microsomes. *Biochemistry*; 11(23), 4243-4251.

Schneider, S., Wubbolts, M.G., Sanglard, D. and Witholt, B. (1998) Biocatalyst engineering by assembly of fatty acid transport and oxidation activities for *in vivo* application of cytochrome P-450BM-3 monooxygenase. *Applied and Environmental Microbiology*; 64(10), 3784-3790.

Schuler, M.A. and Werck-Reichhart, D. (2003) Functional genomics of P450s. *Annual Review of Plant Biology*; 54, 629-67.

Schwarz, D., Kisselev, P., Ericksen, S.S., Szklarz, G.D., Chernogolov, A., Honeck, H., Schunck, W.-H. and Roots, I. (2004) Arachidonic and eicosapentaenoic acid metabolism by human CYP1A1: highly steroselective formation of 17(*R*), 18(*S*)-epoxyeicosatetraenoic acid. *Biochemical Pharmacology*; 67(8), 1445-1457.

Scott, E.E., Spatzenegger, M. and Halpert, J.R. (2001) A truncation of 2B subfamily cytochromes P450 yields increased expression levels, increased solubility, and decreased aggregation while retaining function. *Archives of Biochemistry and Biophysics*; 395(1), 57-68.

Shah, V., Doncel, G.F., Seyoum, T., Eaton, K.M., Zalenskaya, I., Hagver, R., Azim, A. and Gross, R. (2005) Sophorolipids, microbial glycolipids with anti-

human immunodeficiency virus and sperm-immobilising activities. *Antimicrobial Agents and Chemotherapy*; 49(10), 4093-4100.

Sharp, P. M. and Li, W.-H. (1987) The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*; 15(3), 1281-1295.

Sheehan, D.J., Hitchcock, C.A. and Sibley, C.M. (1999) Current and emerging azole antifungal agents. *Clinical Microbiology Reviews*; 12(1), 40-79.

Shibata, Y., Matsui, K., Kajiwara, T. and Hatanaka, A. (1995) Fatty acid hydroperoxide lyase is a heme protein. *Biochemical and Biophysical Research Communications*; 207(1), 438-443.

Shyadehi, A.Z., Lamb, D.C., Kelly, S.L., Kelly, D.E., Schunck, W.-H., Wright, J.N., Corina, D. and Akhtar, M. (1996) The mechanism of the acyl-carbon bond cleavage reaction catalyzed by recombinant sterol 14α-demethylase of *Candida albicans* (other names: lanosterol 14α-demethylase, P-450_{14DM}, and CYP51). *The Journal of Biological Chemistry*; 271(21), 12445-12450.

Sigle, R.O., Titus, M.A., Harada, N. and Nelson, S.D. (1994) Baculovirus-mediated high-level expression of human placental aromatase (CYP19A1). *Biochemical and Biophysical Research Communications*; 201(2), 694-700.

SignalP - http://www.cbs.dtu.dk/services/SignalP/

Simpson, A.E. (1997) The cytochrome P450 4 (CYP4) family. *General Pharmacology*; 28(3), 351-359.

SMART http://smart.embl-heidelberg.de/

Smith, J., Safdar, N., Knasinski, V., Simmons, W., Bhavnani, S.M., Ambrose, P.G. and Andes, D. (2006) Voriconazole therapeutic drug monitoring. *Antimicrobial Agents and Chemotherapy*; 50(4), 1570-1572.

Song, W.-C., Funk, C.D. and Brash, A.R. (1993) Molecular cloning of an allene oxide synthase: a cytochrome P450 specialized for the metabolism of fatty acid hydroperoxides. *PNAS*; 90(18), 8519-8523.

Sorenson, H.P. and Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Eschericha coli. Microbial Cell Factories*; 4, 1-8.

Strijbis, K., van Roermund, C.W.T., Visser, W.F., Mol, E.C., van den Burg, J., MacCallum, D.M., Odds, F.C., Paramonova, E., Krom, B.P., Distel, B. (2008) Carnitine-dependent transport of acetyl coenzyme A in *Candida albicans* is essential for growth on nonfermentable carbon sources and contributes to biofilm formation. *Eukaryotic Cell*, 7(4), 610-618.

Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis, R.W. (1979) High frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *PNAS*; 76(3), 1035-1039.

Strushkevich, N., Usanov, S.A. and Park, H.-W. (2010) Structural basis of human CYP51 inhibition by antifungal azoles. *Journal of Molecular Biology*; 397(4), 1067-1078.

Stumpe, M., Kandzia, R., Göbel, C., Rosahl, S. and Feussner, I. (2001) A pathogen-inducible divinyl ether synthase (CYP74D) from elicitor-treated potato suspension cells. *FEBS Letters*; 507(3), 371-376.

Stumpe, M., Göbel, C., Demchenko, K., Hoffmann, M., Klösgen, R.B., Pawlowski, K. and Feussner, I. (2006) Identification of an allene oxide synthase (CYP74C) that leads to formation of α-ketol from 9-hydroperoxides of linoleic and linolenic acid in below-ground organs of potato. *The Plant Journal*; 47(6), 883-896.

Sudbery, P., Gow, N. and Berman, J. (2004) The distinct morphogenic states of *Candida albicans*. *TRENDS in Microbiology*; 12(7), 317-324.

Sugiyama, H., Ohkuma, M., Masuda, Y., Park, S.-M., Ohta, A. and Takagi, M. (1995) *In vivo* evidence for non-universal usage of the codon CUG in *Candida maltosa*. *Yeast*; 11(1), 43-52.

SUPERFAMILY http://supfam.cs.bris.ac.uk/SUPERFAMILY/

Swiss Institute of Bioinformatics http://www.isb-sib.ch

Tanaka, A., Yamamura, M., Kawamoto, S. and Fukui, S. (1977) Production of uricase by *Candida tropicalis* using n-alkane as a substrate. *Applied and Environmental Microbiology*; 34(4), 342-346.

Tang, Z., Salamanca-Pinzón, S.G., Wu, Z.-L., Xiao, Y. and Guengerich, F.P. (2010) Human cytochrome P450 4F11: heterologous expression in bacteria, purification and characterization of catalytic function. *Archives of Biochemistry and Biophysics*; 494(1), 86-93.

Tartoff, K.D. and Hobbs, C.A. (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Research Labs Focus*; 9, 12.

Tehlivets, O., Scheuringer, K. and Kohlwein, S.D. (2007) Fatty acid synthesis and elongation in yeast. *Biochimica et Biophysica Acta*; 1771(3), 255-270.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*; 25(24), 4876-4882.

TIGRFAMs http://www.jcvi.org/cms/research/projects/tigrfams/overview/

Tijet, N., Helvig, C., Pinot, F., Le Bouquin, R., Lesot, A., Durst, F., Salaün, J.-P. and Benveniste, I. (1998) Functional expression in yeast and characterization of a clofibrate-inducible plant cytochrome P-450 (CYP94A1) involved in cutin monomers synthesis. *The Biochemical Journal*; 332(2), 583-589.

TMHMM - http://www.cbs.dtu.dk/services/TMHMM/

Traystman, M.D., Higuchi, M., Kasper, C.K., Antonarakis, S.E. and Kazazian, H.H. Jr. (1990) Use of denaturing gradient gel electrophoresis to detect point mutations in the factor VIII gene. *Genomics*; 6(2), 293-301.

Tschumper, G. and Carbon, J. (1980) Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP*1 gene. *Gene*; 10(2), 157-166.

UniProt www.uniprot.org

Uppuluri, P. and Chaffin, W.L. (2007) Defining *Candida albicans* stationary phase by cellular and DNA replication, gene expression and regulation. *Molecular Microbiology*; 64(6), 1572-1586.

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

Usuda, K., Toritsuka, N., Matsuo, Y., Kim, D.H. and Shoun H. (1995) Denitrification by the fungus *Cylindrocarpon tonkinense*: anaerobic cell growth and two isozyme forms of cytochrome P-450nor. *Applied and Environmental Microbiology*; 61(3), 883-889.

Vamecq, J., de Hoffmann, E. and van Hoof, F. (1985) The microsomal dicarboxylyl-CoA synthetase. *The Biochemical Journal*; 230(3), 683-693.

van Bogaert, I.N.A., Develter, D., Soetaert, W. and Vandamme, E.J. (2007) Cloning and characterization of the NADPH cytochrome P450 reductase gene (CPR) from *Candida bombicola*. *FEMS Yeast Research,* 7(6), 922-928.

van Bogaert, I.N.A., De Mey, M., Develter, D., Soetaert, W. and Vandamme, E.J. (2008) Importance of the cytochrome P450 monooxygenase CYP52 family

for the sophorolipid-producing yeast *Candida bombicola*. *FEMS Yeast Research*; 9(1), 87-94.

van Bogaert, I.N., Groeneboer, S., Saerens, K. and Soetaert, W. (2011) The role of cytochrome P450 monooxygenases in microbial fatty acid metabolism. *The FEBS Journal*; 278(2), 206-221.

van den Brink, H.J.M., van Gorcom, R.F.M., van den Hondel, C.A.M.J.J. and Punt, P.J. (1998) Cytochrome P450 enzyme systems in fungi. *Fungal Genetics and Biology*; 23(1), 1-17.

Vanden Bossche, H., Koymans, L. and Moereels, H. (1995) P450 inhibitors of use in medical treatment: focus on mechanisms of action. *Pharmacology and Therapeutics*; 67(1), 79-100.

van der Klei, I.J. and Veenhuis, M. (1997) Yeast peroxisomes: function and biogenesis of a versatile cell organelle. *TRENDS in Microbiology*; 5(12), 502-509.

van Roermund, C.W.T., Elgersma, Y., Singh, N., Wanders, R.J.A. and Tabak, H.F. (1995) The membrane of peroxisomes in *Saccharomyces cerevisiae* is impermeable to NAD(H) and acetyl-CoA under *in vivo* conditions. *The EMBO Journal*; 14(14), 3480-3486.

van Roermund, C.W.T., Hettema, E.H., Kal, A.J., van den Berg, M., Tabak, H.F. and Wanders, R.J.A. (1998) Peroxisomal β-oxidation of polyunsaturated fatty acids in *Saccharomyces cerevisiae*: isocitrate dehydrogenase provides NADPH for reduction of double bonds at even positions. *The EMBO Journal*; 17(3), 677-687.

van Roermund, C.W.T., Visser, W.F., Ijlst, L., van Cruchten, A., Boek, M., Kulik, W., Waterham, H.R. and Wanders, R.J.A. (2008) The human peroxisomal ABC half transporter ALDP functions as a homodimer and accepts acyl-CoA esters. *The FASEB Journal*; 22(12), 4201-4208.

van Roermund, C.W.T., Visser, W.F., Ijlst, L., Waterham, H.R. and Wanders, R.J.A. (2011) Differential substrate specificities of human ABCD1 and ABCD2 in peroxisomal fatty acid β-oxidation. *Biochimica et Biophysica Acta*; 1811(3), 148-152.

Ventura, S. and Villaverde, A. (2006) Protein quality in bacterial inclusion bodies. *TRENDS in Biotechnology*; 24(4), 179-185.

Verma, R., Boletti, E. and George, A.J.T. (1998) Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological Methods*; 216(1-2), 165-181.

Vermilion, J.L. and Coon, M.J. (1978) Purified liver microsomal NADPH-cytchrome P450 reductase. *The Journal of Biological Chemistry*; 253, 2694-2704.

Vogelstein, B. and Gillepsie, D. (1979) Preparative and analytical purification of DNA from agarose. *PNAS*; 76(2), 615-619.

von Wachenfeldt, C., Richardson, T.H., Cosme, J. and Johnson, E.F. (1997) Microsomal P450 2C3 is expressed as a soluble dimer in *Escherichia coli* following modifications of its N-terminus. *Archives of Biochemistry and Biophysics*; 339(1), 107-114.

Wada, Y., Itabashi, T., Sato, H., Kawamura, M., Tada, A. and Tamai, M. (2005) Screening for mutations in CYP4V2 gene in Japanese patients with Bietti's crystalline corneoretinal dystrophy. *American Journal of Ophthalmology*; 139(5), 894-899.

Wanders, R.J.A., Visser, W.F., van Roermund, C.W.T., Kemp, S. and Waterham, H.R. (2007) The peroxisomal ABC transporter family. *Pflugers Archiv: European Journal of Physiology*; 453(5), 719-734.

Wang, M.-H., Stec, D.E., Balazy, M., Mastyugin, V., Yang, C.S., Roman, R.J. and Schwartzman, M.L. (1996) Cloning, sequencing and cDNA-directed expression of the rat renal CYP4A2: arachidonic acid ω-hydroxylation and 11,12-epoxidation by CYP4A2 protein. *Archives of Biochemistry and Biophysics*; 336(2), 240-250.

Wang, R.W., Prasad, K.H., Lu, A.Y.H., Thomas, P.E., Guengerich, F.P. and Vyas, K.P. (1991) Biotransformation of lovastatin. *Archives of Biochemistry and Biophysics*; 290(2), 355-361.

Warman, A.J., Roitel, O., Neeli, R., Girvan, H.M., Seward, H.E., Murray, S.A., McLean, K.J., Joyce, M.G., Toogood, H., Holt, R.A., Leys, D., Scrutton, N.S. and Munro, A.W. (2005) Flavocytochrome P450 BM3: an update on structure and mechanism of a biotechnologically important enzyme. *Biochemical Society Transactions*; 33(4), 747-753.

Warrilow, A., Ugochukwu, C., Lamb, D., Kelly, D. and Kelly, S.L. (2008) Expression and characterization of CYP51, the ancient sterol 14-demethylase activity for cytochromes P450 (CYP), in the white-rot fungus *Phanerochaete chrysosporium*. *Lipids*; 43(12), 1143-1153.

Waterman, M.R., Jenkins, C.M. and Pikuleva, I. (1995) Genetically engineered bacterial cells and applications. *Toxicology Letters*; 82/83, 807-813.

Watkins, P.B., Wrighton, S.A., Maurel, P., Schuetz, E.G., Mendez-Picon, G., Parker, G.A. and Guzelian, P.S. (1985) Identification of an inducible form of cytochrome P-450 in human liver. *PNAS*; 82(18), 6310-6314.

Waxman, D.J. (1991) P450-catalysed steroid hydroxylation: assay and product identification by thin-layer chromatography. *Methods in Enzymology*; 206, 462-476.

Wellesen, K., Durst, F., Pinot, F., Benveniste, I., Nettesheim, K., Wisman, E., Steiner-Lange, S., Saedler, H. and Yephremov, A. (2001) Functional analysis of the *LACERATA* gene of *Arabidopsis* provides evidence for different roles of fatty acid ω-hydroxylation in development. *PNAS*; 98(17), 9694-9699.

Wentzel, A., Ellingsen, T.E., Kotlar, H.-K., Zotchev, S.B. and Throne-Holst, M. (2007) Bacterial metabolism of long-chain *n*-alkanes. *Applied Microbiology and Biotechnology*; 76(6), 1209-1221.

Werck-Reichhart, D. and Feyereisen, R. (2000) Cytochromes P450: a success story. *Genome Biology*; 1(6), 3003.1-3003.9.

Wester, M. R., Stout, C.D. and Johnson, E.F. (2002). Purification and crystallization of N-terminally truncated forms of microsomal cytochrome P450 2C5. *Methods in Enzymology*; 357, 73-79.

Westhead, D.R., Parish, J.H. and Twyman, R.M. (2002) *Instant Notes: Bioinformatics.* Taylor & Francis.

Whiteway, M. and Bachewich, C. (2007) Morphogenesis in *Candida albicans*. *Annual Review of Microbiology*; 61, 529-553.

Williams, D.C., van Frank, R.M., Munt, W.L. and Burnett, J.P. (1982) Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science*; 215(4533), 687-689.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F. and McRee, D.E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural

adaptations for membrane binding and functional diversity. *Molecular Cell*; 5(1), 121-131.

Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2007) The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Research*; 35, D308-D313.

Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. and Osawa, S. (1985) UGA is read as tryptophan in *Mycoplasma capricolum*. *PNAS*; 82(8), 2306-2309.

Yamano, S., Tatsuno, J. and Gonzalez, F.J. (1990) The CYP2A3 gene product catalyzes coumarin 7-hydroxylation in human liver microsomes. *Biochemistry*; 29(5), 1322-1329.

Yeats, C., Lees, J., Reid, A., Kellam, P., Martin, N., Liu, X. And Orengo, C. (2008) Gene3D: comprehensive structural and functional annotation of genomes. *Nucleic Acids Research*; 36, D414-D418.

Yokogawa, T., Suzuki, T., Ueda, T., Mori, M., Ohama, T., Kuchino, Y., Yoshinari, S., Motoki, I., Nishikawa, K., Osawa, S. and Watanabe, K. (1992) Serine tRNA complementary to the nonuniversal serine codon CUG in *Candida cylindracea*: evolutionary implications. *PNAS*; 89(16), 7408-7411.

Yoshida, Y. (1988) Cytochrome P450 of fungi: primary target for azole antifungal agents. *Current Topics in Medical Mycology*; 2, 388-418.

Yoshida, Y., Aoyama, Y., Noshiro, M. and Gotoh, O. (2000) Sterol 14-demethylase P450 (CYP51) provides a breakthrough of the discussion on the evolution of cytochrome P450 gene superfamily. *Biochemical and Biophysical Research Communications*; 273(3), 799-804.

Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*; 38(3), 95-109.

Zhang, L., Takaya, N., Kitazume, T., Kondo, T. and Shoun, H. (2001) Purification and cDNA cloning of nitric oxide reductase cytochrome P450nor (*CYP55A4*) from *Trichosporon cutaneum*. *European Journal of Biochemistry*; 268(11), 3198-3204.

Zhou, D., Linnenbach, A.J., Liu, R., Luzietti, R.A., Harris, J.J., Booth-Genthe, C.L. and Grimm, S.W. (2010) Expression and characterization of dog cytochrome P450 2A13 and 2A25 in baculovirus-infected insect cells. *Drug Metabolism and Disposition*; 38(7), 1015-1018.

Zhou, H. and Lorenz, M.C. (2008) Carnitine acetyltransferase are required for growth on non-fermentable carbon sources but not for pathogenesis in *Candida albicans*. *Microbiology*; 154(2), 500-509.

Zimmer, T., Kaminski, K., Scheller, U., Vogel, F. and Schunck, W.-H. (1995) *In vivo* reconstitution of highly active *Candida maltosa* cytochrome P450 monooxygenase systems in inducible membranes of *Saccharomyces cerevisiae*. *DNA and Cell Biology*; 14(7), 619-628.

Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. and Schunck, W.-H. (1996) The CYP52 multigene family of *Candida maltosa* encodes functionally diverse n-alkane-inducible cytochromes P450. *Biochemical and Biophysical Research Communications*; 224(3), 784-789.

Zubay, G., Morse, D.E., Schrenk, W.J. and Miller, J.H. (1972) Detection and isolation of the repressor protein for the tryptophan operon of *Escherichia coli*. *PNAS*; 69(5), 1100-1103.