CONSTRUCTION OF SELF-SUFFICIENT CYP153 CHIMERAS

BY

CHARLENE RANDALL

SUBMITTED IN ACCORDANCE WITH THE REQUIREMENTS FOR THE DEGREE

MAGISTER SCIENTIAE

IN THE

DEPARTMENT OF MICROBIAL, BIOCHEMICAL AND FOOD BIOTECHNOLOGY
FACULTY OF NATURAL AND AGRICULTURAL SCIENCES
UNIVERSITY OF THE FREE STATE
BLOEMFONTEIN 9300
SOUTH AFRICA

JANUARY 2010

SUPERVISOR: PROF. M.S. SMIT
CO-SUPERVISOR: PROF. J. ALBERTYN

“Imagination is more important than knowledge…”

Albert Einstein

“Anyone who has never made a mistake has never tried anything new.”

Albert Einstein

Acknowledgements

The financial assistance of the **National Research Foundation (NRF)** towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

I would like to express my gratitude towards the following people:

Prof. M.S. Smit for her enthusiasm and guidance and for always finding something positive in negative results.

Prof. J. Albertyn for his sense of humour and logical approach to research.

My father **Anthony**, my mother **Delene** and my sister **Margaux** for their encouragement and support when I needed it most.

Dr Nathlee Abbai for donating *CYP153A* fragments amplified during her Ph.D. research project.

Members of the **Biocatalysis Research Group** for their friendship, support and laughter.

Members of the **Molecular Biology Lab** for the conversions and fun times that made long days a little shorter.

My **friends** in the department for their support and conversations in the corridor.

Sarel Marais for his help with the GC analysis.

My Heavenly Father for His unchanging love. He makes all things possible.

Table of Contents

List of Abbreviations

°	Degrees
℃	Degrees Celsius
℃/min	Degrees Celsius per minute
x *g*	Times acceleration due to gravity
[2Fe-2S]	Iron-sulphur cluster
3'	Three-prime
5'	Five-prime
δ-ALA	5-Aminolevulinic acid hydrochloride
ε	Extinction coefficient
μg.mL^{-1}	Microgram(s) per millilitre
μL	Microlitre(s)
μm	Micrometre(s)
μM	Micromolar
A_{420}	Absorbance at 420 nanometres
A_{450}	Absorbance at 450 nanometres
A_{490}	Absorbance at 490 nanometres
AciA/PFOR(P450RhF)	AciA heme domain – P450RhF PFOR domain fusion
AdRed	Adrenodoxin reductase
Adx	Adrenodoxin
BLAST	Basic Local Alignment Search Tool
bp	Basepair(s)
BSA	Bovine serum albumin
CLERY	Combinatorial Libraries Enhanced by Recombination in Yeast
CO-difference	Carbon monoxide-difference
CPR	NADPH-cytochrome P450 reductase
CYP	Cytochrome P450
CYP153A6/PFOR(CYP116B3)	CYP153A6 heme domain – CYP116B3 PFOR domain fusion
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I

dNTPs	Deoxyribonucleoside triphosphates
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FdR	Ferredoxin reductase
Fdx	Ferredoxin
$FeCl_3$	Ferric chloride
$FeSO_4$	Ferrous sulphate
FID	Flame ionisation detector
FMN	Flavin mononucleotide
GC	Gas chromatography
H_2	Hydrogen
HCl	Hydrochloric acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobasepair(s)
k_{cat}	Turnover number
KCN	Potassium cyanide
kDa	Kilodalton(s)
kPa	Kilopascal(s)
kpsi	Kilopound(s) per square inch
l	Path length
LB	Luria-Bertani
m	Metre(s)
M	Molar
mg	Milligram(s)
Mg^{2+}	Magnesium ions
$MgCl_2$	Magnesium chloride
$mg.L^{-1}$	Milligram(s) per litre
min	Minute(s)
min^{-1}	Per minute
min/kb	Minute(s) per kilobasepair
mL	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
$mM^{-1}cm^{-1}$	Per millimolar per centimetre
Mn^{2+}	Manganese ions
NaCl	Sodium chloride

NADH	β-Nicotinamide adenine dinucleotide-reduced
NADPH	β-Nicotinamide adenine dinucleotide phosphate-reduced
NaOH	Sodium hydroxide
NCBI	National Centre for Bioinformatics
ng	Nanogram(s)
nm	Nanometre(s)
nmol.min^{-1}.mg protein^{-1}	Nanomole(s) per minute per milligram protein
OD_{600}	Optical density at 600 nanometres
ORF	Open-reading frame
P420	Pigment 420
P450	Pigment 450
$P450_{balk}$/PFOR(P450RhF)	$P450_{balk}$ heme domain – P450RhF PFOR domain fusion
$P450_{bzo}$/PFOR(P450RhF)	$P450_{bzo}$ heme domain – P450RhF PFOR domain fusion
$P450_{cam}$/PFOR(P450RhF)	$P450_{cam}$ heme domain – P450RhF PFOR domain fusion
PCR	Polymerase chain reaction
Pd	Putidaredoxin
PdR	Putidaredoxin reductase
PFOR	Phthalate family oxygenase reductase
pmol	Picomole(s)
pmol.mg protein^{-1}	Picomole(s) per milligram protein
PMSF	Phenylmethylsulphonyl fluoride
RNAs	Ribonucleic acids
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sec	Seconds
sec/kb	Second(s) per kilobasepair
SHIPREC	Sequence Homology-Independent Protein Recombination
SISDC	Sequence-Independent Site-Directed Chimeragenesis
sp.	Species
SRSs	Substrate recognition sites
TAE	Tris-Acetate-EDTA
T_m	Melting temperature
Tris	2-Amino-2-(hydroxymethyl)-1,3-propandiol

Tris-HCl	2-Amino-2-(hydroxymethyl)-1,3-propandiol, hydrochloric acid
tRNAs	Transfer RNAs
U	Units
V_{max}	Maximum velocity
v/v	Volume per volume
w/v	Weight per volume
w/w	Weight per weight
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1

Introduction to Present Study

1.1 Introduction to P450s

Cytochrome P450 monooxygenases (P450 or CYP) are a diverse superfamily of heme-containing enzymes that were first discovered in 1955 in rat liver microsomes (Klingenberg, 1958). Currently, there are more than 11 000 genes encoding P450 proteins that have been identified in archae, various bacteria and eukaryotes (http://drnelson.utmem.edu/). Two P450s have also recently been identified in the mimivirus (Lamb *et al.*, 2009).

P450 is the abbreviation for "Pigment 450" and this name is derived from the maximum absorbance peak (Soret band) at approximately 450 nm, which is exhibited by P450s in a reduced, carbon monoxide-bound form (Danielson, 2002). This spectroscopic property is the result of the thiolate group of an absolutely conserved cysteine residue which serves as the fifth ligand of the hexacoordinated active site heme iron (Sono *et al.*, 1996). This property is used to determine if a P450 is folded correctly, and therefore active, as only P450s which are catalytically active, having a correctly incorporated heme and therefore a correctly folded heme domain, exhibit the peak at 450 nm. P450s that have an incorrectly bound heme, and therefore a misfolded heme domain, exhibit a Soret band at 420 nm, with the enzyme usually being catalytically inactive. In some cases peaks occur at both 420 nm and 450 nm, where both functional and non-functional forms of the protein are present. This is referred to as low-level functional expression (Kubota *et al.*, 2005).

P450 enzymes are named and grouped into families and subfamilies according to amino acid identity. P450s are designated by the abbreviation "CYP" followed by a number indicating the family (more than 40% amino acid identity), a letter indicating the subfamily (usually more than 55% identity) and another number indicating a particular enzyme in the subfamily (Hannemann *et al.*, 2007). Proteins with more than 97% amino acid identity are variants of the same enzyme. P450 names are italicised when referring to the gene encoding that particular protein (e.g. *CYP153A6* refers to the gene encoding the CYP153A6 protein).

Although P450s from different families often share less than 20% amino acid identity, these enzymes have a conserved structural framework (Fig. 1.1) consisting of an amino-terminal half which is rich in beta-sheets and a carboxy-terminal half which is rich in alpha-helices (Danielson, 2002). The most variable regions of the P450 structure form part of the substrate recognition sites (SRSs) involved in the binding of substrates (Gotoh, 1992). This accounts for the ability of P450s, as a superfamily, to accept a broad range of substrates.

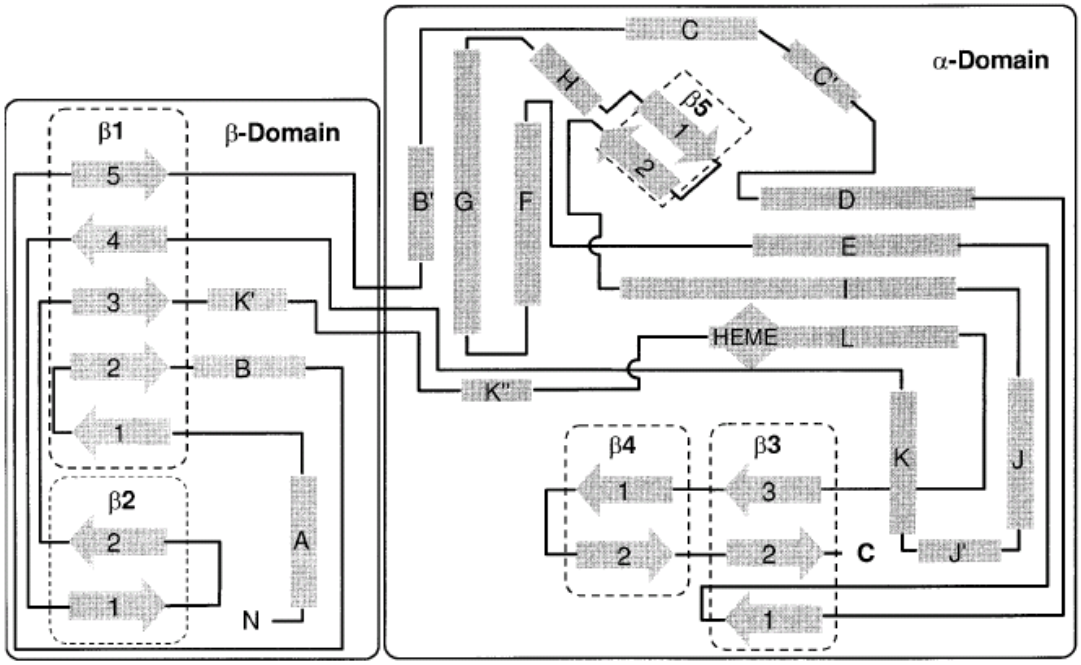


Figure 1.1 Topographic map of the secondary structure elements of the conserved P450 structural framework. The grey arrows represent beta-strands and the grey boxes represent alpha-helices. The β-domain consists mostly of beta-strands whereas the α-domain consists mostly of alpha-helices (Graham & Peterson, 1999).

P450s have a range of catalytic functions (Fig. 1.2) including the biosynthesis and metabolism of endogenous compounds (e.g. hormones, fatty acids and defensive compounds) and the metabolism of xenobiotics (e.g. polycyclic aromatic hydrocarbons and pesticides). Some of these xenobiotic compounds are converted to reactive intermediates which may react with biological compounds and lead to the formation of cancer (Sono *et al.*, 1996).

P450s are able to catalyse diverse reactions including hydroxylations, epoxidations, heteroatom oxidations and reductions in a stereo- and regioselective manner by the insertion of a single oxygen atom into an activated or unactivated bond (Graham & Peterson, 1999; Danielson,

2002). A typical hydroxylation reaction catalysed by P450s can be summarised with the following equation:

$$R - H + O_2 + NAD(P)H + H^+ \rightarrow R - OH + H_2O + NAD(P)^+$$

R – H represents an activated or unactivated bond in the substrate (Sono *et al.*, 1996). In the substrate-free form of the enzyme, a water molecule serves as the sixth ligand to the hexacoordinated heme iron. When a substrate molecule, R, binds to the active site, this water molecule is displaced and the iron becomes pentacoordinated. An electron is then transferred to the heme from a cofactor, NADH (β-Nicotinamide adenine dinucleotide-reduced) or NADPH (β-Nicotinamide adenine dinucleotide phosphate-reduced), which reduces the iron and results in the binding of a molecule of oxygen (Danielson, 2002). A second electron is transferred to the heme, two protons are taken up, and the bond between the two oxygen atoms is cleaved. One oxygen atom is released as a water molecule and the remaining oxygen atom is inserted into a bond in the substrate, R – H, resulting in the hydroxylated product, R – OH, which is released from the active site.

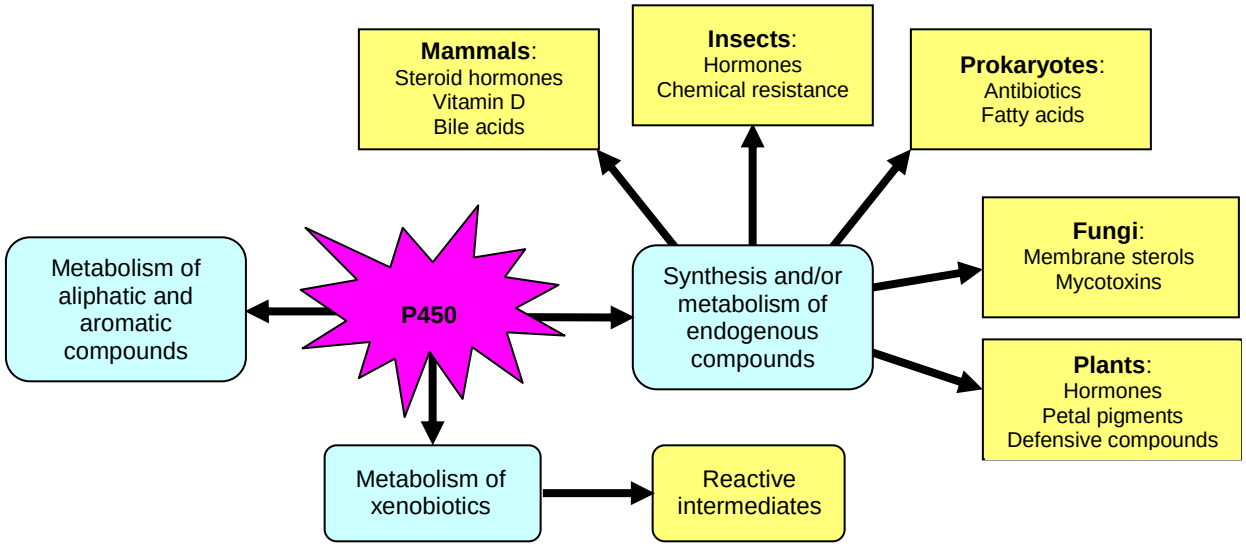


Figure 1.2 Schematic representation of the diverse functions of P450s in various organisms (Yuan *et al.*, 2005; Chefson & Auclair, 2006; Hannemann *et al.*, 2007).

All prokaryotic P450s are soluble, cytosolic enzymes whereas most eukaryotic P450s are bound to inner mitochondrial membranes (mitochondrial P450s) or to the membranes of the endoplasmic reticulum (microsomal P450s) (Danielson, 2002). Most P450s require one or more

redox partner proteins to transfer the electrons required for catalysis from the cofactor to the heme iron. Hannemann *et al.* (2007) grouped P450s into ten classes according to the redox partner proteins involved in electron transport. The eukaryotic mitochondrial P450 systems and most bacterial systems are grouped together in Class I (Fig 1.3(a) and (b)). These electron transport systems consist of two separate redox partner proteins: an FAD-containing ferredoxin reductase (FdR) and a ferredoxin (Fdx) containing an iron-sulphur cluster of the [2Fe-2S]-type. Electrons are transferred from the cofactor to the ferredoxin *via* the ferredoxin reductase; ferredoxin then transfers the electrons to the P450.

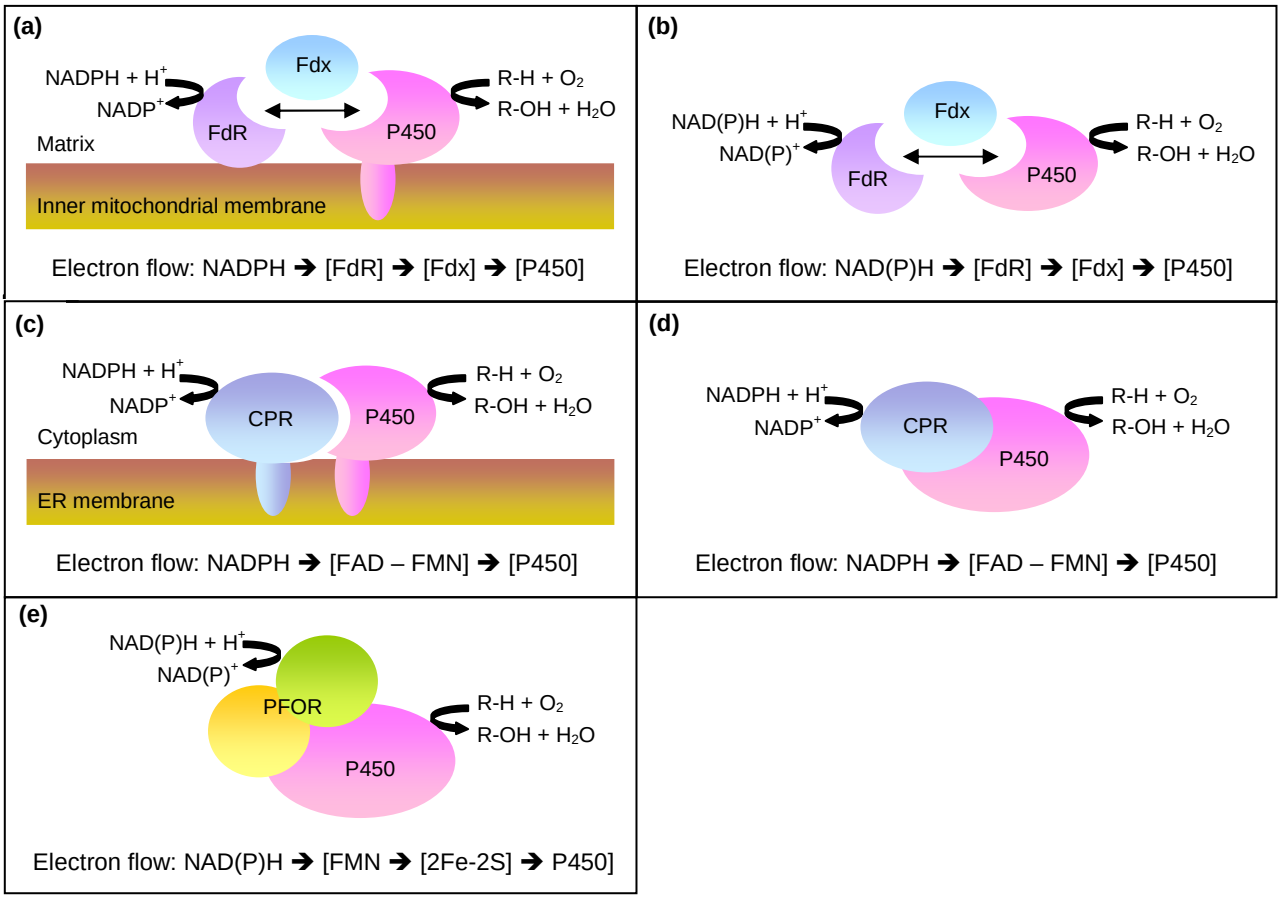


Figure 1.3 Schematic representations of P450 electron transport systems. **(a)** Eukaryotic mitochondrial Class I P450 system; **(b)** soluble bacterial Class I P450 system; **(c)** eukaryotic microsomal Class II P450 system; **(d)** Class VIII P450 system which is a soluble, fused version of the Class II system; **(e)** electron transport system of Class VII P450s (Adapted from Hannemann *et al.*, 2007).

The eukaryotic microsomal P450s (Fig 1.3(c)) and some bacterial P450s are grouped in Class II. These P450s have an NADPH-cytochrome P450 reductase (CPR) containing flavin adenine

dinucleotide (FAD) and flavin mononucleotide (FMN) prosthetic groups which transfer electrons from NADPH to the heme (Hannemann *et al.*, 2007).

An interesting electron transport system is that of P450s in Class VII and Class VIII. These are soluble fusion P450s, which means that the redox partners and the P450 domains are expressed together as a single protein, making these P450s catalytically self-sufficient as they do not require additional separate redox partners for electron transport (Hannemann *et al.*, 2007). Electron transport systems of P450s belonging to Class VIII consist of a CPR, similar to that of Class II systems, which is fused to the C-terminus of a soluble P450 domain (Fig 1.3(d)). Electrons from the cofactor are transferred to the P450 domain *via* the FAD and FMN prosthetic groups. $P450_{BM3}$ (CYP102A1) from *Bacillus megaterium*, the fastest known P450, is a member of this class.

The electron transport system of P450s belonging to Class VII (Fig 1.3(e)) is composed of a reductase domain of the phthalate family of mono- and dioxygenases (PFOR) fused to a Class I P450 domain (De Mot & Parret, 2002). The reductase domain contains an FMN-binding domain, an NAD(P)H-binding domain and a [2Fe-2S] ferredoxin domain (Correll *et al.*, 1992). Electrons are transferred to the P450 domain *via* the FMN and the iron-sulphur cluster. The first reported P450 belonging to this class is P450RhF (CYP116B2) from *Rhodococcus* sp. strain NCIMB 9784 (Roberts *et al.*, 2002). A number of P450RhF homologues have since been identified in *Burkholderia*, *Ralstonia* and *Gibberella* species, and in a strain of *Rhodococcus ruber* (Liu *et al.*, 2006; Hannemann *et al.*, 2007).

1.2 Aim of the Study

Many of the reactions catalysed by P450s are difficult or impossible to accomplish with synthetic organic chemistry, making these enzymes potentially useful in industry, particularly for the synthesis of chemicals. One such difficult reaction is the terminal hydroxylation of alkanes which has been of interest to a number of research groups, including ours. This reaction, which converts alkanes to 1-alkanols, is the first step in the degradation of alkanes and enables the host organism to utilise aliphatic alkanes as a sole carbon source (van Beilen *et al.*, 2006). Chemists have thus far been unable to develop chemical catalysts that can catalyse the terminal hydroxylation of alkane chains with the same exquisite regioselectivity as enzymes. There are only two extensively studied microbial P450 families that are able to catalyse this reaction: CYP52, which is found in yeast (Craft *et al.*, 2003) and CYP153, which is found in bacteria (van

Beilen *et al.*, 2006). One advantage that enzymes of the CYP153 family have over those of the CYP52 family is that they are less prone to over-oxidation of the alcohol products to carboxylic acids. For the purposes of this study, we focused on the CYP153 family.

Enzymes in the CYP153 family also catalyse, in addition to the terminal hydroxylation of alkanes, the terminal hydroxylation of alicyclic and alkyl-substituted substrates and the epoxidation of linear and cyclic compounds (Sieber *et al.*, 2001). One of the best-characterised enzymes in the CYP153 family is CYP153A6 from *Mycobacterium* sp. HXN-1500. This enzyme catalyses the hydroxylation of C6 to C11 alkanes with a regiospecificity of 95% for the terminal carbon position (Funhoff *et al.*, 2006) and is also able to convert limonene to perillyl alcohol, a compound used in the treatment of cancer (van Beilen *et al.*, 2005).

P450s in this family are Class I enzymes so their electron transport system consists of three separate, soluble proteins. This is not an optimal situation for the application of CYP153 enzymes in industry as the redox partners have to be cloned and expressed along with the P450. Furthermore, electron transfer *via* two redox partner proteins is not very efficient. A preferred situation would be a fusion arrangement similar to that of $P450_{BM3}$ or the CYP116B enzymes. With this arrangement, electron transfer occurs quicker than with the separate protein system and may result in an enhanced reaction velocity (Munro *et al.*, 2007). In addition, expression of the electron transport system proteins is relatively simple and the electron transport system is less complex because of the fact that these proteins are encoded by a single gene. However, no naturally self-sufficient CYP153 enzymes have been identified. Nodate and co-workers therefore constructed an artificial fusion between the reductase domain (PFOR) of P450RhF (CYP116B2) from *Rhodococcus* sp. NCIMB 9784 and three different Class I P450 domains, including $P450_{balk}$ (CYP153A13a) from *Alcanivorax borkumensis* SK2 (Nodate *et al.*, 2006). The resulting proteins were functional self-sufficient P450s. This raised the question whether fusion of other Class I P450 domains to the PFOR reductase domains of other Class VII P450s would also result in functional enzymes (De Mot & Parret, 2002). With this in mind, the first aim of this study was to construct a self-sufficient terminal alkane hydroxylase by fusing the P450 domain of CYP153A6 to the PFOR domain of CYP116B3 from *Rhodococcus ruber* DSM 44319, which shares 89% amino acid identity with the PFOR domain of *Rhodococcus* sp. NCIMB 9784.

There are also other properties of P450s, besides redox partners, that currently limit their application in industry. These include low stability, low activity and limited substrate specificity. The ultimate goal of our research is the directed evolution of P450s to modify some of these

properties, using the constructed self-sufficient P450 as a starting point. The first step of directed evolution involves the generation of genetic diversity and one approach which can be used to generate this genetic diversity is the creation of chimeras. Kubota *et al*. (2005) used cassette PCR to construct 8 new catalytically active chimeric self-sufficient CYP153As based on the $P450_{balk}$/PFOR(P450RhF) fusion described above. The second aim of this study was to follow the same approach to construct chimeras based on the envisaged CYP153A6/PFOR(CYP116B3) fusion.

The construction of fusions and cassette PCR are two methods that can be used to create chimeras, but they are not the only methods; therefore, the following literature review will focus on these and other methods that have been applied to the chimeragenesis of P450s.

Chapter 2

Literature Review

Chimeragenesis of Cytochrome P450 Monooxygenases

2.1 Introduction

Cytochrome P450 monooxygenases (P450 or CYP) are a superfamily of heme-containing proteins that are found in all domains of life. These enzymes have a range of catalytic functions, including the biosynthesis and metabolism of endogenous compounds and the metabolism of xenobiotics (Sono *et al.*, 1996). P450s are able to catalyse diverse reactions including hydroxylations, epoxidations, heteroatom oxidations and reductions in a stereo- and regioselective manner *via* the insertion of a single atom of oxygen into an activated or unactivated bond in a substrate (Graham & Peterson, 1999; Danielson, 2002).

Most P450s require one or two proteins known as redox partners to transfer the electrons required for catalysis from the cofactors NADH or NADPH to the active-site heme iron. One type of electron transport system found in some bacteria and some eukaryotes is composed of three separate proteins: the P450 and two redox partners known as the ferredoxin reductase and the ferredoxin (Hannemann *et al.*, 2007). In eukaryotes, the ferredoxin reductase and the P450 are attached to the inner mitochondrial membrane (i.e. mitochondrial P450s). A second type of system that is found in some bacteria, but mainly in eukaryotic organisms, is composed of two separate proteins: the P450 and one redox partner known as the NADPH-cytochrome P450 reductase. In eukaryotes, both of these proteins are anchored to the membrane of the endoplasmic reticulum (i.e. microsomal P450s). A third type of electron transport system is that of the catalytically self-sufficient P450s, where the P450 and the redox partner(s) are fused together to form a single protein. These P450s do not require additional separate redox partners for electron transfer. This fusion arrangement may enhance the velocity and efficiency of catalysed reactions (Munro *et al.*, 2007).

Many of the reactions catalysed by P450s are difficult reactions to catalyse, even with the use of organic chemistry, making these enzymes potentially useful as biocatalysts in industry for the synthesis of value-added products. However, their application in industry is currently limited due to the fact that P450s have evolved to carry out specific biological tasks, in a specific environment, and are generally not suited to industrial conditions, exhibiting low stability, low activity and a dependence on expensive cofactors for catalysis (Zhao & Zha, 2004). Directed evolution has often been employed to target these and other P450 properties on a nucleic acid level.

Directed evolution is a repetitive process that results in the evolution of proteins on a scale of days or weeks as opposed to decades, centuries or millennia (Zhao & Zha, 2004). The directed evolution process consists of repetitive cycles or “rounds”. Each round consists of six steps (Fig. 2.1). The process begins with the selection of genes to target (“parent” genes). A library of variants is then generated and ligated to a plasmid for expression. The genes are expressed in bacteria, allowing the high-throughput screening of bacterial colonies and the identification of those exhibiting the property of interest. The improved gene(s) are then isolated and serve as the parent genes for the next round. The cycle is repeated until proteins exhibiting the desired property have been obtained.

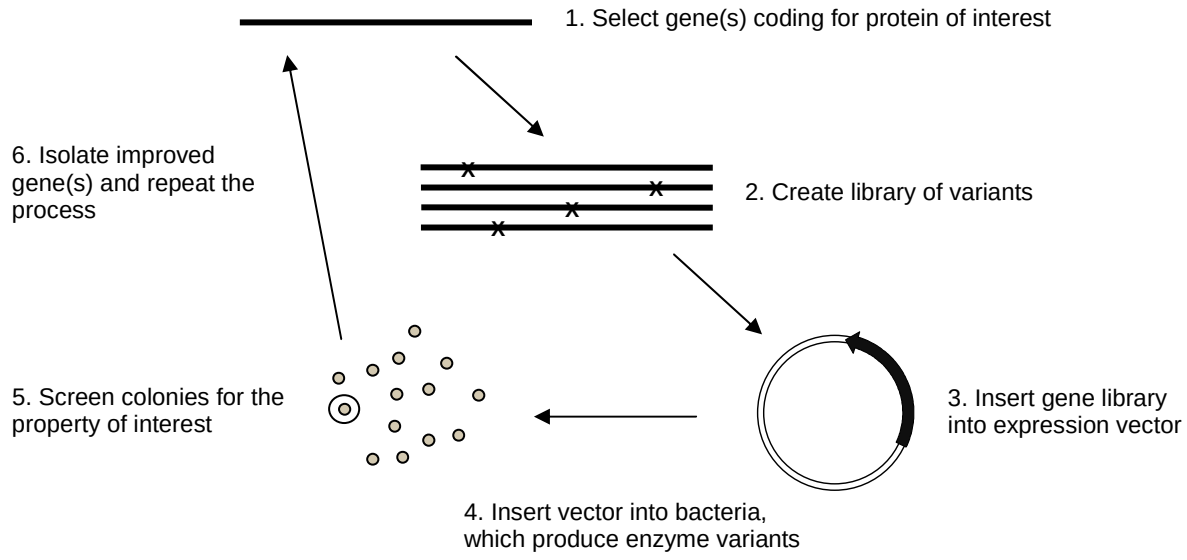


Figure 2.1 Steps involved in a typical directed evolution experiment (http://www.che.caltech.edu/groups/fha/).

One of the most important steps in the directed evolution process is the creation of a library of variants. The method used for generating this genetic diversity determines the quality of the library, which in turn determines the number of rounds required to obtain proteins exhibiting the property of interest. An approach that is often used for generating genetic diversity is chimeragenesis. Chimeragenesis refers to the linking of genes or gene fragments originating from different sources, while maintaining the correct open-reading frame, resulting in chimeric proteins (chimeras) which combine the biochemical properties and functions of different proteins into a single structure (Domanski & Halpert, 2001). The genes or gene fragments which are linked may or may not be closely related.

The chimeras that can be constructed range from very simple chimeras to more complex chimeras. Simple chimeras can be constructed by introducing point mutations into an existing protein based on the structure of a second protein, by linking large gene fragments encoding entire domains originating from two or three parent proteins, or by linking entire genes encoding proteins with separate activities (Nixon *et al.*, 1998). Complex chimeras are constructed using homologous or nonhomologous recombination, which involves the linking of smaller gene fragments encoding secondary structural elements originating from a number of parent proteins. Simple chimeras may also serve as the parent proteins for the construction of complex chimeras.

A number of different methods have been developed for the construction of chimeras. This literature review summarises the methods that have been applied specifically to the construction of chimeric P450s.

2.2 Methods for Constructing Simple Chimeras

2.2.1 Fusions

Fusions are simple chimeras that are generated when genes or gene fragments encoding two different proteins or domains from two different proteins are linked with the use of restriction enzymes or by overlap extension PCR (polymerase chain reaction), resulting in the expression of the genes or gene fragments as a single protein (Nixon *et al.*, 1998). The stop codon, if present, is eliminated from the first gene or fragment by PCR. The second gene or fragment is appended, in frame, to the 3'-end of the first, usually *via* a short DNA (deoxyribonucleic acid) sequence known as a linker (Fig. 2.2). The linker encodes a few amino acids which serve as a

"spacer", making it more likely that the two proteins or protein domains will be able to fold independently and behave as expected. The resulting chimera is expected to have properties derived from both of the original proteins.

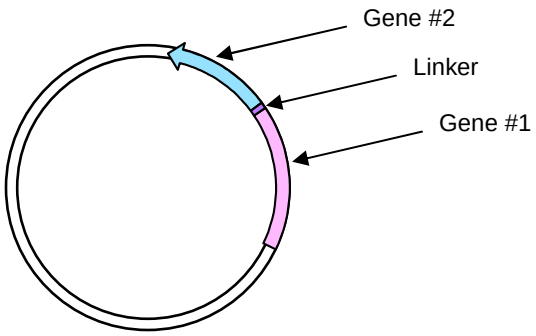


Figure 2.2 Schematic representation of a plasmid containing genes or gene fragments which will result in a fusion protein when expressed. The first gene or fragment is shown in pink, the second is shown in blue and the linker between them is shown in purple.

P450 fusions have been generated for a number of reasons. Eiben *et al.* (2007) constructed a fusion between the heme domain of $P450_{BM3}$ (CYP102A1) from *Bacillus megaterium* and the reductase domain of CYP102A3 from *Bacillus subtilis*, both of which are Class VIII fusion P450s (65% amino acid identity), with the aim of obtaining a thermostable self-sufficient P450. The resulting chimera exhibited 38% and 88% of the activity of $P450_{BM3}$ and CYP102A3, respectively, towards 12-*para*-nitrophenoxydodecanoic acid, but was thermostable at 51℃, a temperature at which both of the parent proteins were denatured.

Shimoji *et al.* (1998) constructed a P450 fusion consisting of approximately 50% membrane-bound mammalian CYP2C9 and 50% $P450_{cam}$ (CYP101A), a soluble Class I P450 from *Pseudomonas putida*, with the aim of investigating structure-function relationships. These proteins have less than 15% amino acid identity. The N-terminus of CYP2C9 was replaced with the N-terminus of $P450_{cam}$, resulting in the solubilisation of the membrane-bound P450. By utilising the putidaredoxin reductase and the putidaredoxin of the $P450_{cam}$ electron transport system, the fusion was able to convert 4-chlorotoluene to 4-chlorobenzyl alcohol with an activity of 0.167 nmoles per minute per nanomole of P450, which was higher than the activity of 0.078 nmoles per minute per nanomole of P450 for $P450_{cam}$ towards this substrate. The activity of the fusion was comparable to the CYP2C9 activity of 0.158 nmoles per minute per milligram of protein from microsomal fractions of expressed CYP2C9.

Sukumaran *et al.* (2002) constructed two fusions between the membrane-bound human P450 CYP2E1 and $P450_{cam}$ as a strategy for expressing CYP2E1 as a soluble protein in both *E. coli* and *Pseudomonas.* The first fusion was constructed by replacing the first 145 amino acids at the N-terminus of CYP2E1 with the first 130 amino acids of the $P450_{cam}$ N-terminus. The resulting fusion was soluble but exhibited a Soret band at 443 nm, possibly indicating low protein stability and degradation. The second fusion was constructed by replacing 28 amino acids at the C-terminus of CYP2E1 in the first fusion with 75 amino acids from the C-terminus of $P450_{cam}$. The resulting protein was soluble and exhibited a Soret band at 450 nm, indicating a correctly folded protein.

Gilardi and co-workers also generated a solubilised CYP2E1 by replacing the N-terminus of the CYP2E1 heme domain (residues 1 to 81) with the N-terminus of $P450_{BM3}$ (CYP102A1), the self-sufficient soluble P450 from *Bacillus megaterium* (Gilardi *et al.*, 2002). The heme domain was then fused to the reductase domain of $P450_{BM3}$, generating an artificial self-sufficient mammalian P450 exhibiting peaks at both 450 nm and 420 nm. Dodhia *et al.* (2006) used a similar approach to construct fusions between the heme domains of the human P450s CYP2C9, CYP2C19 and CYP3A4 and the reductase domain of $P450_{BM3}$. The resulting self-sufficient fusions had turnover rates comparable to those reported in literature for the native reconstituted mammalian P450 systems, but showed increased solubility.

A number of groups have created fusions using the reductase domain of a self-sufficient P450 to generate artificial self-sufficient P450s, with the aim of increasing the rate of electron transfer from the cofactor to the heme and thereby the rate of catalysis, or to decrease the complexity of the electron transport system for simpler P450 expression, often with larger-scale applications in mind. A reductase domain that has often been used for this purpose is that of P450RhF (CYP116B2) from *Rhodococcus* sp. NCIMB 9784 (Roberts *et al.*, 2002). Nodate and co-workers generated three self-sufficient chimeras by fusing this reductase domain to the P450 domains of three Class I P450s: $P450_{balk}$ (CYP153A13a) from *Alcanivorax borkumensis* SK2, $P450_{bzo}$ (CYP203A) from an environmental metagenomic library and $P450_{cam}$ (Nodate *et al.*, 2006). All three fusion proteins exhibited Soret bands at 450 nm. The $P450_{cam}$/PFOR(P450RhF) fusion resulted in 100% conversion of 0.5 mM (+)-camphor to 5'-*exo*-hydroxyl camphor after a period of 24 hours. The $P450_{bzo}$/PFOR(P450RhF) fusion resulted in complete conversion of 0.5 mM 4-hydroxybenzoate to 3,4-dihydroxybenzoate after a period of 4 hours. The $P450_{balk}$/PFOR(P450RhF) fusion produced 800 $mg.L^{-1}$ 1-octanol from octane after a period of 24 hours, allowing the first direct identification of the function of P450s in the CYP153A subfamily as that of terminal alkane hydroxylases.

Li *et al.* (2007) fused the reductase domain of P450RhF to PikC, a P450 involved in the biosynthesis of pikromycin by *Streptomyces venezuelae*. The natural redox partner of this P450 is unknown, but it is able to function with spinach ferredoxin reductase and ferredoxin as the redox partners. The constructed fusion resulted in a four-fold increase in the catalytic activity towards its substrate when compared to the wild-type PikC with spinach redox partners, indicating that this fusion arrangement stabilised the interaction between the heme domain and the redox partner(s), resulting in an enhanced electron transfer efficiency.

Although a number of P450 fusions have been constructed, not all fusions are successful. An example of such a fusion is one constructed by Fujita and co-workers between the P450 domain of a CYP153A protein from *Acinetobacter* sp. OC4 (AciA) and the P450RhF reductase domain (Fujita *et al.*, 2009). Fujii *et al.* (2006) cloned the genes encoding the P450 domain, the ferredoxin and the ferredoxin reductase of this CYP153A and expressed them in *E. coli*. This P450 complex was able to produce 2 250 $mg.L^{-1}$ 1-octanol from *n*-octane after 24 hours of incubation. The AciA/PFOR(P450RhF) fusion, however, although producing a Soret band at 450 nm, showed extremely poor activity towards octane as a substrate and a negligible amount of 1-octanol was produced (Fujita *et al.*, 2009).

2.2.1.1 Triple Fusions

Triple fusions are fusion variants that are generated when genes or gene fragments encoding three different proteins or domains from three different proteins are linked *via* short linkers to form a single protein. The stop codon is eliminated from the first two genes or gene fragments by PCR and these are then appended, in frame, to the 5'-end of the third gene or gene fragment.

The first triple fusion P450 was constructed by Harikrishna *et al.* (1993) between the cholesterol side-chain cleaving $P450_{scc}$, a human P450, and its redox partners, adrenodoxin reductase (AdRed) and adrenodoxin (Adx). Two variants of this fusion were constructed. With the first fusion, the adrenodoxin was positioned adjacent to the P450 ($P450_{scc}$-Adx-AdRed), but with the second fusion the adrenodoxin reductase was positioned adjacent to the P450 ($P450_{scc}$-AdRed-Adx). With both of these fusions, the linkers between the genes encoded five amino acids. The amount of pregnenolone produced from 22R-hydroxycholesterol by the first fusion was similar to that produced by the separate protein system. The second fusion, however, produced more pregnenolone, with an apparent V_{max} of 9.1 ng produced per millilitre of medium per 24 hours compared to the V_{max} of 1.7 ng per millilitre per 24 hours for the separate protein system.

Sibbesen *et al.* (1996) used a similar approach to construct a triple fusion protein between $P450_{cam}$ from *Pseudomonas putida* and its natural redox partners, putidaredoxin reductase (PdR) and putidaredoxin (Pd). Three fusions were constructed with the putidaredoxin reductase positioned adjacent to the P450 ($P450_{cam}$-PdR-Pd). The linkers between $P450_{cam}$ and PdR of these fusions encoded seven, seventeen and twenty-one amino acids. The linker between PdR and Pd remained constant, encoding seven amino acids. The rate of cytochrome c reduction by the fusion system was found to be slower than that of the native, separate protein system, with similar results being obtained for all three fusions. A fusion with the arrangement of PdR-Pd-$P450_{cam}$ was then constructed, with the linker region between the PdR and the Pd encoding seven amino acids while that between the Pd and the P450 encoded four amino acids. This fusion exhibited the highest activity, with a k_{cat} value of 30 min^{-1}, being three times more active than the best fusion obtained with the first fusion arrangement. At low P450 concentrations, this fusion was found to have a higher catalytic activity than that of the native system, but at concentrations higher than 0.3 μM, the native system expressed higher activity.

$P450_{c27}$ is a human P450 which catalyses the 25-hydroxylation of vitamin D_3 and the 27-hydroxylation of sterols. Dilworth *et al.* (1996) constructed a fusion between this P450 and its natural redox partners, adrenodoxin (Adx) and adrenodoxin reductase (AdR). The fusion was constructed in such a way that the adrenodoxin was positioned adjacent to the P450 (AdRed-Adx-$P450_{c27}$). The resulting fusion was able to convert 1α-hydroxyvitamin D_3 (1α-OH-D_3) to 1α,25-$(OH)_2$-D_3 and 1α,27-$(OH)_2$-D_3 four times more efficiently than the native separate protein system. With the natural substrate, vitamin D_3, the hydroxylation efficiency was lower than for 1α-hydroxyvitamin D_3, but was still 1.7-fold higher than that of the native separate protein system.

2.2.2 Cassette PCR

Cassette PCR is a method that was developed for the retrieval of proteins from environmental metagenomic sources as chimeric genes (Okuta *et al.*, 1998). It consists of two PCR steps (Fig. 2.3). The first PCR step involves the amplification of the internal fragments of genes from a specific subfamily of proteins, using degenerate primers that have been designed according to conserved amino acid sequences near the N- and C-termini of proteins in that subfamily. The template for this PCR is total DNA that has been extracted from mixed bacterial cultures. A cloned gene from the same subfamily serves as a “scaffold” for the construction of the chimeric genes and the 5’- and 3’-ends of this gene are PCR amplified. The reverse primer used for

amplifying the 5'-end and the forward primer used for amplifying the 3'-end are complementary to the degenerate primers used for amplifying the internal gene fragments. The three products of these two reactions are purified and mixed to serve as the template for the second PCR step.

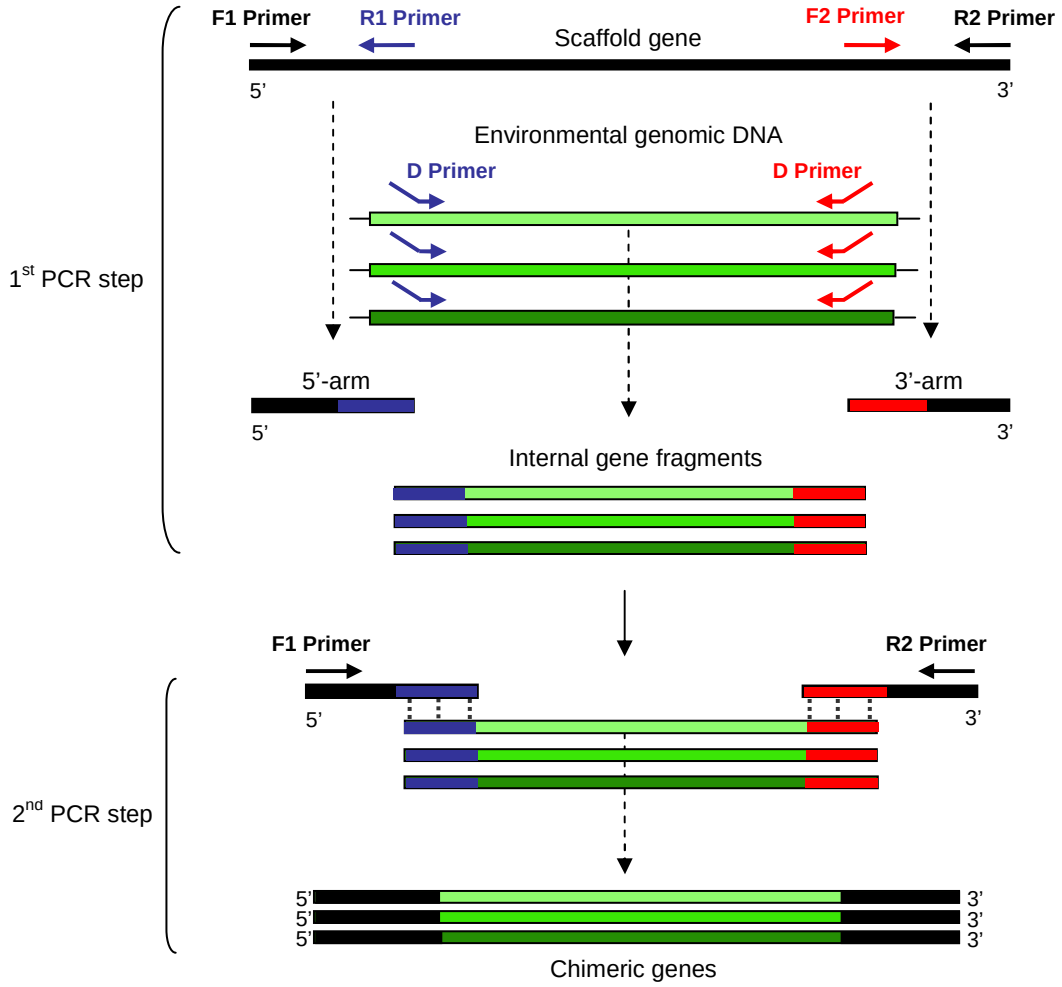


Figure 2.3 Scheme for cassette PCR. The degenerate primers used to amplify the internal gene fragments from environmental DNA are represented by "D Primer". The primers used for the amplification of the 5'- and 3'-ends of the scaffold gene are represented by "F1 Primer" and "R1 Primer", and "F2 Primer" and "R2 Primer", respectively. The primers used for assembly of the chimeric genes in the second PCR step are represented by "F1 Primer" and "R2 Primer" (Adapted from Okuta *et al.*, 1998).

During the second PCR step, the complementary sequences of the PCR products allow them to anneal each other and the sequences are extended by overlap extension PCR using the forward and reverse primers used to amplify the 5'- and 3'-arm fragments, respectively. This results in the three products being combined to form a single chimeric gene with the structure of (5'-arm)-(Central Gene Fragment)-(3'-arm).

Kubota and co-workers applied cassette PCR to the construction of chimeric *CYP153A* genes (Kubota *et al.*, 2005). They identified sixteen new P450s belonging to the CYP153A subfamily by amplifying the central fragment of the genes from total DNA prepared with enrichments of petroleum-contaminated soil, petroleum-contaminated groundwater and coastal seawater. The degenerate primers were designed according to the amino acid sequences of conserved domains in the CYP153A subfamily: MFIAMDPP near the N-terminus and HRCMGNRL, containing the heme-binding cysteine, near the C-terminus. The 5'- and 3'-arm fragments were amplified from the gene encoding $P450_{balk}$ (CYP153A13a) from *Alcanivorax borkumensis* SK2. Eight of the sixteen constructed chimeric genes resulted in functional CYP153A chimeric proteins. This was the only report of cassette PCR being applied to P450s, but this method has successfully been applied to the construction of catechol 2,3-dioxygenase chimeras (Okuta *et al.*, 1998).

The disadvantage of cassette PCR is that it requires information about conserved amino acid sequences to design primers, requiring a total of two conserved regions, one at the N-terminus and one at the C-terminus (Okuta *et al.*, 1998). The advantage of this method is that it allows the isolation of genes from the environment without requiring the isolation of microorganisms.

2.3 Methods for Constructing Complex Chimeras

2.3.1 DNA Shuffling

DNA shuffling is an *in vitro* recombination method that results in a library of chimeric genes (Stemmer, 2002) and consists of three main steps (Fig. 2.4). The first step is the random fragmentation of genes using Deoxyribonuclease I (DNase I). This enzyme cleaves DNA adjacent to pyrimidine residues and, in the presence of magnesium ions (Mg^{2+}), cleaves the two strands of double-stranded DNA independently, resulting in a pool of DNA fragments of varying lengths (Reid, 2000).

The second step is the random reassembly of the DNA fragments in a primerless PCR which is a standard PCR performed in the absence of primers (Stemmer, 1994a). During this step, the DNA fragments are denatured and the resulting single-stranded fragments "prime" each other by hybridisation based on sequence homology. This allows the extension of fragments with a 5'-overhang (Fig. 2.5), but not a 3'-overhang (Reid, 2000). It is in this second step of DNA shuffling

where recombination occurs: when a fragment originating from one gene primes a fragment originating from a different gene, there is a template switch which results in a crossover between the two genes (Stemmer, 1994a). As the DNA fragments are reassembled, point mutations are introduced at a rate of approximately 0.7%, which increases the diversity of the reassembled products.

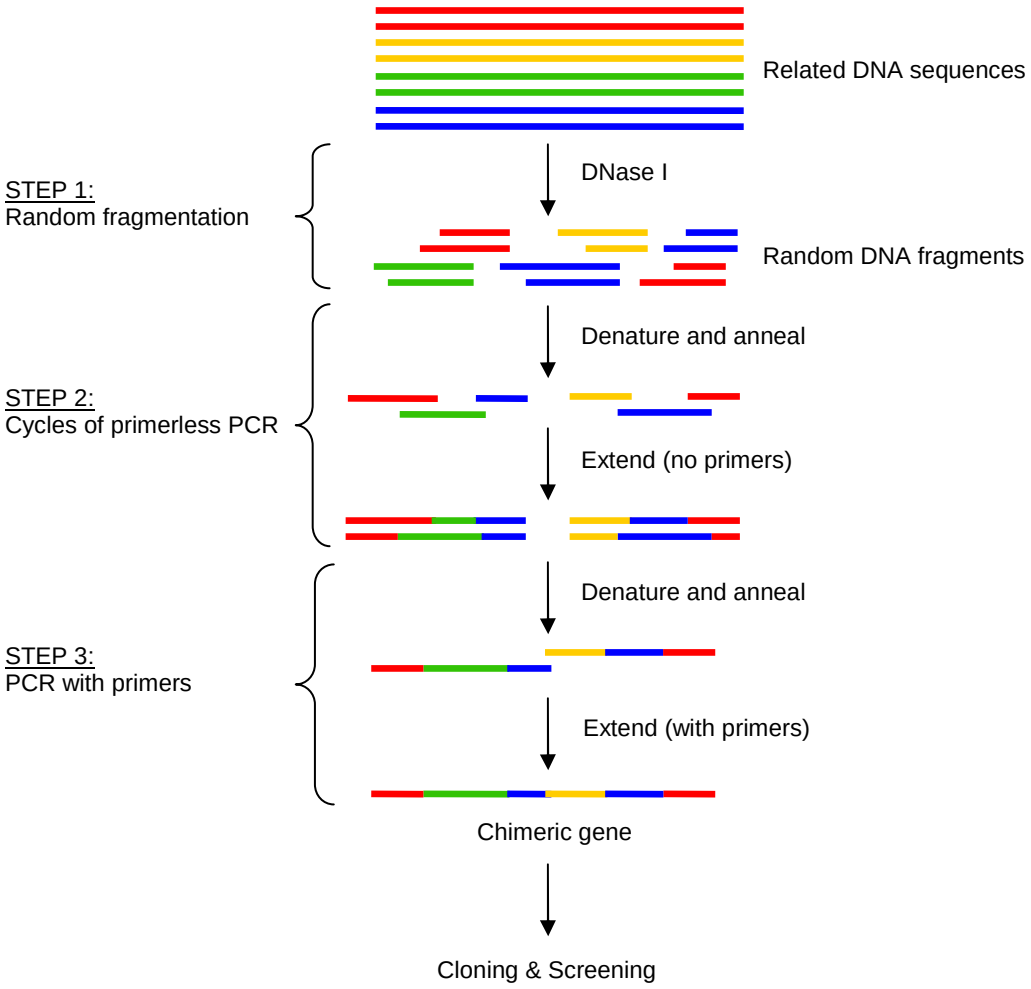


Figure 2.4 Schematic representation of the steps involved in a general DNA shuffling experiment (Adapted from Minshull & Stemmer, 1999).

The third step of DNA shuffling is the amplification of the reassembled products by PCR with primers corresponding to the 5'- and 3'-ends of the genes, to obtain full-length chimeric genes (Stemmer, 1994b). This is followed by cloning in an expression vector for screening and selection (Gillam, 2005). Clones that show improvements in the property of interest are then used as "parent" genes for a new round of DNA shuffling (Stemmer, 1994a).

Figure 2.5 Schematic representation of the two possibilities that result from DNA fragments priming each other. Those with 5'-overhangs will result in extension whereas those with 3'-overhangs cannot be extended.

The DNA shuffling method requires some optimisation with respect to the concentration of DNA used for the reassembly and amplification steps, and the size of the fragments used for reassembly (Rosic *et al.*, 2007). DNA fragments of ten to fifty basepairs (bp) can successfully be reassembled into functional, full-length genes of more than two kilobasepairs (kb), but the size of the fragments selected for reassembly is dependent on the desired number of crossovers in the chimeras (Stemmer, 1994a).

2.3.1.1 Family Shuffling with DNase I

In the original DNA shuffling method, the parent genes were variants of a single gene generated by point mutations introduced using error-prone PCR or site-directed mutagenesis (Rosic *et al.*, 2007). When variants of a single gene are used as the parent genes, the recombined genes may accumulate the point mutations present in the parent genes in different combinations, but the active clones obtained usually differ from the parent sequences by between one and three amino acids only (Stemmer, 2002). One of the major disadvantages of this method is that the accumulation of mutations which are beneficial or which will result in improvements in a particular property may occur quite slowly (Crameri *et al.*, 1998). In order to overcome this disadvantage, DNA shuffling was expanded to family shuffling where the parent genes are homologous genes from nature. These genes could either be related genes from one species or a single gene cloned from related species (Stemmer, 2002). The chimeric products of cassette PCR may also serve as parent genes for family shuffling (Kagami *et al.*, 2004).

This method has the advantage that sequences important for structure and function tend to be conserved within a particular protein family and by shuffling homologous genes the chance of obtaining active proteins is increased. In addition, beneficial mutations that have been selected for in nature can be accumulated in the progeny genes, providing the genetic diversity required

for directed evolution. The progeny genes tend to have a high number of mutations when compared to the parent sequences, but most of the mutations introduced into the progeny genes will be present in at least one of the parent genes (Minshull & Stemmer, 1999; Gillam, 2007).

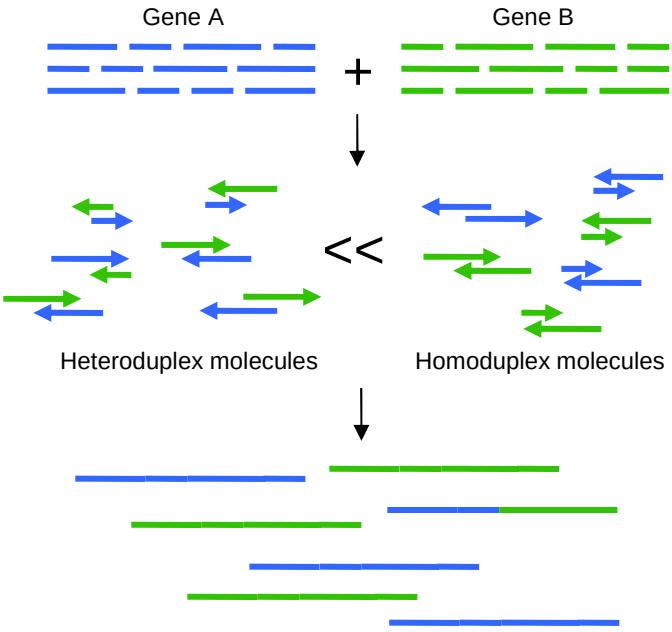


Figure 2.6 Schematic representation of the situation where the probability of forming homoduplex molecules is greater than that of forming heteroduplex molecules. When this occurs, a large number of parental sequence duplexes will reform (Adapted from Kagami *et al.*, 2004).

One of the disadvantages of family shuffling is that it results in very large libraries, requiring extensive screening (Gillam, 2005). For example, if only two parent genes are used that differ at twenty amino acid positions, a library of 2^{20} or one million different chimeras is created and these all need to be screened (Stemmer, 2002). The second disadvantage is that a successful family shuffling experiment requires parent genes with at least 70% sequence identity for recombination to occur (Farinas *et al.*, 2001). In addition, libraries created by family shuffling tend to contain a high proportion of parental forms which are reconstituted by PCR-based reassembly (Kagami *et al.*, 2004). This occurs when the probability of homoduplex formation is greater than the probability of heteroduplex formation (Fig. 2.6). A homoduplex forms when DNA fragments originating from one gene hybridise to each other, resulting in extension, whereas a heteroduplex forms when the DNA fragments originate from different genes (Kikuchi *et al.*, 1999). Homoduplex formation occurs as a result of incompletely digested parental DNA, or as a result of high numbers of amino acids that differ between the parent genes, and can decrease the recombination frequency of DNA shuffling to less than 1% (Kagami *et al.*, 2004).

2.3.1.2 Family Shuffling with Restriction Enzymes

The family shuffling method was modified by Harayama and co-workers in order to decrease the probability of homoduplex formation and thereby increase the proportion of chimeras in a shuffled library (Kikuchi *et al.*, 1999). Instead of using DNase I to randomly fragment the genes for shuffling, each gene is cleaved with a different restriction enzyme or a combination of restriction enzymes. The resulting fragments are mixed and reassembled by primerless PCR. This approach results in a higher frequency of recombination as extension will not occur when fragments originating from one gene hybridise to each other (Fig. 2.7). Extension will occur only when fragments from different genes hybridise, resulting in 5'-overhangs (Reid, 2000).

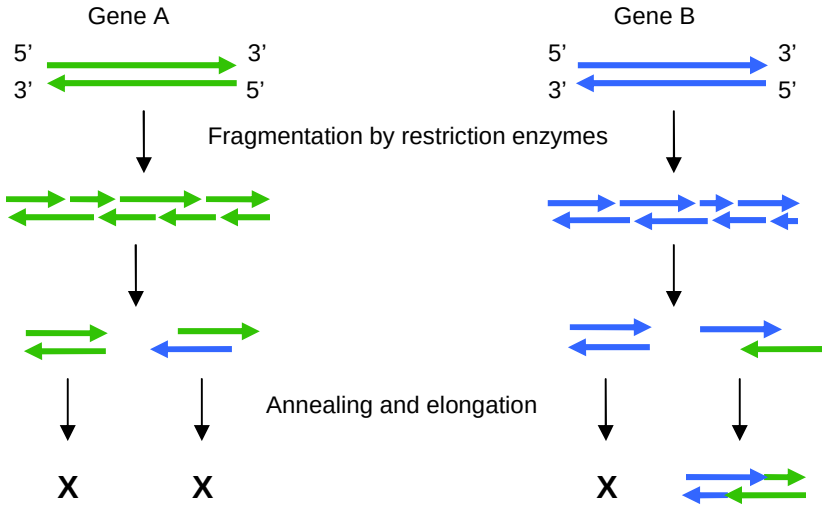


Figure 2.7 Schematic representation of the steps involved in family shuffling using restriction enzymes. Fragments originating from one parent gene or fragments that anneal to form 3'-overhangs will not be extended (Adapted from Kagami *et al.*, 2004).

Gillam and co-workers used this approach to shuffle homologous mammalian P450 genes of the CYP2 family. The first library was constructed using the genes encoding CYP2C9, CYP2C11 and CYP2C19 (Rosic *et al.*, 2007). The DNA was cleaved with *Mnl*I or an *Mse*I-*Hin*fI combination and the resulting fragments, which were less than 300 bp in length, were reassembled and amplified. Fifty-four clones were randomly sampled and their sequences determined. No parental forms were detected, indicating minimal parental sequence contamination in the library. The estimated number of crossovers obtained per 1.5 kb DNA sequence was between three and seven and the mutation rate was between five and eleven point mutations per 1.5 kb sequence. Fifteen percent of the clones tested exhibited Soret bands

at 450 nm. Five hundred clones were screened for the production of indigo from indole and four clones were identified that exhibited similar or higher levels of indigo pigment production to those of the parental P450s.

The second library was constructed using the genes encoding CYP2C8, CYP2C9, CYP2C18 and CYP2C19, which were digested with either an *Alu*I-*Bsa*JI or an *Mse*I-*Fnu*4HI restriction enzyme combination, generating between seven and fifteen DNA fragments (Huang *et al.*, 2007). Fragments that were smaller than 300 bp were used for the reassembly step and, once again, no parental sequences were detected in randomly sampled clones. When compared with the first shuffled library, the number of crossovers obtained per 1.5 kb increased to between seven and eleven and the mutation rate decreased to between one and two point mutations per 1.5 kb sequence. Fifty-four percent of the sampled clones exhibited Soret bands at 450 nm. Ninety-six clones were screened for indigo production, but only one showed an elevated level compared to the parental forms.

The advantage of this method is that it can result in increased recombination frequency and can decrease the parental sequence background in the library. This method can sometimes recombine genes that the conventional family shuffling approach cannot recombine (Zhao & Zha, 2004). The disadvantage of using restriction enzymes for fragmentation is that recombination is not as random as it is with DNase I as the restriction sites are not random, resulting in chimeric genes with decreased diversity. Another disadvantage is that this method results in a low number of crossovers (Kagami *et al.*, 2004).

2.3.2 CLERY

CLERY (Combinatorial Libraries Enhanced by Recombination in Yeast) is a variant of family shuffling that combines the PCR-based steps with *in vivo* recombination in yeast (Abécassis *et al.*, 2000). This method consists of four main steps (Fig 2.8). The first step involves the random digestion of vectors containing the DNA of interest with DNase I in the presence of manganese ions (Mn^{2+}), which results in the cleavage of double-stranded DNA and the generation of small DNA fragments. The fragments are reassembled by primerless PCR, followed by amplification of the reassembled products with primers designed according to vector sequences flanking the DNA. The amplified products are then mixed with a yeast expression vector which is linearised at the expression site, and this mixture is used to co-transform yeast. The gap-repair system of

yeast results in a circularised vector by the insertion of the PCR fragment into the expression site of the linearised vector *via* homologous recombination. This results in a library of chimeras expressed in yeast.

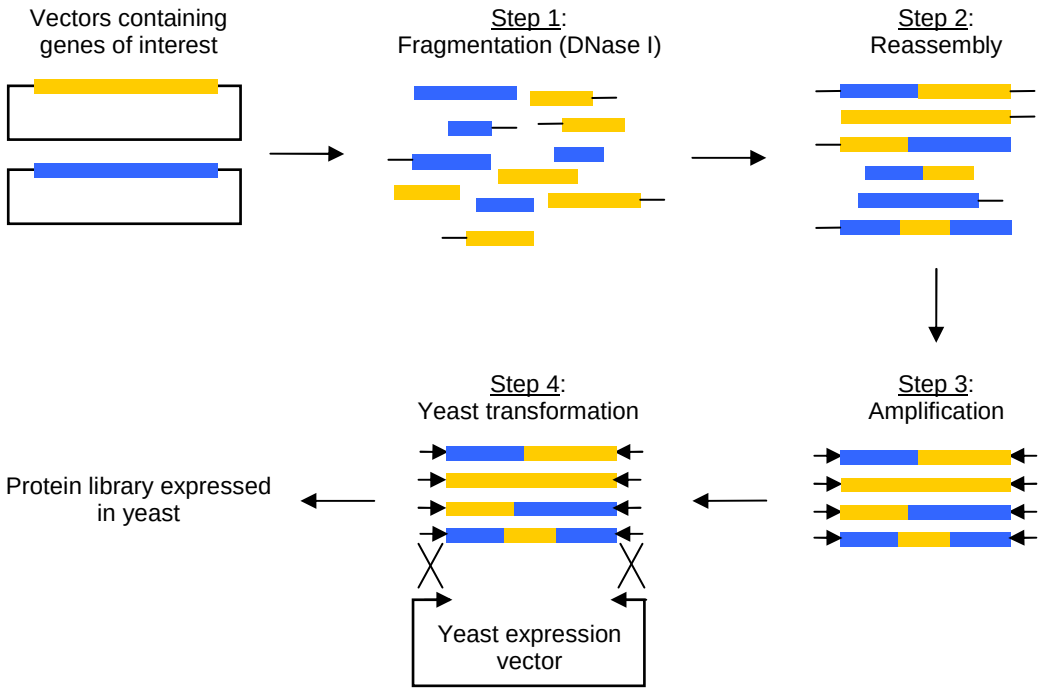


Figure 2.8 Schematic representation of the CLERY process (Adapted from Abécassis *et al.*, 2000).

Abécassis *et al.* (2000) used this approach to generate a chimeric library of the human P450s CYP1A1 and CYP1A2, expressed in *Saccharomyces cerevisiae*. These P450s are involved in the metabolic activation of carcinogens and the genes encoding these enzymes share 74% nucleotide identity. Sequence and statistical analysis of randomly selected clones from the library revealed that 86% of the genes were chimeric, with the average number of fragments composing the genes being 5.4. There was an almost equal representation of the parental sequences in the chimeras, with 55.8% of each chimeric gene consisting of *CYP1A2*. The average number of point mutations introduced into chimeric genes encoding functional proteins was 8.3 whereas chimeric genes encoding non-functional proteins had an average of 14 mutations, with at least one stop codon being generated in each of these genes. Clones were screened for activity towards naphthalene, a good substrate for both parental enzymes, and 11.8% of these clones expressed a detectable activity.

CLERY has a number of advantages. The use of yeast as a host for eukaryotic P450s allows the direct expression and selection or screening of active clones without requiring intermediate steps in *E. coli* (Abécassis *et al.*, 2000). Homologous recombination results in efficient cloning and introduces a different molecular technique into the family shuffling process. Each clone in the library may contain multiple chimeric gene vectors, increasing the complexity of the library. The disadvantage of this method is that the level of protein expression obtained with yeast may be quite low compared to the level of expression that can be obtained with bacterial systems (Gillam, 2005).

2.3.3 SHIPREC

SHIPREC (Sequence Homology-Independent Protein Recombination) is a method developed for the *in vitro* recombination of distantly related or unrelated proteins (Sieber *et al.*, 2001). The SHIPREC method begins with the construction of a gene dimer (Fig. 2.9). The gene encoding one protein is fused to the gene encoding a second protein *via* a linker sequence that contains unique restriction enzyme sites. The dimer is then fragmented by DNase I digestion in the presence of manganese ions (Mn^{2+}), which results in DNase I cutting the two strands of DNA at approximately the same position. A library of random fragments is generated and fragments that are the length of one parent gene plus the length of the linker are separated from the others and treated with S1 nuclease. This enzyme degrades single-stranded DNA, producing blunt-ended fragments. This is followed by blunt-end ligation, resulting in circularised DNA fragments, which are then linearised by restriction enzyme digestion in the linker sequence. The gene that was initially at the 5'-end of the dimer is then positioned at the 3'-end of the linearised fragment. The resulting chimeric genes are amplified by PCR with one primer from the terminal end of each of the two parent genes, and are then cloned in an expression vector for screening and selection.

SHIPREC was applied to the construction of chimeras between the human P450 CYP1A2, which is membrane-bound, and the heme domain of $P450_{BM3}$ (Sieber *et al.*, 2001). These P450s share 16% amino acid sequence identity. In the full-length variants, 43% of the crossovers occurred in the first third of the chimera, whereas approximately 28% occurred in each of the remaining thirds. Carbon monoxide-difference spectra were performed on a total of 116 variants and 80% of these exhibited Soret bands at 450 nm. These chimeras consisted mainly of $P450_{BM3}$ with a crossover to CYP1A2 occurring only at the far C-terminus of each chimera, suggesting that crossovers occurring within the core structure of P450s result in disruption of the P450 structure (Gillam, 2005). Two thousand chimeras were screened for activity towards 7-ethoxyresorufin, a CYP1A2 substrate towards which $P450_{BM3}$ does not show any activity. Only

two chimeras showed activity towards the substrate and they were found to be more soluble than the wild-type CYP1A2 (Sieber *et al.*, 2001).

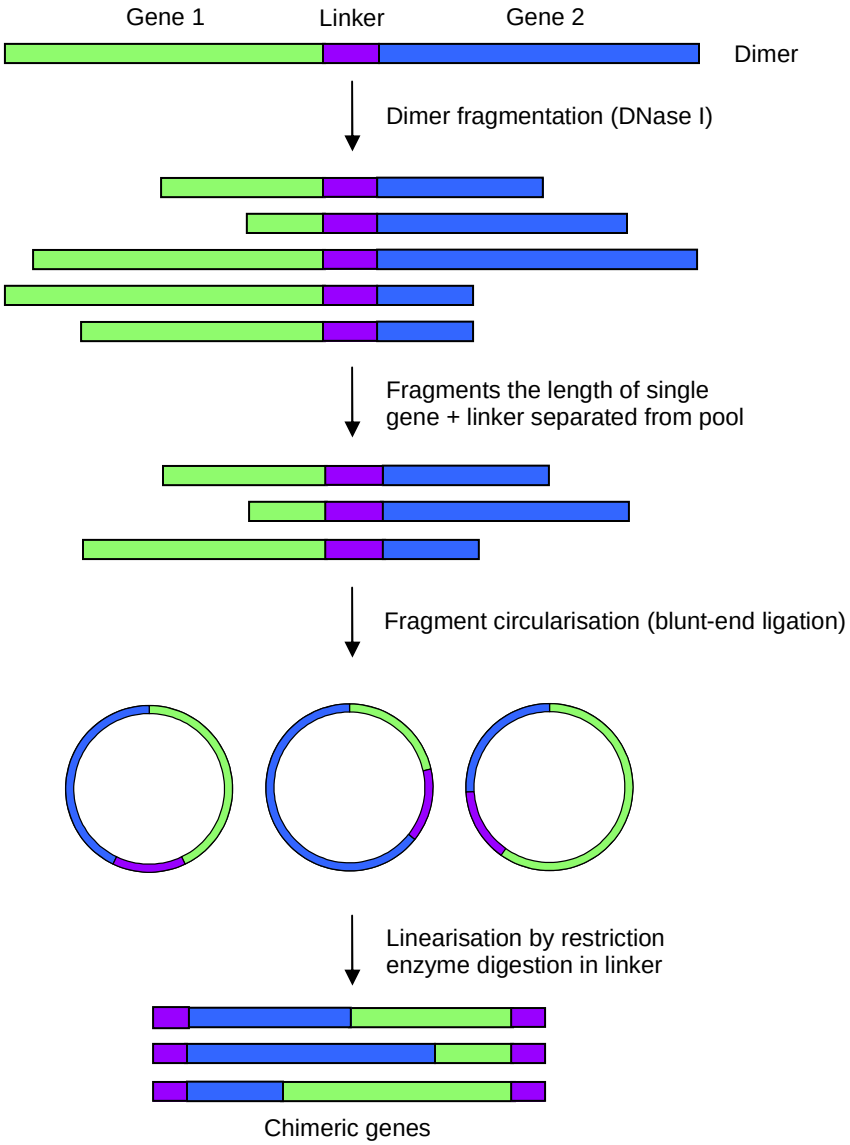


Figure 2.9 Schematic representation of the SHIPREC method (Adapted from Sieber *et al.*, 2001).

The advantage of this method is that genes that have no sequence homology can be recombined. Another advantage is that the step at which fragments of the expected size are selected ensures that the amino acids that meet at the crossover are in structurally related sites in the two parent proteins (Sieber *et al.*, 2001; Kagami *et al.*, 2004). The disadvantage of this method is that only a single crossover occurs between the parent genes. As a result, this method is limited to the recombination of only two parent genes.

2.3.4 SISDC

SISDC (Sequence-Independent Site-Directed Chimeragenesis) is a method that can be used for the chimeragenesis of related, distantly related or unrelated proteins (Hiraga & Arnold, 2003). This method consists of four steps (Fig. 2.10).

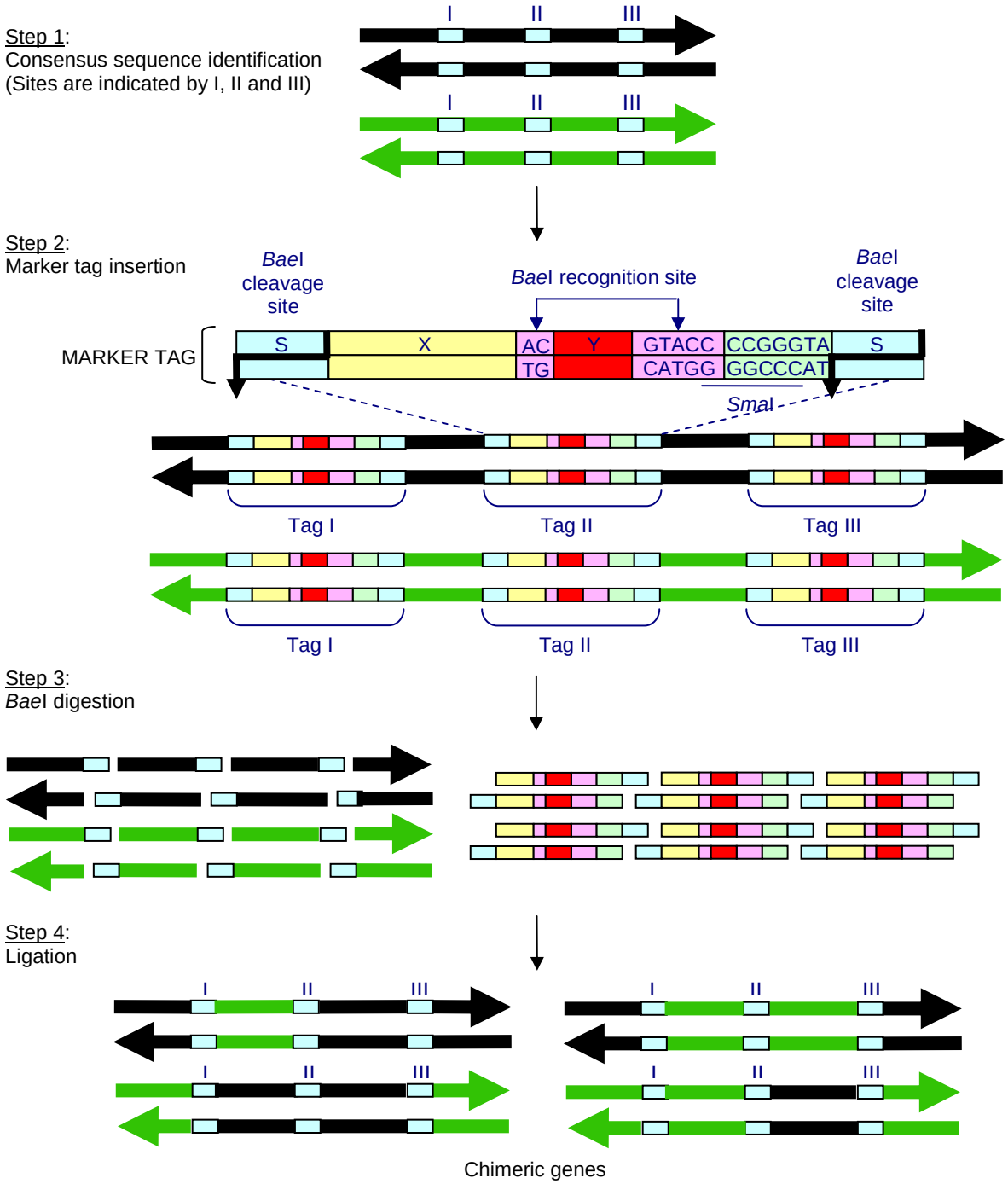


Figure 2.10 Schematic representation of the SISDC method (Adapted from Hiraga & Arnold, 2003).

The first step of this method involves the alignment of the nucleotide sequences of two or more parent genes, with the aim of identifying consensus sequences (Hiraga & Arnold, 2003). These are regions of identity within the parental sequences which, when translated, will result in one or two amino acids that are identical in the parent sequences. These regions can then be selected as crossover sites.

The second step involves the insertion of sequences, referred to as marker tags, into the crossover sites by PCR (Hiraga & Arnold, 2003). The marker tags targeting each crossover site have unique sequences, which can be divided into four regions (Fig. 2.10, Step 2). All marker tags contain the recognition sequence of a type IIb restriction enzyme, often *Bae*I. These restriction enzymes digest double-stranded DNA on both sides of the recognition sequence, resulting in two overhangs. Each marker tag contains two variable regions, an upstream region (referred to as X) and a middle region (referred to as Y). Tags targeting different crossover sites must not contain identical sequences in these regions. Adjacent to the type IIb endonuclease recognition site is the recognition sequence for *Sma*I, which is used in the final step.

The third step involves the digestion of the parent genes using the type IIb endonuclease, resulting in the removal of the marker tags and the generation of sticky ends. In the fourth and final step, the fragments from different genes, which have been purified to remove the tags, are mixed and are allowed to ligate, resulting in the formation of a library of chimeric genes that have been assembled in the correct sequence order. During this final step *Sma*I treatment will remove any marker tags which may still be present in the library.

As previously mentioned, the marker tags are inserted into the crossover sites by PCR. To achieve this, each gene is amplified as a series of fragments using primers that are composed of a part of the marker tag sequence, including the X and Y sequences, and a part which is complementary to the parent gene sequence (Fig. 2.11). Each amplified fragment therefore contains part of a marker tag, which facilitates the reassembly of the parent genes by primerless PCR as the complementary X and Y regions of a particular marker tag prime each other. The result is that the genes are reassembled in the correct order, with the marker tag being inserted between two copies of the consensus sequence. The number of primers required to insert the marker tags is dependent on the number of parent genes and the number of fragments to be shuffled: Primer no. = 2 x P x E where P = Number of parent genes and E = Number of fragments or elements (Hiraga & Arnold, 2003).

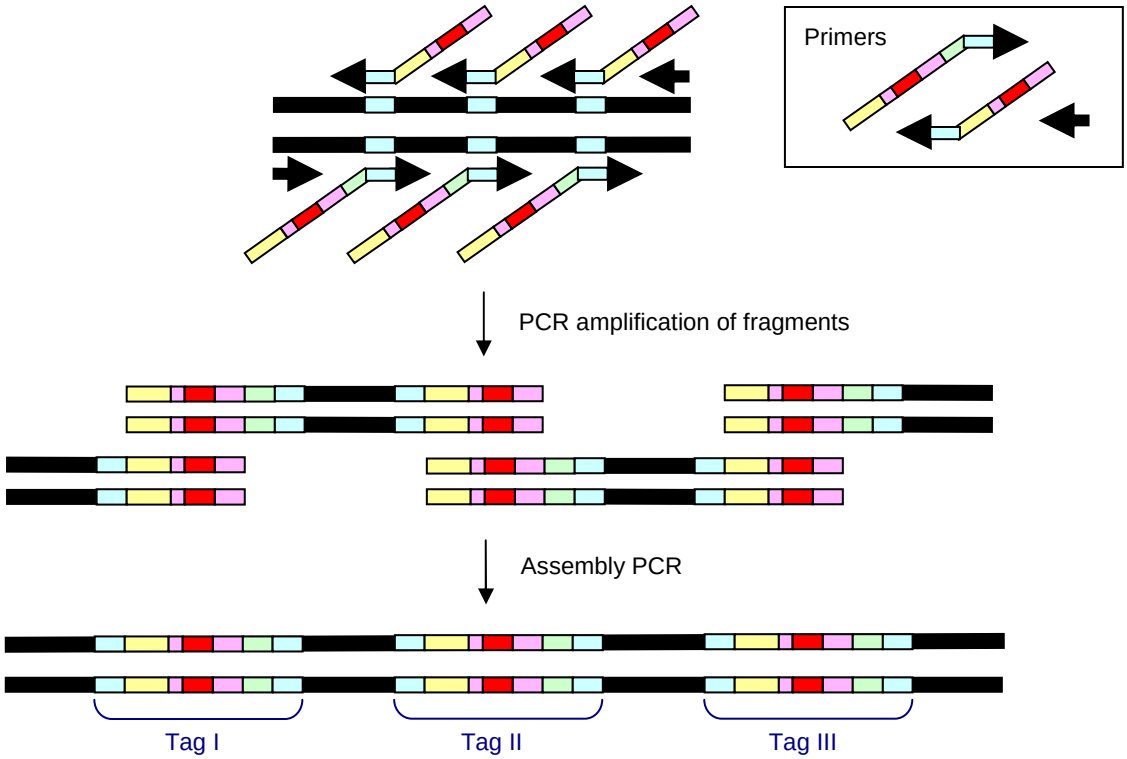


Figure 2.11 Insertion of marker tags into consensus sequence sites by PCR (Adapted from Hiraga & Arnold, 2003).

One of the main drawbacks of chimeragenesis is that often very large numbers of misfolded and therefore inactive chimeric proteins are generated, which can make screening a very tedious process. As a way of overcoming this problem, Arnold and co-workers developed a computer algorithm called SCHEMA. This algorithm uses the three-dimensional data of proteins to identify protein fragments that can be exchanged while minimising disruptive interactions that result in misfolded proteins. This minimises the number of misfolded proteins, but maximises the number of crossovers between parent genes (Bernhardt, 2004).

The SISDC method was developed by Arnold and co-workers specifically for use with SCHEMA (Otey *et al.*, 2006). They used the combination of SISDC and SCHEMA to generate a chimeric protein library of three cytochrome P450 monooxygenase heme domains: $P450_{BM3}$ (CYP102A1) from *Bacillus megaterium* and CYP102A2 and CYP102A3 from *Bacillus subtilis*. The heme domains of these three P450s have an average amino acid identity of 65%. SCHEMA identified seven crossover sites that would minimise disruptions, dividing each of the three heme domains into eight fragments. The resulting twenty-four fragments were shuffled using SISDC with *Bsa*XI as the type IIb restriction enzyme. A library containing 6 561 different chimeric gene sequences was generated and these genes were expressed in *Escherichia coli* and subsequently subjected

to screening and various assays. Approximately 3 000 of the expressed proteins were properly folded P450s. On average, the chimeric proteins differed from the parental proteins by 72 amino acids. Some of the chimeras were more thermostable than the parent proteins, with the most thermostable chimera differing by 84 amino acids from its closest parent, but having a melting temperature 7℃ higher than the most stable parent.

Landwehr *et al.* (2007) reported further analysis of the chimeric library by reconstituting 14 chimeric heme domains with the three parental reductase domains and determining the activities of these P450s. The CYP102A2 and CYP102A3 reductase domains share between 52% and 55% sequence identity with the reductase domain of CYP102A1. Functional P450 chimeras were obtained in all cases and for all the substrates tested (eleven in total) the best-performing enzyme was a chimera.

SISDC has a number of advantages. The marker tags are designed so that the fragments ligate in the correct gene sequence order, with a single ligation step being required for gene reassembly (Hiraga & Arnold, 2003). This method can be used to shuffle two or more genes, and the molar concentration of fragments from each gene can be controlled, thereby decreasing bias during the ligation step. In addition, multiple crossovers per gene are possible. The disadvantage of this method is that the gene sequence of each parent gene is required as consensus sequences need to be identified. This method requires a large number of primers to facilitate marker tag insertion.

2.4 Concluding Remarks

A number of different methods have been developed for the construction of chimeras, ranging from those which require only two or three starting genes to those which can be applied to a larger number of sequences. Many of the methods require the genes of interest to be closely related (at least 70% amino acid identity), but some methods have been developed that can be applied to relatively unrelated or distantly related sequences (from less than 15% nucleotide identity to about 60% amino acid identity). Each method has its own advantages and disadvantages (Table 2.1).

The methods that have been described in this literature review have successfully been applied to the chimeragenesis of P450s (Table 2.2), both related and unrelated, often resulting in chimeras with improved properties when compared to the wild-type protein. Ultimately, the

method of choice for constructing chimeric P450s, as with all proteins, is determined by the parent genes to be used and the downstream application(s) of the resulting chimera(s).

Table 2.1 Advantages and disadvantages of chimeragenesis methods that have been applied to P450s

Method	Advantages	Disadvantages
Fusion & Triple Fusion	·Genes or fragments are linked in frame ·High-throughput screening is not required	·Limited to two or three genes or fragments ·Restriction enzymes that do not cleave within the genes or gene fragments must be identified
Cassette PCR	·Isolation of environmental genes without isolating bacteria	·Requires sequence information
Family Shuffling with Restriction Enzymes	·Decreases parental contamination in library ·Recombines genes that family shuffling cannot	·Recombination is not random ·Limited number of crossovers per gene
CLERY	·Does not require intermediate steps in *E. coli* ·Increases library complexity	·Low level of protein expression
SHIPREC	·Recombines unrelated genes ·Crossovers occur mostly at structurally related sites	·Single crossover ·Limited to the recombination of two parent genes ·May introduce insertions or deletions at crossover point
SISDC	·Can be applied to two or more genes ·Molar concentration of fragments can be controlled, decreasing bias ·Multiple crossovers per gene	·Requires sequence information ·Requires many primers

Table 2.2 Summary of the successful application of chimeragenesis to P450s

Method	P450s	Results	Reference
Fusion	**Heme domains:** ·CYP17A (Bovine) ·CYP4A1 (Rat) **Reductase domain:** ·NADPH-P450 reductase (Rat)	·Self-sufficient mammalian P450s ·Some reactions required addition of phospholipid & cytochrome b_5	Fisher *et al.* (1992)
Fusion	**Heme domain:** ·CYP3A4 (Human) **Reductase domain:** ·NADPH-P450 reductase (Rat)	·Self-sufficient mammalian P450 · Some reactions required addition of lipids, detergents & cytochrome b_5	Shet *et al.* (1993)
Fusion	**Heme domain:** ·CYP2B11 (Dog) **Reductase domain:** ·NADPH-P450 reductase (Rat)	·Self-sufficient mammalian P450 ·Activity towards androstenedione similar to that of reconstituted systems	Harlow & Halpert (1996)
Fusion	**Heme domain:** **N-terminus:** ·$P450_{cam}$ (Bacterium) **C-terminus:** ·CYP2C9 (Human) ➔*Less than 15% nucleotide identity*	·Soluble P450 ·Higher activity than $P450_{cam}$ towards 4-chlorotoluene, comparable to that of CYP2C9	Shimoji *et al.* (1998)
Fusion	**Heme domain:** ·CYP2D6 (Human) **Reductase domain:** ·NADPH-P450 reductase (Human)	·Self-sufficient mammalian P450 ·Activities towards substrates similar to those with co-expressed CYP2D6 and reductase	Deeni *et al.* (2001)
Fusion	**Heme domain:** **N-terminus:** ·$P450_{cam}$ (Bacterium) **Central part:** ·CYP2E1 (Human) **C-terminus:** ·$P450_{cam}$ (Bacterium)	·Soluble P450	Sukumaran *et al.* (2002)
Fusion	**Heme domain:** **N-terminus:** ·$P450_{BM3}$ (Bacterium) **Central part & C-terminus:** ·CYP2E1 (Human) **Reductase domain:** ·$P450_{BM3}$ (Bacterium)	·Soluble, self-sufficient P450 ·Turnover rates comparable to native reconstituted systems	Gilardi *et al.* (2002)
Fusion	**Heme domains:** ·$P450_{cam}$ (Bacterium) ·$P450_{bzo}$ (Bacterium) ·$P450_{balk}$ (Bacterium) **Reductase domain:** ·P450RhF (Bacterium)	·Self-sufficient P450s capable of converting typical substrates	Nodate *et al.* (2006)
Fusion	**Heme domain:** ·PikC (Bacterium) **Reductase domain:** ·P450RhF (Bacterium)	·Four-fold increase in activity towards substrate compared to activity with separate spinach redox partners	Li *et al.* (2007)

Method	P450s	Results	Reference
Fusion	**Heme domain:** ·$P450_{BM3}$ (Bacterium) **Reductase domain:** ·CYP102A3 (Bacterium) ➔ *65% amino acid identity*	·Self-sufficient P450 thermostable at 51℃	Eiben *et al.* (2007)
Triple Fusion	**Heme domain:** ·$P450_{scc}$ (Human) **Redox partners:** ·AdR & Adx ($P450_{scc}$)	·Pregnenolone production apparent Vmax = 9.1 ng per mL per day. Native system Vmax = 1.7 ng per mL per day	Harikrishna *et al.* (1993)
Triple Fusion	**Heme domain:** ·$P450_{cam}$ (Bacterium) **Redox partners:** ·PdR & Pd ($P450_{cam}$)	·Catalytic activity higher than native system at P450 concentrations lower than 0.3 μM	Sibbesen *et al.* (1996)
Triple Fusion	**Heme domain:** ·$P450_{c27}$ (Human) **Redox partners:** ·AdR & Adx ($P450_{c27}$)	·Four-fold higher activity towards 1α-hydroxyvitamin D_3 than native system ·1.7-fold higher activity towards vitamin D_3 than native system	Dilworth *et al.* (1996)
Cassette PCR	**Central fragments:** ·CYP153A (Bacteria) **N- and C-termini:** ·$P450_{balk}$ (Bacterium)	·50% correctly folded chimeric P450s	Kubota *et al.* (2005)
Family Shuffling with Restriction Enzymes	**Heme domain:** ·CYP2C9 (Human) ·CYP2C11 (Human) ·CYP2C19 (Human) ➔ *More than 80% amino acid identity*	·15% correctly folded P450s ·0.8% of screened clones produced indigo at similar or higher levels than parental P450s	Rosic *et al.* (2007)
Family Shuffling with Restriction Enzymes	**Heme domain:** ·CYP2C8 (Human) ·CYP2C9 (Human) ·CYP2C18 (Human) ·CYP2C19 (Human) ➔ *More than 84% nucleotide identity*	·54% correctly folded P450s ·1% of screened clones produced indigo at higher levels than parental P450s	Huang *et al.* (2007)
CLERY	**Heme domain:** ·CYP1A1 (Human) ·CYP1A2 (Human) ➔ *74% nucleotide identity*	·11.8% of screened clones exhibited detectable activity towards naphthalene	Abécassis *et al.* (2000)
SHIPREC	**Heme domain:** ·CYP1A2 (Human) ·$P450_{BM3}$ (Bacterium) ➔ *16% amino acid identity*	·80% correctly folded P450s ·0.1% of screened clones exhibited activity towards 7-ethoxyresorufin, a CYP1A2 substrate ·Increased P450 solubility	Sieber *et al.* (2001)

Method	P450s	Results	Reference
SISDC	**<u>Heme domain</u>:** ·$P450_{BM3}$ (Bacterium) ·CYP102A2 (Bacterium) ·CYP102A3 (Bacterium) ➔ *65% amino acid identity*	·45% correctly folded P450s ·Some more thermostable (by 7℃) than parental P450s ·When reconstituted with reductase domains, best-performing P450 towards 11 substrates was a chimera	Otey *et al.* (2006) Landwehr *et al.* (2007)

2.5 Goals of the Study

The goal of this research was to:

- Construct a self-sufficient terminal alkane hydroxylase by fusing the gene encoding the heme domain of a Class I P450, CYP153A6 from *Mycobacterium* sp. HXN-1500, to the gene fragment encoding the PFOR-type reductase domain of a Class VII P450, CYP116B3 from *Rhodococcus ruber* DSM 44319.
- Express and characterise the resulting fusion P450 using *E. coli* as the host organism.
- Amplify the internal fragments of *CYP153A* genes from environmental samples and use cassette PCR to replace the internal fragment of the *CYP153A6* gene with these fragments, with the aim of generating diverse self-sufficient terminal alkane hydroxylases that could serve as possible parent genes for family shuffling.
- Express and characterise the resulting chimeras using *E. coli* as the host organism.

Chapter 3

Materials and Methods

3.1 Bacterial Strains

Bacterial strains used in this study are listed in Table 3.1.

Table 3.1 *Escherichia coli* strains used in the study

Strain	Genotype	Source	Purpose
XL-10 Gold	Tet^r Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*^{q}*ZΔM15*Tn*10* (Tet^r) Amy Cam^r]*a*	Stratagene	Cloning Plasmid propagation
BL21(DE3)pLysE	F- *ompT hsdS*B (rB-mB-) *gal dcm* (DE3) pLysE (CamR)	Invitrogen	Expression
Rosetta-gami 2(DE3)pLysS	Δ(*ara-leu*)*7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rpsL* (DE3) F'[*lac*$^+$ *lacI*^{q} *pro*] *gor522*::Tn*10 trxB* pLysSRARE2 (Cam^R, Str^R, Tet^R)	Novagen	Expression

3.2 Plasmids used in the Study

The pCom8-PFR1500 plasmid (van Beilen *et al.*, 2005), a derivative of pCom8 (Smits *et al.*, 2001) containing the genes of the *CYP153A6* operon from *Mycobacterium* sp. strain HXN-1500 (Fig. 3.1), was originally obtained from Dr Jan B. van Beilen (Institute of Biotechnology, Zürich, Switzerland) transformed in *Pseudomonas putida* GPo12 (pGEc47ΔB, pCom8-PFR1500). After attempts to extract this plasmid were unsuccessful, it was reamplified by Dr Khajamohiddin Syed, a postdoctoral researcher in our laboratory at the time.

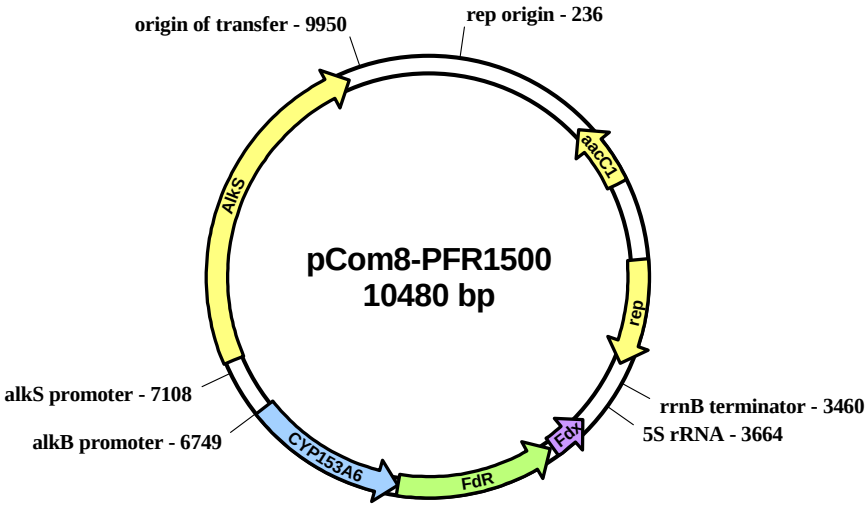


Figure 3.1 Schematic representation of pCom8-PFR1500. The genes encoding the P450, the ferredoxin reductase and the ferredoxin proteins of the CYP153A6 electron transport system are indicated in blue, green and purple, respectively.

Plasmids pET28a-CYP116B3 and pET28a-PFOR were provided by Dr Vlada B. Urlacher (University of Stuttgart, Stuttgart, Germany). The pET28a-CYP116B3 plasmid contains the DNA encoding CYP116B3 from *Rhodococcus ruber* DSM 44319, inserted between the *Eco*RI and *Hin*dIII restriction sites of the pET28a(+) plasmid (Fig. 3.2) (Liu *et al.*, 2006). The pET28a-PFOR plasmid contains the DNA encoding only the reductase domain (PFOR) and linker region of CYP116B3, inserted between the *Eco*RI and *Hin*dIII sites. In both cases, the insertion of the DNA was performed in such a way that the sequence encoding an N-terminal His-tag was included in the open-reading frame (ORF) (Fig. 3.3).

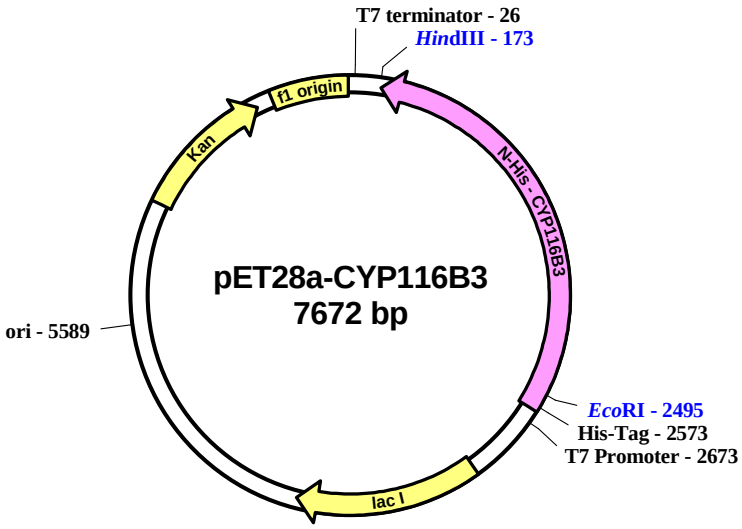


Figure 3.2 Schematic representation of pET28a-CYP116B3. The ORF of the His-tagged CYP116B3 protein is indicated in pink.

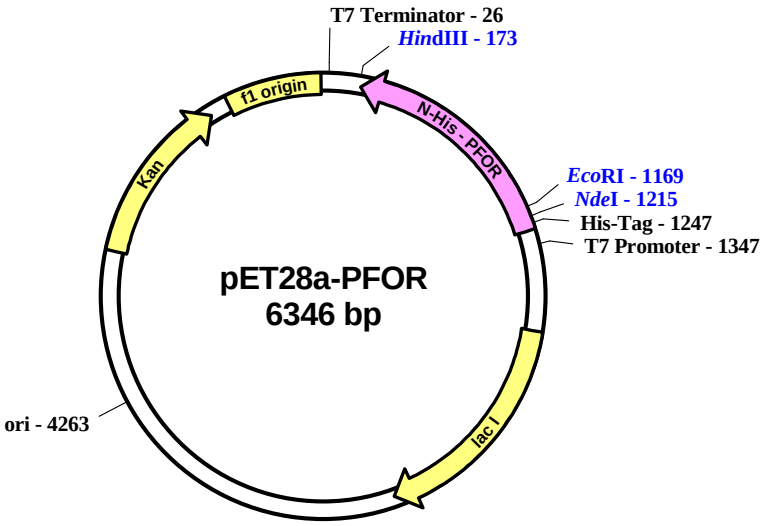


Figure 3.3 Schematic representation of pET28a-PFOR. The ORF of the His-tagged reductase (PFOR) domain of CYP116B3 is indicated in pink.

The pET28b(+) and the pET22b(+) plasmids used in the study were obtained from Novagen (Fig. 3.4).

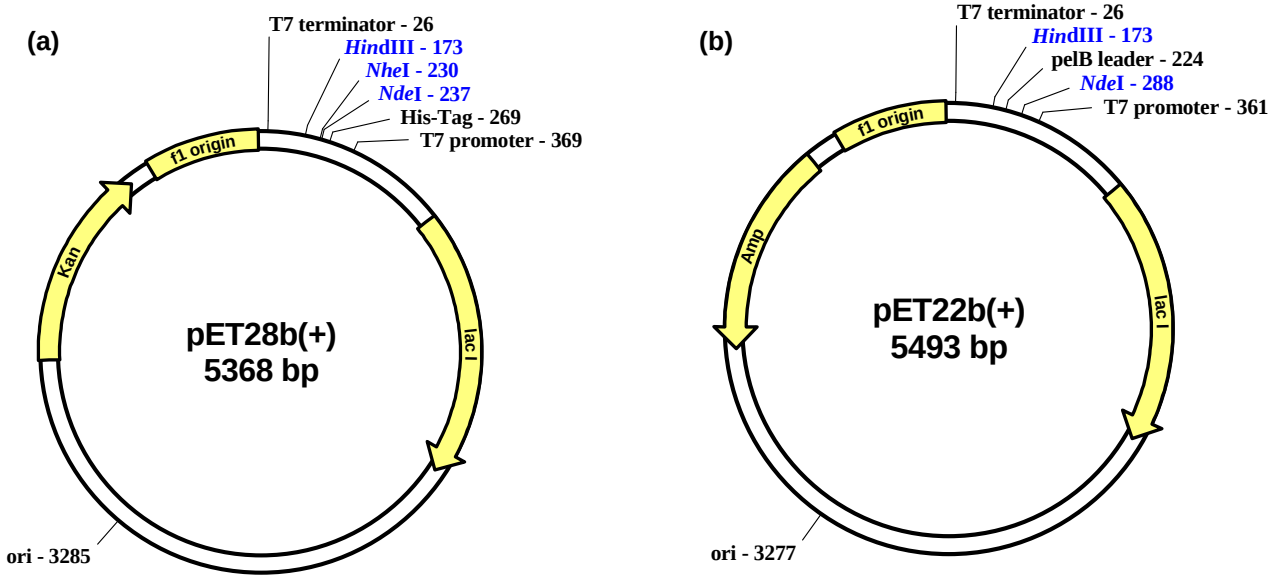


Figure 3.4 Schematic representations of **(a)** the pET28b(+) and **(b)** the pET22b(+) plasmids.

3.3 Chemicals, Enzymes and Consumables

Chemicals and antibiotics were obtained from Fluka, Merck, Sigma-Aldrich, Roche or Peqlab Biotechnologie GmbH, unless otherwise stated.

Restriction enzymes, T4 DNA Ligase, Quick Blunting Kit, dNTPs (deoxyribonucleoside triphosphates) and DNA extraction and purification kits were obtained from Fermentas, New England Biolabs and Separation Scientific.

Primers used in the study were obtained from Integrated DNA Technologies (IDT) or Bioneer (Table 3.2 and 3.3). All the designed primers were analysed using the OligoAnalyser algorithm from IDT (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/).

DNA and protein molecular weight markers were obtained from Fermentas (Fig. 3.5).

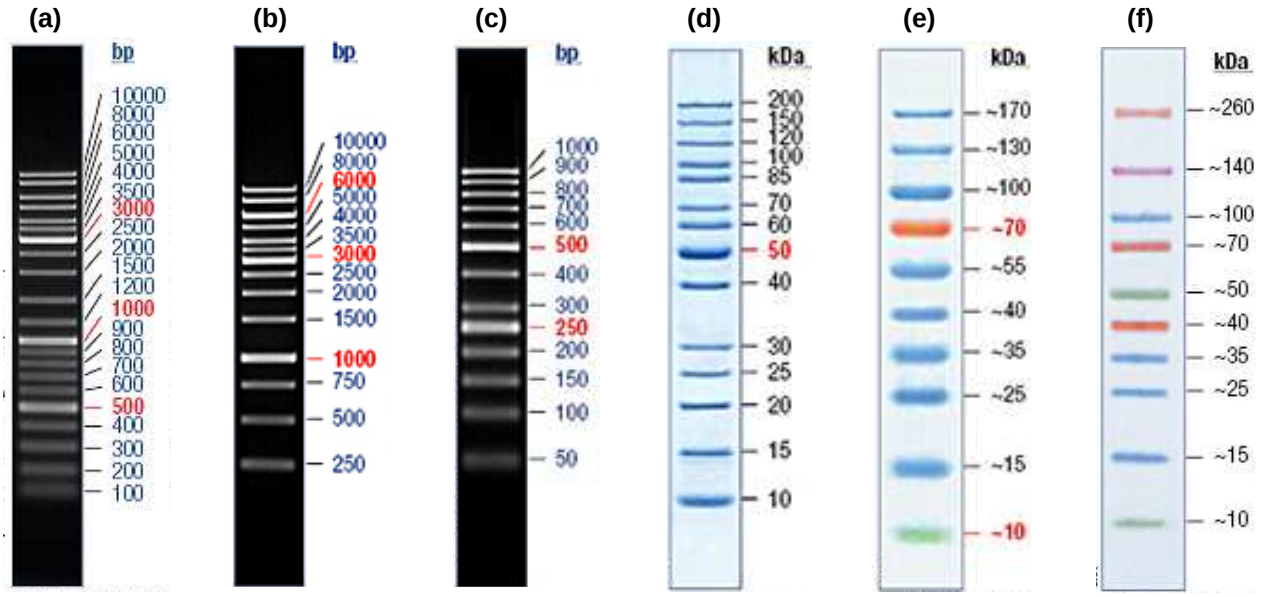


Figure 3.5 DNA and protein molecular weight markers. **(a)** GeneRuler™ DNA Ladder Mix; **(b)** GeneRuler™ 1 kb DNA Ladder; **(c)** GeneRuler™ 50 bp DNA Ladder; **(d)** PageRuler™ Unstained Protein Ladder; **(e)** PageRuler™ Prestained Protein Ladder; **(f)** Spectra™ Multicolour Broad Range Protein Ladder (http://www.fermentas.com/catalog/electrophoresis).

Table 3.2 Cloning primers used in the study. Introduced restriction sites are underlined and italicised. Degenerate bases are shown in bold.

Primer Name	Sequence (5' – 3')	Restriction Site
CF	ATGTT**Y**AT**H**GC**N**ATGGA**Y**CC**N**C [a]	
P450rv3	GCA**S**CGGTGGATGCCGAAGCC**R**AA [b]	
pCom-153F	CG*CATATG*GAAATGACGGTGGCCGCCAGCGAC	*Nde*I
pCom-153R	CG*GAATTC*GGCGTTGATGCGCACGGGCAG	*Eco*RI
5R	GNGGRTCCATNGCDATRAACATGGGAAGCGAT	
3F	TTYGGCTTCGGCATCCACCGSTGC	
HXN-1500-op-FW	CG*GCTAGC*GAAATGACGGTGGCCGCCAG [c]	*Nhe*I
HXN-1500-op-RV	CG*AAGCTT*CTAATGTTGTGCAGCTGGTGTCCGTACGACG [c]	*Hin*dIII
CYP116-pET-F	CG*CATATG*ATGAGTGCATCAGTTCCGGCGTC	*Nde*I
CYP116-pET-R	CG*AAGCTT*TCAGAGTCGGAGGGTCAGTC	*Hin*dIII
Red-pET-F	CG*CATATG*GTGCTGCAACGGCAG	
CYP116LF	ACCATCGGAGAACCCGCCGCCCGGTCGGTGTCACGCACC	
CYP116RF	GACCGGTTGATGCCGGTGCAGCACGAATTCGGATCCGCGA	

Notes: a = Kubota *et al.* (2005)
b = van Beilen *et al.* (2006)
c = van Beilen *et al.* (2005)
Degenerate bases: **H** = A/C/T; **N** = G/A/T/C; **Y** = C/T; **R** = A/G; **S** = C/G

3.4 Media and Growth Conditions

All *E. coli* strains were grown in Luria-Bertani (LB) media [composition: 0.5% (w/v) yeast extract, 1% (w/v) tryptone, 1% (w/v) NaCl; pH adjusted to 7.4 with 5 M NaOH] or on LB agar plates containing 1.5% (w/v) bacteriological agar (Sambrook *et al.*, 1989), supplemented with 30 $\mu g.mL^{-1}$ kanamycin or 100 $\mu g.mL^{-1}$ ampicillin and, in the case of expression hosts, 34 $\mu g.mL^{-1}$ chloramphenicol.

Cultures were cultivated at 30℃ on a rotary shaker at 130 rpm, unless otherwise stated. Plasmid-containing *E. coli* XL-10 Gold was cultivated at 37℃ on a rotary shaker at 160 rpm for 16 hours, in 5 mL LB media supplemented with 30 $\mu g.mL^{-1}$ kanamycin or 100 $\mu g.mL^{-1}$ ampicillin.

Table 3.3 Sequencing primers used in the study

Primer Name	Sequence (5' – 3')	Template DNA
CYP116-F1	GAATACGTCGACCGCTTCGTCGAA	*CYP116B3*
CYP116-F2	CGCAAAGCTGCTCGTCGTCAACG	*CYP116B3*
CYP116-R1	CGATCGAGTACTTGCGCATCGTC	*CYP116B3*
CYP116-R2	GGTCAGTTCCTCGAGGAAGATCTG	*CYP116B3*
CYP116red-F1	CACATCGATGTCGACCTCGGTGC	*PFOR*
CYP116red-R1	GAGGTGAAGTGCTCGACGTGCA	*PFOR*
CYP116red-R2	GATGGTGACGGGATGCTGCCGTTG	*PFOR*
seqCYP153-5R	GTCTCGACTGCCATGATG	*CYP153A6*
seqCYP153-3F	CAACAGACTCGCCGAACTA	*CYP153A6*
CYP153A6-F1	GCATCACCATCATGGACGACAAC	*CYP153A6*
CYP153A6-F2	CCCAAGAACGATCTCATCTCGATG	*CYP153A6*
CYP153A6-F3	ATCAACGCCTGATGATCCACAC	*CYP153A6*
CYP153A6-R1	CACAGGATGTTGAGCTGTAGTTCG	*CYP153A6*
CYP153A6-R2	CACGATGTTTCCGAGATATTCCTC	*CYP153A6*
FerRed-F1	CTATCAGCTTCGATCACCTAGTGC	*FdR*
FerRed-F2	GATAGGATTGCTCCCCAACGTCGA	*FdR*
FerRed-R1	CAAGACGGTGGTTTGTGCTTCGAG	*FdR*
FerRed-R2	CTCGAACCCTCTTTGTCGAAGGCT	*FdR*
FerRed-R3	ACTGTGGTTGCGCACCGAATC	*FdR*
FerRed-R4	GATCCTGTCTACGTCGGTCACTGT	*FdR*
Ferr-R1	GTGTAGTCGATGTAGGTGATCTTC	*Fdx*
Ferr-R2	AACCAGTCTGCATCGACGTGCACA	*Fdx*
SL1	CAGTCCAGTTACGCTGGAGTC	pSMART HC Kan insert
SR2	GGTCAGGTATGATTTAAATGGTCAGT	pSMART HC Kan insert
T7	TAATACGACTCACTATAGGG	pGem-T Easy insert
Sp6	TACGATTTAGGTGACACTATAG	pGem-T Easy insert
T7 Promoter	TAATACGACTCACTATAGGG	pET insert
T7 Terminator	GCTAGTTATTGCTCAGCGG	pET insert

3.5 General Molecular Techniques

3.5.1 PCR Amplification

PCR was performed with *Taq* DNA Polymerase (New England Biolabs), Kapa HiFi™ Polymerase (Kapa Biosystems) or Expand High Fidelity^{PLUS} PCR System (Roche) in standard 50 µL reactions consisting of the following: 1x reaction buffer containing $MgCl_2$ (Fidelity Buffer in the case of Kapa HiFi™, Vial 2 in the case of Expand High Fidelity^{PLUS}), 0.2 mM dNTP mix (0.3 mM in the case of Kapa HiFi™), 2 µM of each primer (0.3 µM in the case of Kapa HiFi™, 0.4 µM in the case of Expand High Fidelity^{PLUS}), approximately 100 ng template DNA, 1.25 U *Taq* DNA polymerase (0.5 U Kapa HiFi™ DNA Polymerase, 2.5 U Expand High Fidelity^{PLUS}) and deionised water. PCR was performed in a 2720 Thermal Cycler (Applied Biosystems) using standard PCR programs (Tables 3.4 to 3.6).

Table 3.4 Standard PCR program for *Taq* DNA Polymerase

Step	Temperature	Time	Number of cycles
Initial denaturation	94℃	2 min	1
Denaturation	94℃	45 sec	
Annealing	58 – 64℃ [a]	1 min	30
Elongation	72℃	30 sec – 1 min [b]	
Final elongation	72℃	7 min	1

Notes: a = Annealing temperature based on (T_m - 2℃)

b = Elongation time based on 1 min/kb

Table 3.5 Standard PCR program for Kapa HiFi™ DNA Polymerase

Step	Temperature	Time	Number of cycles
Initial denaturation	95℃	2 min	1
Denaturation	98℃	20 sec	
Annealing	68 – 87℃ [a]	15 sec	25
Elongation	72℃	1 min – 3 min 15 sec [b]	
Final elongation	72℃	5 min	1

Notes: a = Annealing temperature based on (T_m - 2℃)

b = Elongation time based on 30 sec/kb

Table 3.6 Standard PCR program for Expand High Fidelity^{PLUS} PCR System

Step	Temperature	Time	Number of cycles
Initial denaturation	94℃	2 min	1
Denaturation	94℃	20 sec	
Annealing	52 - 60℃ [a]	30 sec	30
Elongation	68 / 72℃ [b]	1 min 30 sec – 3 min [c]	
Final elongation	68 / 72℃ [b]	7 min	1

Notes: a = Annealing temperature based on (T_m - 5℃)

b = Elongation temperature: 72℃ for amplicons up to 3 kb, 68℃ for amplicons larger than 3 kb

c = Elongation time based on 30 sec/kb

3.5.2 Restriction Enzyme Digestions

Restriction enzyme digestions of plasmid DNA were carried out according to the manufacturer's specifications in volumes of 10 µL. In the case of double digests, the required ratios of enzymes and the recommended buffers for digestion were determined using Fermentas Double Digest (http://www.fermentas.com/doubledigest/index.html).

3.5.3 Visualisation and Purification of PCR and Restriction Enzyme Digestion Products

PCR amplicons and restriction enzyme digestions were electrophoresed on 0.8% (w/v) agarose gels, unless otherwise stated, containing GoldView Nucleic Acid Stain (SBS Genetech) in TAE (Tris-Acetate-EDTA) electrophoresis buffer [composition: 40 mM Tris, 2 mM EDTA, 20 mM glacial acetic acid; pH 8.5]. Gels were visualised with a ChemiDoc XRS (Bio-Rad Laboratories) or, in the case of DNA for cloning purposes, with a DarkReader^{TM} transilluminator (Bio-Rad Laboratories), facilitating the excision of the DNA from the gel. The DNA was purified using a BioSpin Gel Extraction Kit (Separation Scientific) as specified by the manufacturer.

3.5.4 Nucleic Acid Quantification

DNA was quantified with an Eppendorf Biophotometer.

3.5.5 Ligations

All ligation reactions were carried out with T4 DNA Ligase (Fermentas) according to the manufacturer's specifications, in a volume of 10 µL for 1 hour at room temperature, unless otherwise stated. A ratio of insert to plasmid backbone of at least 3:1 was used in all ligation reactions.

3.5.6 Cloning and *E. coli* Transformations

Blunt-ended PCR amplicons were phosphorylated using the Quick Blunting Kit (New England Biolabs) according to the manufacturer's instructions, before being ligated for 1 hour at room temperature to the pSMART HC Kan blunt cloning vector (Lucigen) for sub-cloning. PCR amplicons with A-overhangs were ligated overnight at 4℃ to the pGem-T Easy vector (Promega) for sub-cloning.

Competent *E. coli* cells were prepared in aliquots of 100 µL using the method of Inoue *et al.* (1990) and transformed according to the method described by Sambrook *et al.* (1989). The cells were plated on LB agar supplemented with 30 $\mu g.mL^{-1}$ kanamycin or 100 $\mu g.mL^{-1}$ ampicillin, and 34 $\mu g.mL^{-1}$ chloramphenicol where applicable (see section 3.4, p.37). Cells transformed with pGem-T Easy were plated on LB agar supplemented with 60 $\mu g.mL^{-1}$ ampicillin, 9.6 $\mu g.mL^{-1}$ IPTG (Isopropyl-β-D-1-thiogalactopyranoside) and 40 $\mu g.mL^{-1}$ X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). The plates were incubated at 37℃ for 16 hours. Positive transformants were selected and inoculated into 5 mL LB media supplemented with the appropriate antibiotic for plasmid propagation (see section 3.4, p.37).

3.5.7 Plasmid Extraction

Plasmid DNA was extracted with a BioSpin Plasmid DNA Extraction Kit (Separation Scientific), as instructed by the manufacturer. The size of plasmid inserts were confirmed by restriction enzyme analysis (see section 3.5.2, p.40).

3.5.8 Nucleotide Sequence Analyses

DNA sequencing was performed at the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, unless otherwise stated. Both strands of all plasmid inserts were sequenced using the BigDye™ Terminator Cycle Sequencing Kit v3.1 (Applied

Biosystems) according to the manufacturer's instructions. The primers used for sequencing had a final concentration of 3.2 pmol in a total volume of 10 µL. Plasmid DNA at a final concentration of 150 – 300 ng was used as the template for sequencing reactions. A standard PCR program was used for all the reactions (Table 3.7). Sequencing was performed with a 3130*xl* Genetic Analyzer (Applied Biosystems). pSMART HC Kan and pGem-T Easy inserts were sequenced using primers SL1 and SR2, and T7 and Sp6, respectively (Table 3.3, p.38). pET plasmid inserts were sequenced using the T7 Promoter and T7 Terminator primers. Internal sequencing primers were used where necessary to obtain sufficient sequence overlaps to allow contig assembly.

Nucleotide sequences were analysed using Vector NTI Advance 10 and Geneious Pro 4.7. Nucleotide sequences were compared with sequences in the database at NCBI using BLAST (http://www.ncbi.nlm.nih.gov/blast/BLAST.cgi). Amino acid sequence alignments were performed using DNAssist Version 3.0. Phylogenetic trees were generated with MEGA Version 4, using the neighbour-joining method. Plasmid diagrams were constructed using pDRAW Version 1.1.101 (Acaclone Software).

Table 3.7 Standard sequencing PCR program

Step	Temperature	Time	Number of cycles
Initial denaturation	96℃	1 min	1
Denaturation	96℃	10 sec	
Annealing	50℃	5 sec	25
Elongation	60℃	4 min	

3.6 Construction of the *CYP153A6/PFOR(CYP116B3)* Fusion

3.6.1 PCR Amplification of the *CYP153A6* Gene

The gene encoding the CYP153A6 heme domain was amplified without its stop codon from pCom8-PFR1500 with Kapa HiFi^{TM} DNA Polymerase, using primers pCom-153F and pCom-153R (Table 3.2, p.37). These primers introduced an *Nde*I and an *Eco*RI restriction site at the 5'- and 3'-end of the amplicon, respectively. A PCR program with an annealing temperature of 68℃ and an elongation time of 1 min was used (Table 3.5, p.39). The amplicon was purified (see section 3.5.3, p.40) and sub-cloned in pSMART HC Kan (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with

kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by *Eco*RI and *Pvu*I digestion.

3.6.2 Ligation of the *CYP153A6* Gene to pET28a-PFOR

The *CYP153A6* insert was excised from the pSMART HC Kan plasmid *via* double digestion with *Nde*I and *Eco*RI and was then ligated to pET28a-PFOR between the *Nde*I and *Eco*RI restriction sites, placing the *CYP153A6* gene at the 5'-end of the *CYP116B3* linker region. *E. coli* XL-10 Gold cells were transformed with the ligated plasmid and plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the size of the insert was confirmed by double digestion with *Nde*I and *Eco*RI. This construct is referred to as pET28a-CYP153A6/PFOR.

3.7 Cloning of the *CYP153A6* Operon

3.7.1 PCR Amplification of the *CYP153A6* Operon

The complete operon encoding CYP153A6, the ferredoxin reductase (FdR) and the ferredoxin (Fdx) was amplified as a single PCR product from pCom8-PFR1500 using Kapa HiFi^{TM} DNA Polymerase, with primers HXN-1500-op-FW and HXN-1500-op-RV (Table 3.2, p.37). These primers introduced an *Nhe*I and a *Hin*dIII restriction site at the 5'- and 3'-ends of the amplicon, respectively. A PCR program with an annealing temperature of 68℃ and elongation time of 2 min was used (Table 3.5, p.39). The amplicon was purified (see section 3.5.3, p.40) and sub-cloned in pSMART HC Kan (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by double digestion with *Nhe*I and *Hin*dIII.

3.7.2 Ligation of the *CYP153A6* Operon to pET28b(+)

The *CYP153A6* operon insert was excised from the pSMART HC Kan plasmid *via* double digestion with *Nhe*I and *Hin*dIII and was ligated to the pET28b(+) plasmid between the *Nhe*I and *Hin*dIII restriction sites. *E. coli* XL-10 Gold cells were transformed with the plasmid and plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by

double digestion with *Nhe*I and *Hin*dIII. This construct is referred to as pET28b-CYP153A6_FdR_Fdx.

3.8 Creation of pET22b constructs

3.8.1 pET22b-CYP153A6/PFOR

The *CYP153A6/PFOR(CYP116B3)* fusion gene insert was excised from pET28a-CYP153A6/PFOR *via* double digestion with *Nde*I and *Hin*dIII and was ligated to the pET22b(+) plasmid between the *Nde*I and *Hin*dIII restriction sites. *E. coli* XL-10 Gold cells were transformed with the ligated plasmid and plated on LB agar supplemented with ampicillin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by double digestion with *Nde*I and *Hin*dIII. This construct is referred to as pET22b-CYP153A6/PFOR.

3.8.2 pET22b-CYP116B3 and pET22b-PFOR

3.8.2.1 PCR Amplification of the *CYP116B3* Gene and the DNA Encoding the CYP116B3 PFOR Domain

The *CYP116B3* gene was PCR amplified from pET28a-CYP116B3 with the Expand High Fidelity^{PLUS} PCR System using the primers CYP116-pET-F and CYP116-pET-R (Table 3.2, p.37). These primers introduced an *Nde*I site and a *Hin*dIII restriction site at the 5'- and 3'-ends of the amplicon, respectively. A PCR program with an annealing temperature of 60℃, an elongation temperature of 72℃ and an elongation time of 1 min 30 sec was used (Table 3.6, p.40).

The DNA encoding the PFOR domain was also amplified with the Expand High Fidelity^{PLUS} PCR System, using the above-mentioned PCR program. The primers used for this reaction were Red-pET-F and CYP116-pET-R (Table 3.2, p.37) and introduced an *Nde*I and a *Hin*dIII restriction site at the 5'- and 3'-ends of the amplicon, respectively.

The amplicons were purified (see section 3.5.3, p.40) and sub-cloned using pGem-T Easy (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with ampicillin, IPTG and X-gal (see section 3.5.6, p.41). Plasmid DNA was

extracted from positive transformants (see section 3.5.7, p.41) and the insert sizes were confirmed by double digestion with *Nde*I and *Hin*dIII.

3.8.2.2 Ligation of the *CYP116B3* Gene and the PFOR DNA to pET22b(+)

The inserts were excised from pGem-T Easy *via* double digestion with *Nde*I and *Hin*dIII and were ligated to the pET22b(+) plasmid between the *Nde*I and *Hin*dIII restriction sites. *E. coli* XL-10 Gold cells were transformed with the plasmid and plated on LB agar supplemented with ampicillin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by double digestion with *Nde*I and *Hin*dIII. The resulting constructs are referred to as pET22b-CYP116B3 and pET22b-PFOR.

3.8.3 pET22b-CYP153A6_FdR_Fdx

3.8.3.1 PCR Amplification of the *CYP153A6* Operon

The *CYP153A6* operon was amplified as a single PCR product from pET28a-CYP153A6_FdR_Fdx using Kapa HiFi^{TM} DNA Polymerase, with primers pCom-153F and HXN-1500-op-RV (Table 3.2, p.37). These primers introduced an *Nde*I and a *Hin*dIII restriction site at the 5'- and 3'-ends of the amplicon, respectively. A PCR program with an annealing temperature of 68℃ and elongation time of 2 min was used (Table 3.5, p.39). The amplicon was purified (see section 3.5.3, p.40) and sub-cloned using pGem-T Easy (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with ampicillin, IPTG and X-gal (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by double digestion with *Nde*I and *Hin*dIII.

3.8.3.2 Ligation of the *CYP153A6* Operon to pET22b(+)

The *CYP153A6* operon insert was excised from pGem-T Easy *via* digestion with *Nde*I, *Hin*dIII and *Ssp*I and was ligated to the pET22b(+) plasmid between the *Nde*I and *Hin*dIII restriction sites. *E. coli* XL-10 Gold cells were transformed with the plasmid and plated on LB agar supplemented with ampicillin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by double digestion with *Nde*I and *Hin*dIII. This construct is referred to as pET22b-CYP153A6_FdR_Fdx.

3.9 Site-directed Mutagenesis of the CYP116B3 PFOR Domain Linker

3.9.1 PCR Amplification of pET28a-CYP153A6/PFOR

The pET28a-CYP153A6/PFOR plasmid was amplified with Kapa HiFi^{TM} polymerase using primers CYP116LF and CYP116RF (Table 3.2, p.37). These primers were designed according to the desired PFOR linker sequence. A PCR program with an annealing temperature of 87℃ and an elongation time of 3 min 15 sec was used for this reaction (Table 3.5, p.39). The amplicon was purified (see section 3.5.3, p.40), treated with *Dpn*I (see section 3.5.2, p.40) and phosphorylated using the Quick Blunting Kit (see section 3.5.6, p.41). The amplicon was ligated to itself (see section 3.5.5, p.41) and used to transform *E. coli* XL-10 Gold cells. The cells were plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the size of the plasmid was confirmed *via Eco*RI digestion.

3.9.2 Replacement of the DNA Encoding the CYP116B3 PFOR Domain

The DNA encoding the CYP116B3 PFOR domain was excised from the amplified plasmid *via* double digestion with *Eco*RI and *Not*I. This fragment was used to replace the DNA encoding the original PFOR domain in the pET28-CYP153A6/PFOR plasmid, which was excised using *Eco*RI and *Not*I. Following ligation, the plasmid was used to transform *E. coli* XL-10 Gold cells which were then plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the *CYP153A6/PFOR*(*CYP116B3*) insert size confirmed by double digestion with *Nde*I and *Hin*dIII. This construct is referred to as pET28b-CYP153A6/Linker/PFOR.

3.10 Amplification of the Internal Fragments of *CYP153A* Genes from Environmental DNA

During a B.Sc. Honours research project preceding this study, total DNA was extracted from hexadecane enrichments of diesel-contaminated soil that was sampled at a petroleum plant in Aliwal-North (Eastern Cape, South Africa). The DNA was extracted using the phenol/chloroform method described by Sambrook *et al*. (1989) and was stored at -20℃.

For this study, the environmental DNA served as the template for the PCR amplification of the internal fragments of genes encoding P450 proteins from the CYP153A subfamily. The degenerate primers CF and P450rv3 (Table 3.2, p.37) were used for this reaction. PCR was carried out with *Taq* DNA polymerase, using a PCR program with an annealing temperature of 58℃ and an elongation time of 1 min (Table 3.4, p. 39). The amplicon was purified (see section 3.5.3, p.40) and sub-cloned using pGem-T Easy (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with ampicillin, IPTG and X-gal (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by *Eco*RI digestion. Correct plasmids were digested with *Hae*III, a four-base cutter, in order to obtain different restriction digestion profiles. The inserts of selected plasmids representing each profile were sequenced by Inqaba Biotec (South Africa).

3.11 Cassette PCR

3.11.1 First PCR: Amplification of the Internal *CYP153A* Gene Fragments

The *CYP153A* fragments from selected pGem-T Easy plasmids (see section 3.10) were reamplified using the degenerate primers CF and P450rv3 and *Taq* DNA polymerase. As mentioned in section 3.10, a PCR program with an annealing temperature of 58℃ and an elongation time of 1 min was used. The amplicons were purified (see section 3.5.3, p.40) and stored at -20℃ until required.

Additional pGem-T Easy plasmids containing inserts of gene fragments belonging to the CYP153A subfamily were donated by Dr Nathlee Abbai, who was a PhD student in our department at the time. These clones were amplified with the primers CF and P450rv3 from a metagenomic library constructed from samples obtained from the Beatrix Goldmine (Free State, South Africa), 3.2 km below the surface (Abbai, 2009). The inserts of selected plasmids were amplified using these primers and the PCR program mentioned above. The resulting amplicons were purified and stored at -20℃ until required.

3.11.2 First PCR: Amplification of the 5'- and 3'-ends of the *CYP153A6* Gene

The 5'- and 3'-ends of the *CYP153A6* gene were amplified by PCR from pET28a-CYP153A6/PFOR using *Taq* DNA polymerase and a PCR program with an annealing

temperature of 64°C and an elongation time of 30 sec (Table 3.4, p.39). The 5'-end was amplified using primers pCom-153F and 5R, whereas the 3'-end was amplified using primers 3F and pCom-153R (Table 3.2, p.37). The amplicons were purified (see section 3.5.3, p.40) and stored at -20°C until required.

3.11.3 Second PCR: Assembly of the Chimeric *CYP153A* Genes

Approximately 100 ng each of the internal *CYP153A* gene fragments and the 5'- and 3'-end fragments were mixed and served as the template for assembly of full-length genes by PCR. This was done using the Expand High Fidelity^{PLUS} PCR System with primers pCom-153F and pCom-153R. The PCR program had an annealing temperature of 52°C, an elongation temperature of 72°C and an elongation time of 1.5 min (Table 3.6, p.40). The amplicon was purified (see section 3.5.3, p.40) and sub-cloned in pGem-T Easy (see section 3.5.6, p.41). Following transformation, the *E. coli* XL-10 Gold cells were plated on LB agar supplemented with ampicillin, IPTG and X-gal (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants (see section 3.5.7, p.41) and the insert size was confirmed by *Eco*RI digestion.

3.11.4 Replacement of the *CYP153A6* Gene with the Chimeric *CYP153A* Genes

The pET28b-CYP153A6_FdR_Fdx plasmid was digested with *Nde*I and *Pst*I to remove the *CYP153A6* gene. The chimeric genes were excised from pGem-T Easy using *Nde*I and *Pst*I and were ligated to the digested pET28b-CYP153A6_FdR_Fdx plasmid. *E. coli* XL-10 Gold cells were transformed with the ligated plasmid and were plated on LB agar supplemented with kanamycin (see section 3.5.6, p.41). Plasmid DNA was extracted from positive transformants and the insert size was confirmed by digestion with *Nde*I and *Pst*I.

3.12 Biochemical Methods

3.12.1 Protein Expression

All the constructs described in sections 3.6 to 3.11 were used to transform competent *E. coli* BL21(DE3)pLysE cells for expression and characterisation.

Pre-cultures were cultivated at 37°C on a rotary shaker at 160 rpm in 5 mL LB media supplemented with the appropriate antibiotics (see section 3.4, p.37) to an OD_{600} of 0.6 – 0.8

and then kept overnight at 4℃. A 2% (v/v) pre-culture inoculum was used to inoculate 100 mL LB media supplemented with the appropriate antibiotics (see section 3.4, p.37), in 500-mL Erlenmeyer shake flasks. Cultivation was carried out on a rotary shaker at 130 rpm at 30℃ to an OD_{600} of 0.6 – 0.8. Expression was induced by the addition of 0.5 mM IPTG. The culture was supplemented with 0.5 mM δ-ALA (5-Aminolevulinic acid hydrochloride), a heme precursor, and 100 mg $FeCl_3$ to compensate for the lack of iron in *E. coli*. The cultures were transferred to a shaker at 20℃ and incubated for 16 hours before the cells were harvested by centrifugation at 4000 x *g* for 10 min at 4℃ using a Megafuge 1.0R (Heraeus Instruments). The pellets were resuspended in 50 mM potassium phosphate buffer (pH 7.4) containing 0.1 mM EDTA and 0.1 mM PMSF (Phenylmethylsulphonyl fluoride).

3.12.2 Cell Disruption

The cells were disrupted using a Bandelin Sonopuls sonicator (1 min, 80% power, 20% duty cycle) or a One Shot Cell Disrupter System (Constant Cell Disruption Systems) at a pressure of 30 kpsi. The soluble and insoluble fractions were separated by centrifugation at 14 000 x *g* for 20 min at 4℃ using an Eppendorf Centrifuge 5417R.

3.12.3 Protein Quantification and SDS-PAGE Analyses

Protein quantification was carried out using Bradford Reagent (Sigma-Aldrich) according to the manufacturer's protocol, in a 96-well microtiter plate. The absorbance at 595 nm was recorded with a SpectraMax M2 Microtiter Plate Reader (Molecular Devices). Bovine serum albumin (BSA) was used as the standard.

Protein expression in crude cell lysates and in the soluble and insoluble fractions was analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) on 8% resolving gels, using the Mini-PROTEAN® 3 system (Bio-Rad Laboratories).

To prepare the crude cell lysate samples, 1 mL of culture was pelleted by centrifugation for 1 min at 13 200 x *g* (Eppendorf Centrifuge 5415 D). The pellet was resuspended in 100 µL 50 mM potassium phosphate buffer (pH 7.4) containing 0.1 mM EDTA and 0.1 mM PMSF, and was diluted 1:2 by the addition of SDS Reducing Buffer [composition per 10 mL volume: 3.55 mL deionised water, 1.25 mL 0.5M Tris-HCl (pH 6.8), 2.5 mL 100% glycerol, 2 mL 10% (w/v) SDS, 0.2 mL 0.5% (w/v) bromophenol blue and 500 µL β-Mercaptoethanol]. Each sample was then heated at 95℃ for 4 min and stored at -20℃ until required. The soluble and insoluble fractions

were also prepared using this method, with a few modifications: in the case of the soluble fractions a 100 µL sample was diluted directly with the SDS Reducing Buffer prior to heating, whereas with the insoluble fractions the pellet was first resuspended in 50 mM potassium phosphate buffer (pH 7.4) containing 0.1 mM EDTA and 0.1 mM PMSF before being diluted with the SDS Reducing Buffer.

The gels were stained and destained using a modified version of the method described by Fairbanks *et al*. (1971), facilitating the visualisation of the expressed proteins.

3.12.4 Spectroscopic Characterisation and Enzyme Quantification

Carbon monoxide-difference (CO-difference) spectra were performed using both whole cells and the soluble fraction of disrupted cells, in 96-well microtiter plates. Four millilitres of induced culture (see section 3.12.1, pp.48-49) was centrifuged for 1 min at 13 200 x *g* (Eppendorf Centrifuge 5415 D) and the pellet was resuspended in 220 µL 50 mM potassium phosphate buffer (pH 7.4) containing 0.1 mM EDTA and 0.1 mM PMSF. An aliquot (200 µL) of this cell suspension or of the soluble fraction was transferred to the microtiter plate and was reduced with sodium dithionite. A wavelength scan between 400 nm and 500 nm was conducted and the resulting spectrum was recorded. The plate was then placed for 5 min in a sealed container held under positive carbon monoxide pressure. The spectrum obtained between 400 nm and 500 nm was once again recorded and the difference between spectra of the reduced and the reduced with carbon monoxide samples was calculated. The P450 and P420 content was quantified (in units of µM) using the following equations, with extinction coefficients (ε) of 91 $mM^{-1}cm^{-1}$ and 110 $mM^{-1}cm^{-1}$, respectively (Omura & Sato, 1964) and pathlength (l) of 0.596 cm:

$$[P450] = \frac{[A_{450\ nm} - A_{490\ nm}] \times 1000}{\varepsilon . l} \qquad [P420] = \frac{[A_{420\ nm} - A_{490\ nm}] \times 1000}{\varepsilon . l}$$

3.12.5 Reductase Activity Determination

Cytochrome c reductase assays were performed to determine the activity of the reductase domain(s). These assays were carried out at 30℃ in 96-well microtiter plates, in a total volume of 200 µL consisting of the following: 50 mM Tris-HCl (pH 7.5), 3.3 mM KCN, 625 µg/mL cytochrome c, 5 µL soluble fraction and deionised water. NADPH or NADH was added to each test sample to a final concentration of 0.12 mM, but was excluded from the control sample. The increase in absorbance at 550 nm was then monitored for a period of 5 min, after which the

reductase activity was calculated using an extinction coefficient of 21 $mM^{-1}cm^{-1}$ and pathlength of 0.596 cm.

3.12.6 Whole-cell Octane Bioconversions

Cultivation of cultures for whole-cell bioconversions was carried out according to the method used for protein expression (see section 3.12.1, pp.48-49). After the cells were harvested, the pellet from each 100-mL culture (approximately 0.9 g wet weight) was resuspended in 5 mL 0.2 M sodium phosphate buffer (pH 7.2) containing 0.8% (v/v) glycerol and 100 $\mu g.mL^{-1}$ $FeSO_4$. This mixture was known as the bioconversion reaction mixture. One-millilitre volumes of the bioconversion reaction mixture were transferred to 40-mL amber vials with screw caps (Chromatography Research Supplies) and glucose was added to a final concentration of 40 mM. A volume of 333 µL of *n*-octane was added to each vial, after which the vials were incubated on a shaker at 20℃ at an angle of 60°. Individual vials were harvested after specific time intervals and the samples were extracted for gas chromatography (GC) analysis.

3.12.7 Sample Extraction and GC Analysis

Whole-cell CO-difference spectra was performed using 200 µL of the bioconversion reaction mixture from each vial. The remaining reaction mixture was acidified by the addition of 200 µL 1 M HCl, followed by the addition of 500 µL ethyl acetate containing 0.2 mM 2-Nonanol or 1-Undecanol as the internal standard. The mixture was vortexed for 30 sec and centrifuged for 2 min at 13 200 x *g* (Eppendorf Centrifuge 5414 D). Fifty microlitres of the upper organic layer was transferred to GC vials for analysis.

GC analysis was performed on 1-µL samples using a Hewlett-Packard 5890 Series II gas chromatograph equipped with a flame ionisation detector (FID) and a 30 m x 0.53 mm x 1 µm Chrompack® CP-Wax 52 CB column. The carrier gas, hydrogen (H_2), was at a pressure of 75 kPa, with a split ratio of 1:40. The inlet temperature was 200℃. The initial oven temperature was 100℃, held for 1 min. The temperature was then increased at a rate of 10℃/min to 250℃, which was held for 11 min. The FID temperature was 280℃.

Chapter 4

Results

4.1 Construction of the *CYP153A6/PFOR(CYP116B3)* Fusion

The first aim of this study was to construct a self-sufficient terminal alkane hydroxylase by fusing the heme domain of CYP153A6, a terminal alkane hydroxylase from *Mycobacterium* sp. HXN-1500, to the reductase domain (PFOR) of CYP116B3, a self-sufficient P450 from *Rhodococcus ruber* DSM 44319. A similar fusion was constructed by Nodate and co-workers between the heme domain of $P450_{balk}$ (CYP153A13a), a terminal alkane hydroxylase from *Alcanivorax borkumensis* SK2, and the PFOR domain of P450RhF (CYP116B3), a self-sufficient P450 from *Rhodococcus* sp. NCIMB 9784. The resulting $P450_{balk}$/PFOR(CYP116B3) fusion protein was a self-sufficient terminal alkane hydroxylase (Nodate *et al.*, 2006).

In order to construct the CYP153A6/PFOR(CYP116B3) fusion, the gene encoding the CYP153A6 heme domain was amplified, without its stop codon, from pCom8-PFR1500 using primers pCom-153F and pCom-153R. These primers introduced an *Nde*I and an *Eco*RI restriction site at the 5'- and 3'-end of the amplicon, respectively. The primers were designed according to the 5'- and 3'-end of the *CYP153A6* gene in pCom8-PFR1500 which is missing the first nine nucleotides compared to the Genbank nucleotide sequence (accession no. AJ783967). However, the three amino acids encoded by these nucleotides are not essential for the functioning of the protein as all the reported characterisation of the CYP153A6 enzyme has been done using this shortened gene expressed in *Pseudomonas putida* (van Beilen *et al.*, 2005; Funhoff *et al.*, 2006 & 2007).

The expected size of the amplicon was 1 270 bp and a product of the appropriate size was obtained (Fig. 4.1), which was then sub-cloned using the blunt cloning vector pSMART HC Kan. Plasmid DNA extracted from a positive transformant was digested with *Eco*RI, which cleaves pSMART HC Kan on both sides of the insert, generating fragments of approximately 1 800 bp and 1 300 bp. Fragments corresponding to these sizes were obtained (Fig. 4.2, Lane 1). *Pvu*I digestion was utilised for further confirmation. *Pvu*I cuts once in the pSMART HC Kan backbone

and three times in the insert. DNA fragments of the expected sizes of approximately 1 400 bp, 1 000 bp, 600 bp and 100 bp were obtained (Fig 4.2, Lane 2). The 100-bp fragment, however, cannot be seen in Fig. 4.2 as a result of its size.

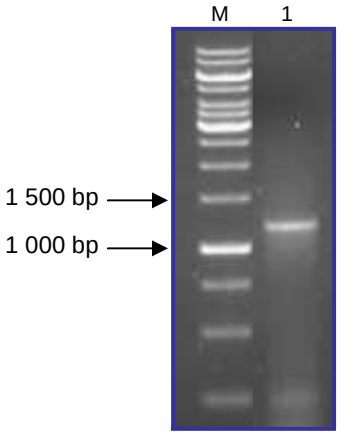


Figure 4.1 Amplification of the *CYP153A6* gene using primers pCom-153F and pCom-153R. 0.8% (w/v) agarose gel stained with GoldView Nucleic Acid Stain showing the PCR product (Lane 1). M = GeneRuler^{TM} 1 kb DNA Ladder.

The plasmid insert was sequenced using primers SL1 and SR2 and the appropriate internal primers listed in Table 3.3 (Chapter 3, p.38). The insert was excised *via* double digestion with *Nde*I and *Eco*RI and was ligated to pET28a-PFOR between the *Nde*I and *Eco*RI restriction sites, placing the *CYP153A6* gene at the 5'-end of the DNA encoding the CYP116B3 PFOR domain. The *CYP153A6* gene was inserted in frame with both the DNA encoding the CYP116B3 PFOR domain and the N-terminal His-tag sequence forming part of the ORF. The resulting construct was pET28a-CYP153A6/PFOR (Fig. 4.3).

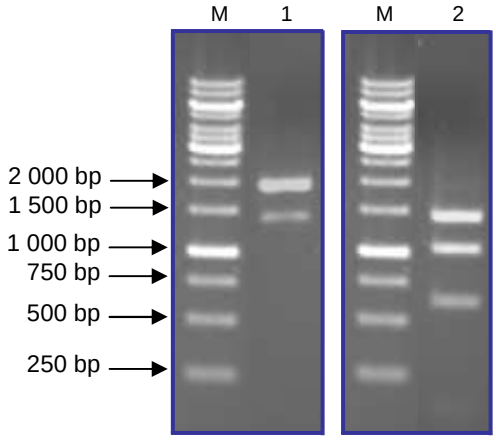


Figure 4.2 Restriction enzyme digestion of pSMART HC Kan containing the *CYP153A6* gene. 0.8% (w/v) agarose gel stained with GoldView Nucleic Acid Stain, showing the DNA fragments obtained using *Eco*RI (Lane 1) and *Pvu*I (Lane 2). M = GeneRuler^{TM} 1 kb DNA Ladder.

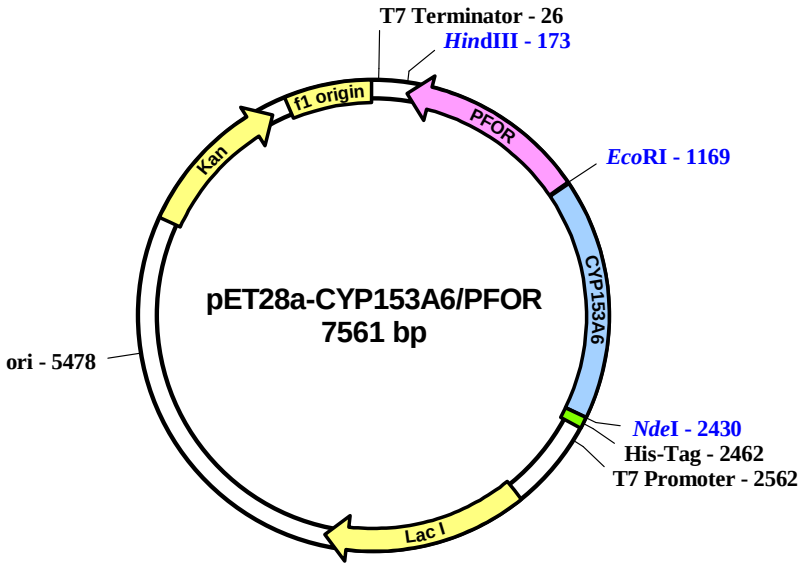


Figure 4.3 Schematic representation of pET28a-CYP153A6/PFOR. The gene encoding the CYP153A6 heme domain is shown in blue and the DNA encoding the PFOR domain of CYP116B3 is shown in pink. The green region represents the upstream region between the start codon and the first basepair of the *CYP153A6* gene, containing the N-terminal His-tag coding sequence. The complete ORF is represented by the green, blue and pink regions combined.

4.2 Cloning of the *CYP153A6* Operon

Prior to 2009, all the work published on CYP153A6 was carried out with *Pseudomonas putida* as the expression host. In order to confirm that the *CYP153A6* gene could be expressed in *E. coli* and that it could be expressed using a pET28 plasmid, and also to generate a control for the heme domain of the CYP153A6/PFOR(CYP116B3) fusion protein, the genes encoding the heme domain, the ferredoxin reductase (FdR) and the ferredoxin (Fdx) of the CYP153A6 electron transport system were amplified as a single amplicon from pCom8-PFR1500. The primers HXN-1500-op-FW and HXN-1500-op-RV were used for this reaction and introduced an *Nhe*I and a *Hin*dIII restriction site at the 5'- and 3'-end of the amplicon, respectively. With the exception of the introduced restriction sites, these primers were identical to the primers used by van Beilen *et al.* (2005) for the construction of pCom8-PFR1500.

An amplicon of approximately 3 000 bp was expected and a product of the appropriate size was obtained (Fig. 4.4). This was sub-cloned using pSMART HC Kan. Plasmid DNA was extracted from a positive transformant and the insert was sequenced using primers SL1 and SR2 and the appropriate internal primers listed in Table 3.3 (Chapter 3, p.38). The insert was excised from

the plasmid *via* double digestion with *Nhe*I and *Hin*dIII and was ligated between these restriction sites in the pET28b(+) plasmid, resulting in construct pET28b-CYP153A6_FdR_Fdx (Fig. 4.5).

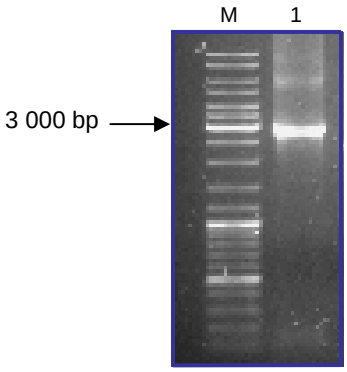


Figure 4.4 Amplification of the *CYP153A6* operon using primers HXN-1500-op-FW and HXN-1500-op-RV. 0.8% (w/v) agarose gel stained with GoldView Nucleic Acid Stain showing the PCR product (Lane 1). M = GeneRuler™ 1 kb DNA Ladder.

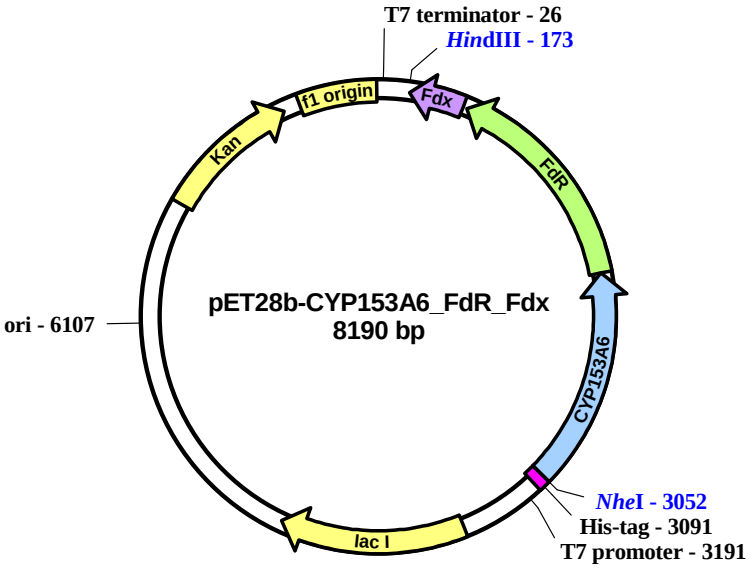


Figure 4.5 Schematic representation of pET28b-CYP153A6_FdR_Fdx. The genes encoding the heme domain, the ferredoxin reductase and the ferredoxin are shown in blue, green and purple, respectively. The pink area represents the upstream region between the start codon and the first basepair of the *CYP153A6* gene, containing the N-terminal His-tag coding sequence. The CYP153A6 ORF is represented by the pink and blue regions combined.

4.3 Sequence Analyses of the *CYP153A6* Genes

Analyses of the nucleotide sequences of the *CYP153A6* genes in both pET28a-CYP153A6/PFOR and pET28b-CYP153A6_FdR_Fdx revealed a mutation resulting in the replacement of a leucine residue with a proline residue within the conserved amino acid sequence of the heme-binding region in the CYP153A subfamily. This mutation could possibly have been incorporated during the reamplification of the pCom8-PFR1500 plasmid. In P450s belonging to this subfamily, this residue is either a valine or a leucine (Fig 4.6). However, proline and leucine may have similar properties and a proline residue in this position does not destroy the activity of CYP153A6 (see section 4.4), even though the heme-binding cysteine is only five residues away.

Figure 4.6 Section of the alignment of selected CYP153A amino acid sequences. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green. The residue indicated by the blue arrow is usually a leucine or a valine residue in this P450 subfamily. The absolutely conserved cysteine residue involved in heme-binding is indicated by the red arrow. “CYP153A6(G)” refers to the Genbank CYP153A6 sequence and “cyp153a6” represents the CYP153A6 heme domain sequence of the CYP153A6/PFOR(CYP116B3) fusion protein and the cloned CYP153A6.

4.4 Analyses of the Expressed Proteins

4.4.1 Bradford Assay

Protein quantification of all the soluble and insoluble fractions was done in triplicate using Bradford Reagent. Bovine serum albumin was used to construct a standard curve.

4.4.2 SDS-PAGE Analyses

Expression of the proteins was analysed by investigating the crude cell lysates and the soluble and insoluble fractions *via* SDS-PAGE on 8% resolving gels. In the case of the soluble and insoluble fractions, approximately 50 µg of protein was loaded in each well.

The CYP153A6/PFOR(CYP116B3) fusion product, consisting of 770 amino acids including an N-terminal His-tag, had an expected molecular weight of approximately 90 kDa and a band corresponding to this size was observed in the crude cell lysate (Fig. 4.7(a), Lane 1). It was expected that the fusion would be expressed in the soluble fraction because both CYP153A6 and CYP116B3 are soluble. However, a band corresponding to the expected size was only observed in the insoluble fraction (Fig. 4.7(b), Lane 2), indicating the formation of inclusion bodies. In the case of the controls, CYP116B3 (expected molecular weight of approximately 90 kDa) and PFOR (expected molecular weight of approximately 40 kDa), bands of the appropriate sizes were observed in the crude cell lysates (Fig. 4.7(a) Lanes 2 and 3) and in both the soluble and insoluble fractions (Fig. 4.7(b), Lanes 3 to 6).

In the case of CYP153A6, a band corresponding to the expected molecular weight of 50 kDa for the heme domain, consisting of 439 amino acids including an N-terminal His-tag, was observed in the crude cell lysate (Fig. 4.8(a), Lane 2). This band was also observed in both the soluble and insoluble fractions (Fig. 4.8(b), Lanes 3 and 4). Bands corresponding to the expected molecular weights of 45 kDa for the ferredoxin reductase (424 amino acids) and 11 kDa for the ferredoxin (106 amino acids), both without N-terminal His-tags, could not be observed by SDS-PAGE analysis of crude cell lysates or soluble and insoluble fractions.

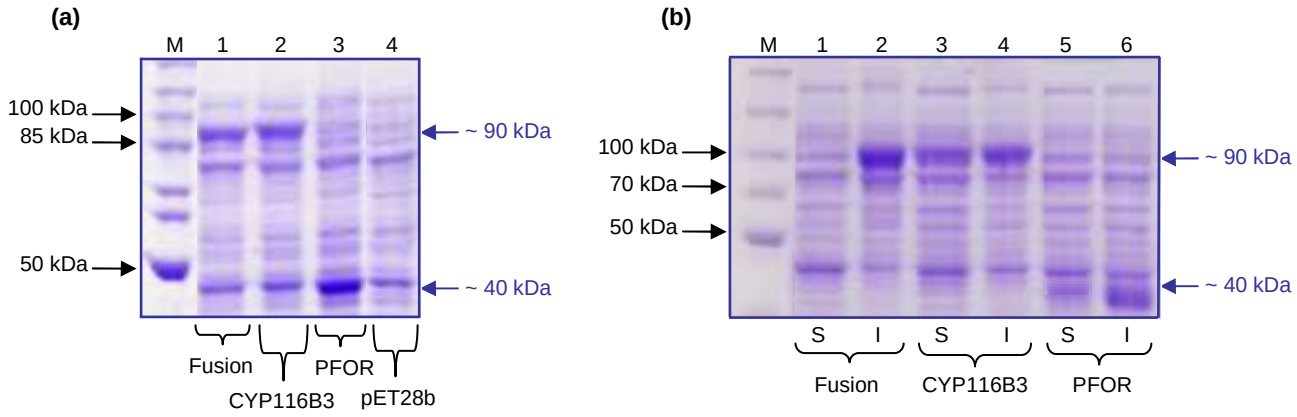


Figure 4.7 SDS-PAGE of **(a)** the crude cell lysates and **(b)** the soluble and insoluble fractions of *E. coli* BL21(DE3)pLysE strains transformed with pET28 plasmids for expression of the CYP153A6/PFOR(CYP116B3) fusion, CYP116B3 and the CYP116B3 PFOR domain. **(a)** Lane 1, CYP153A6/PFOR(CYP116B3) (~ 90 kDa); Lane 2, CYP116B3 (~ 90 kDa); Lane 3, CYP116B3 PFOR domain (~ 40 kDa); Lane 4, empty pET28b plasmid. M = PageRuler™ Protein Ladder. **(b)** Soluble fractions are indicated by "S" and insoluble fractions are indicated by "I". Lanes 1 and 2, CYP153A6/PFOR(CYP116B3); Lanes 3 and 4, CYP116B3; Lanes 5 and 6, CYP116B3 PFOR domain. M = Spectra™ Multicolour Broad Range Protein Ladder.

The expression observed for the above-mentioned proteins is summarised in Table 4.1 (p.60), using positive (+) signs to indicate the level of positive expression and a negative (-) sign to indicate no observed expression.

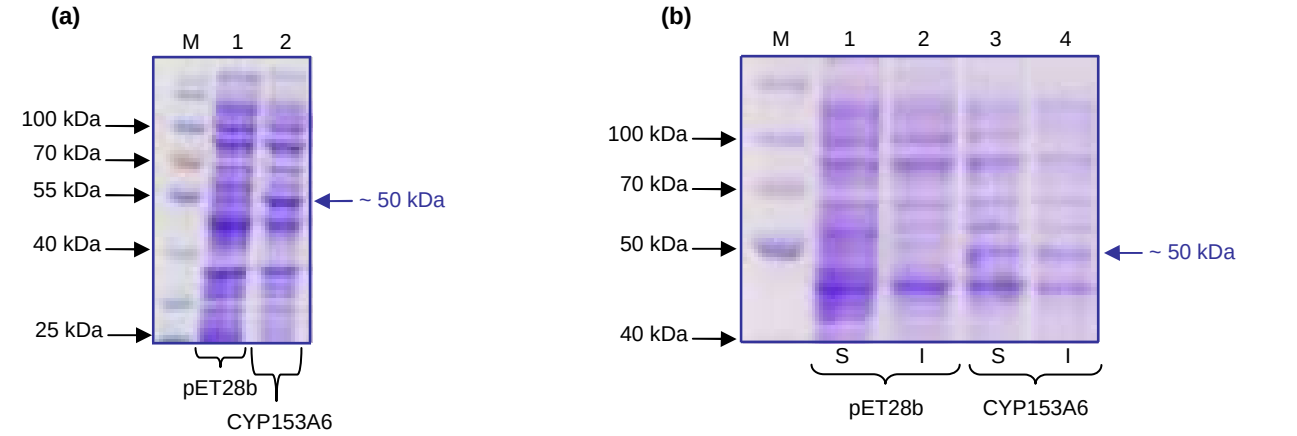


Figure 4.8 SDS-PAGE of **(a)** the crude cell lysates and **(b)** the soluble and insoluble fractions of *E. coli* BL21(DE3)pLysE strains transformed with the pET28 plasmid for expression of CYP153A6 and an empty pET28b plasmid. **(a)** Lane 1, empty pET28b plasmid; Lane 2, CYP153A6 (~ 50 kDa). M = PageRuler™ Pre-stained Protein Ladder. **(b)** Soluble fractions are indicated by "S" and insoluble fractions are indicated by "I". Lanes 1 and 2, empty pET28b plasmid; Lanes 3 and 4, CYP153A6. M = Spectra™ Multicolour Broad Range Protein Ladder.

4.4.3 Quantification of P450 Content

In order to determine if the P450 proteins were correctly folded and also to quantify the P450 content, CO-difference spectra were performed with both whole cells and the soluble fractions. A Soret band at 450 nm was obtained for CYP116B3 with both the whole cells (Fig. 4.9(a)) and the soluble fraction (Fig. 4.9(b)), indicating that this protein is correctly folded and therefore active. In the case of the CYP153A6/PFOR(CYP116B3) fusion, only P420 forms were detected in the whole cells (Fig. 4.9(c)), indicating a misfolded protein which in most cases is inactive. No P450 or P420 forms could be detected in the soluble fraction (Fig. 4.9(d)). A Soret band at 450 nm was obtained for CYP153A6 with both the whole cells (Fig. 4.9(e)) and the soluble fraction (Fig. 4.9(f)). The calculated P450 and P420 concentrations are summarised in Table 4.1.

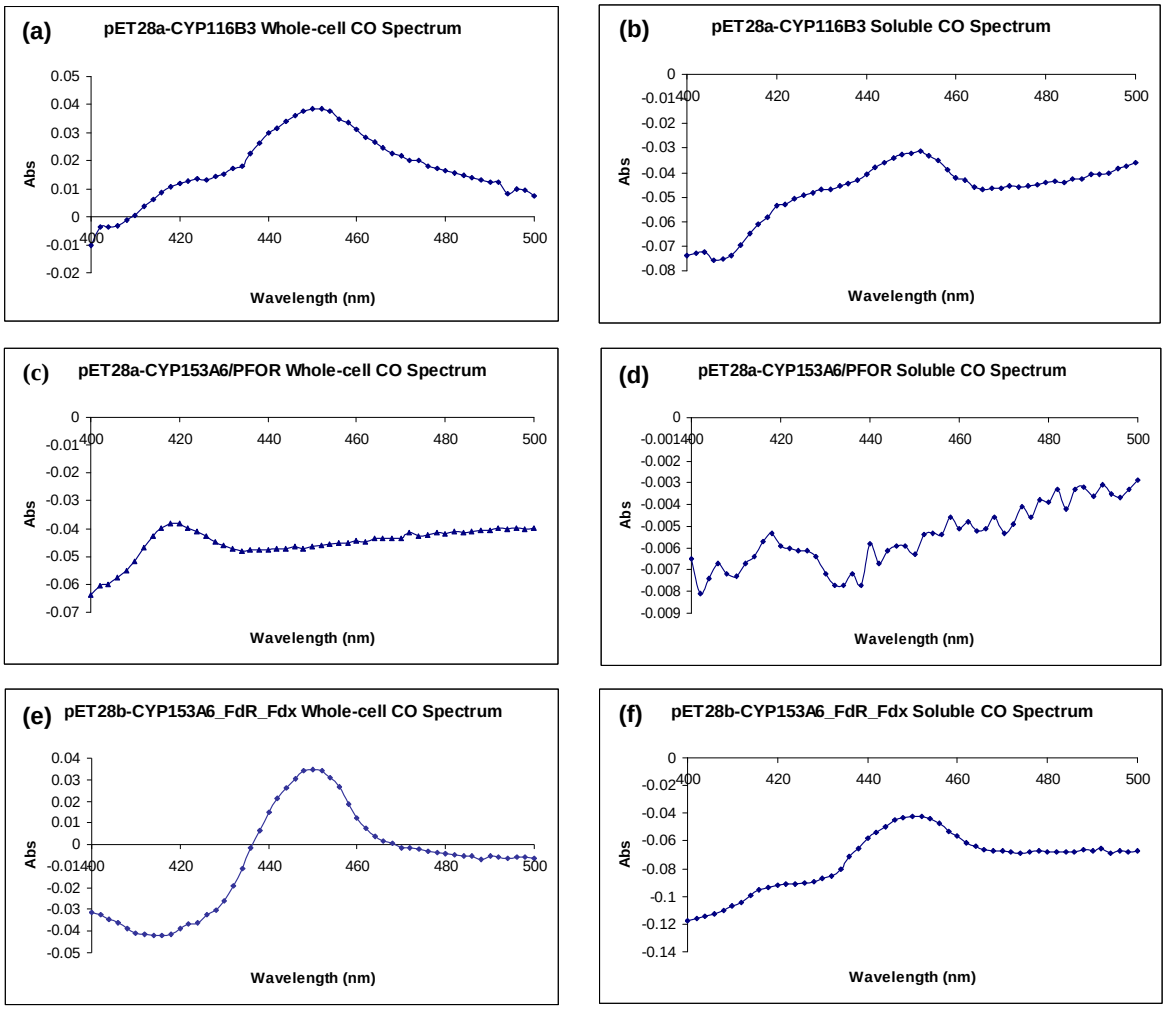


Figure 4.9 CO-difference spectra obtained using whole cells and soluble fractions. **(a)** and **(b),** CYP116B3; **(c)** and **(d),** CYP153A6/PFOR(CYP116B3); **(e)** and **(f)**, CYP153A6.

4.4.4 Reductase Activity Determinations

Cytochrome c reductase assays were conducted to determine the reductase activity of the expressed CYP116B3 PFOR domain as well as the CYP153A6 FdR-Fdx couple. With this assay, cytochrome c serves as an alternative electron acceptor which receives electrons from a cofactor (NADH or NADPH) *via* the reductase system. This results in an increase in absorbance at 550 nm, the rate of which can be used to determine the reductase activity. The cofactor used for CYP116B3, PFOR and the CYP153A6/PFOR(CYP116B3) fusion was NADPH, whereas the cofactor used for the FdR-Fdx couple was NADH.

Table 4.1 Summary of the results obtained for the expression of CYP116B3, the CYP116B3 PFOR domain, the CYP153A6/PFOR(CYP116B3) fusion and CYP153A6 using pET28

Protein	Expression [a]		Whole-cell enzyme content (μM) [b]		Soluble enzyme content (pmol.mg protein^{-1})		Soluble reductase activity (nmol.min^{-1}mg protein^{-1}) [c]
	Soluble	Insoluble	P420	P450	P420	P450	
CYP116B3	++	++	0	0.317	0	8.82	89.95 ± 5.97
PFOR	+	++	N/A	N/A	N/A	N/A	138.62 ± 1.12
Fusion	-	+++	0.038	0	0	0	3.72 ± 0.41
CYP153A6	+	+	0	0.446	0	11.03	2.93 ± 0.87

Notes: a = Observed level of expression from SDS-PAGE analysis: +++ = High; ++ = Intermediate; + = Low; - = No expression
b = 18x concentrated *E. coli* culture (final OD_{600} approximately 50)
c = Average of triplicates; standard deviations are indicated
N/A = No heme domain

The reductase assays were performed using the soluble fractions and were done in triplicate. The average reductase activities and the standard deviations are summarised in Table 4.1. The reductase activity of the PFOR domain was higher than that of CYP116B3. The reductase activity determined for the CYP153A6/PFOR(CYP116B3) fusion was approximately 20-fold lower than that of CYP116B3 and 35-fold lower than that determined for the CYP116B3 PFOR domain. The reductase activity of the FdR-Fdx couple of the CYP153A6 operon was also very

low. This could be the result of the very low level of expression of the ferredoxin reductase and ferredoxin compared to the level of expression of the CYP116B3 PFOR domain. Another possible explanation is that the CYP153A6 electron transport system consists of two separate proteins, an arrangement which probably results in slower and less efficient electron transfer than with a single reductase.

4.5 Modification of the Expression Conditions

In an attempt to minimise the formation of inclusion bodies and to obtain expression of the CYP153A6/PFOR(CYP116B3) fusion in the soluble fraction, the expression conditions were investigated. The first factor that was considered was the post-induction temperature. A post-induction temperature of 20℃ was initially selected for protein expression, which was the temperature used by Nodate *et al.* (2006) for the expression of the $P450_{balk}$/PFOR(CYP116B3) fusion in *E. coli* BL21(DE3) (Nodate *et al.*, 2006). The post-induction temperature was increased to 25℃, the temperature used by Nodate *et al.* (2006) for the expression of two other fusions constructed using the PFOR domain of P450RhF, and to 30℃, which is 2℃ higher than the post-induction temperature used by Fujii *et al.* (2006) for the expression of a CYP153A from *Acinetobacter* sp. OC4 (Fujii *et al.*, 2006). However, no expression of the CYP153A6/PFOR(CYP116B3) fusion was observed in the soluble fraction with SDS-PAGE using these two temperatures and no P450 forms were detected by CO-difference spectra performed with both whole cells and soluble fractions. Expression of the controls was still observed in both the soluble and insoluble fractions.

The second factor that was considered was the concentration of IPTG used for induction. The concentration was lowered from 0.5 mM to 0.4 mM and 0.3 mM, with a post-induction temperature of 20℃, but still no expression of the fusion was observed in the soluble fraction and no P450 forms were detected with whole cells or with the soluble fractions. The expression of the controls remained unchanged.

Liu *et al.* (2006) successfully expressed CYP116B3 in the soluble fraction using a pET28 plasmid with an IPTG concentration of 0.5 mM (Liu *et al.*, 2006), and this was also observed during the first part of this study. CYP153A6 was also successfully expressed in the soluble fraction using an IPTG concentration of 0.5 mM. It was therefore decided that the expression conditions would remain unchanged for the rest of the study: 0.5 mM IPTG for induction, with a post-induction temperature of 20℃.

4.6 Expression in the Absence of an N-terminal His-tag

All of the expression up to this point was done using pET28 plasmids, which resulted in the proteins having an N-terminal His-tag. In order to investigate whether the N-terminal His-tag had any effect on the folding of the proteins, the DNA encoding CYP116B3, the CYP116B3 PFOR domain and the CYP153A6/PFOR(CYP116B3) fusion, as well as the genes of the *CYP153A6* operon were cloned and expressed using the pET22b(+) plasmid, which does not contain the sequence encoding an N-terminal His-tag.

The insert from pET28a-CYP153A6/PFOR was excised *via* double digestion with *Nde*I and *Hin*dIII, and ligated between these sites in the pET22b(+) plasmid, resulting in construct pET22b-CYP153A6/PFOR (Fig. 4.10).

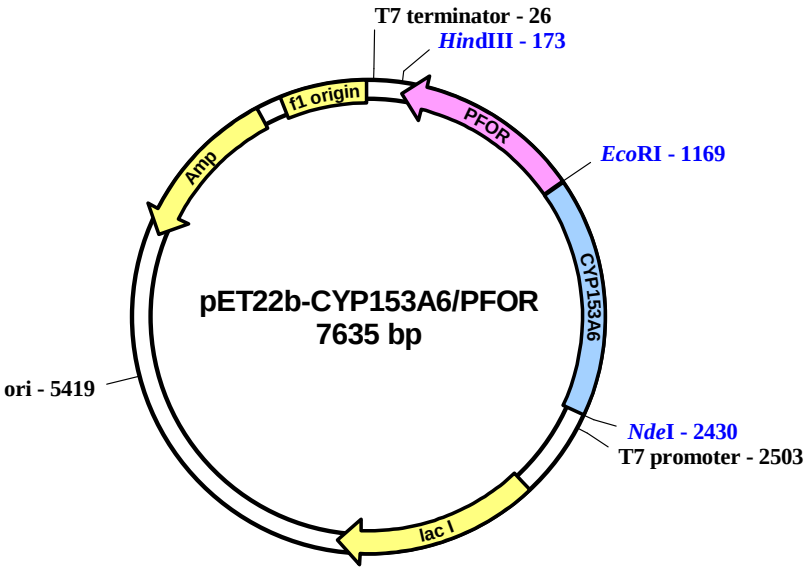


Figure 4.10 Schematic representation of the pET22b-CYP153A6/PFOR construct. The *CYP153A6* gene is shown in blue and the DNA encoding the CYP116B3 PFOR domain is shown in pink.

The pET22b-CYP116B3 plasmid (Fig. 4.11(a)) was constructed by amplifying the *CYP116B3* gene from pET28a-CYP116B3 using primers CYP116-pET-F and CYP116-pET-R. These primers resulted in the replacement of the *Eco*RI restriction site at the 5'-end of the *CYP116B3* gene with an *Nde*I site. This facilitated the removal of the *pelB* signal sequence present in the pET22b(+) plasmid, which is involved in the potential periplasmic localisation of expressed proteins. An amplicon of the expected size of approximately 2 300 bp was obtained (Fig. 4.12, Lane 2) and was sub-cloned using pGem-T Easy. Plasmid DNA was extracted from a positive transformant and the insert was sequenced using the T7 and Sp6 primers and the appropriate

internal primers listed in Table 3.3 (Chapter 3, p.38). The insert was excised from the plasmid *via* double digestion with *Nde*I and *Hin*dIII and was ligated between these restriction sites in the pET22b(+) plasmid.

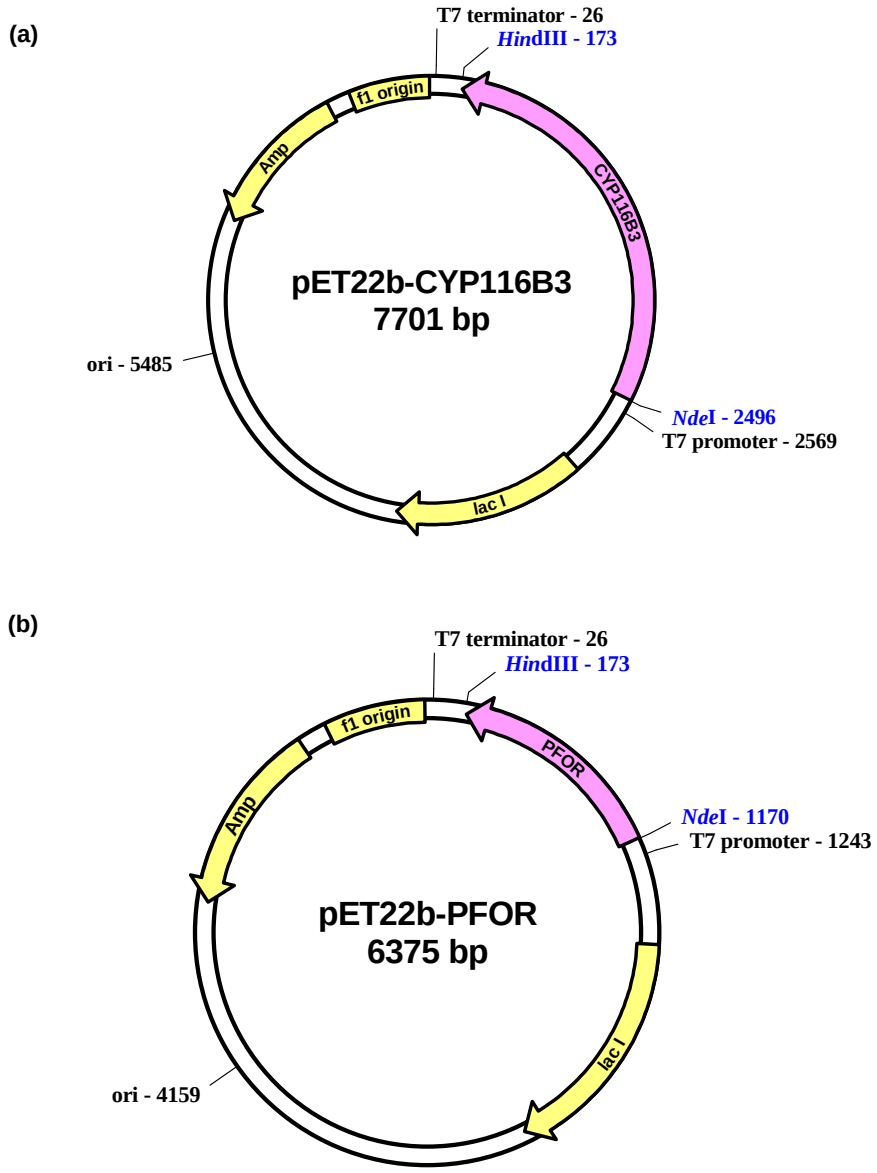


Figure 4.11 Schematic representations of **(a)** the pET22b-CYP116B3 and **(b)** the pET22b-PFOR constructs. The *CYP116B3* gene and the DNA encoding the CYP116B3 PFOR domain are shown in pink.

The pET22b-PFOR plasmid (Fig. 4.11(b)) was constructed using a similar approach. The DNA encoding the CYP116B3 PFOR domain was amplified using primers Red-pET-F and CYP116-pET-R which also introduced an Nde*I* and a Hin*dIII* restriction site at the 5'- and 3'-ends of the amplicon, respectively. An amplicon of the expected size of approximately 1 000 bp was

obtained (Fig. 4.12, Lane 1) and was sub-cloned using pGem-T Easy. Plasmid DNA was extracted from a positive transformant and the insert was sequenced using the primers T7 and Sp6 and the appropriate internal primers listed in Table 3.3 (Chapter 3, p.38). The insert was excised from the plasmid *via* double digestion with *Nde*I and *Hin*dIII and was ligated to pET22b(+) between these restriction sites.

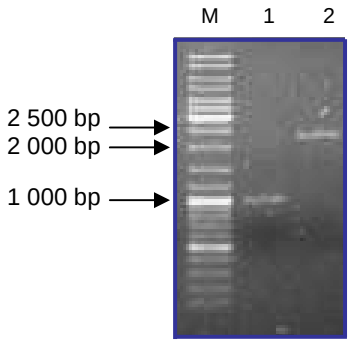


Figure 4.12 Amplification of the DNA encoding CYP116B3 and the CYP116B3 PFOR domain. 0.8% (w/v) agarose gel stained with GoldView Nucleic Acid Stain. Lane 1, DNA encoding the PFOR domain amplified using primers Red-pET-F and CYP116-pET-R; Lane 2, *CYP116B3* gene amplified using primers CYP116-pET-F and CYP116-pET-R. M = GeneRuler^{TM} DNA Ladder Mix.

The pET22b-CYP153A6_FdR_Fdx plasmid was constructed by amplifying the complete *CYP153A6* operon from the pET28b-CYP153A6_FdR_Fdx plasmid using pCom-153F and HXN-1500-op-RV, resulting in the replacement of the *Nhe*I restriction site at the 5'-end of the DNA with an *Nde*I restriction site, as the multiple cloning site of the pET22b(+) plasmid does not contain a *Nhe*I site. An amplicon of the expected size of approximately 3 000 bp was obtained and was sub-cloned using pGem-T Easy. Plasmid DNA was extracted from a positive transformant and the insert was sequenced using the primers T7 and Sp6 and the appropriate internal primers listed in Table 3.3 (Chapter 3, p.38). The insert was excised *via* digestion with *Nde*I, *Hin*dIII and *Ssp*I (Fig. 4.13) and was ligated to the pET22b(+) plasmid between the *Nde*I and *Hin*dIII restriction sites, resulting in construct pET22b-CYP153A6_FdR_Fdx (Fig. 4.14).

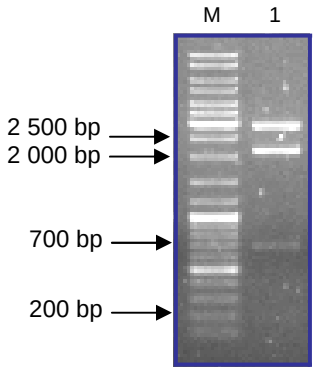


Figure 4.13 Restriction enzyme digestion of the pGem-T Easy plasmid containing the *CYP153A6* operon insert. 0.8% (w/v) agarose gel stained with GoldView Nucleic Acid Stain showing the fragments obtained using *Nde*I, *Hin*dIII and *Ssp*I. M = GeneRuler^{TM} DNA Ladder Mix.

*Ssp*I was used with the *Nde*I and *Hin*dIII restriction enzymes to excise the insert because when only *Nde*I and *Hin*dIII were used, the pGem-T Easy backbone and the insert could not be separated as both were approximately 3 000 bp in size. *Ssp*I cleaves the pGem-T Easy backbone twice, generating fragments of approximately 2 000 bp, 700 bp and 200 bp.

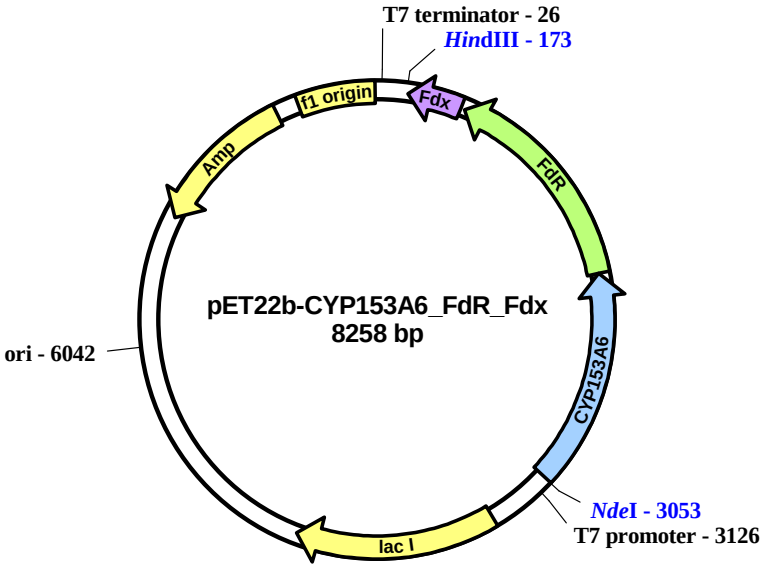


Figure 4.14 Schematic representation of the pET22b-CYP153A6_FdR_Fdx construct. The genes encoding the CYP153A6, the ferredoxin reductase and the ferredoxin are shown in blue, green and purple, respectively.

4.6.1 Analyses of the Expressed Proteins

SDS-PAGE analysis was performed using the soluble and insoluble fractions obtained after cell disruption. Expression of CYP116B3 was observed in both fractions (Fig. 4.15, Lanes 5 and 6), with the level of expression in the insoluble fraction being higher than that observed using the

pET28 plasmid. A CO-difference spectrum performed using whole cells indicated the presence of both P450 and P420 forms, with the P450 concentration being higher than that detected using the pET28 plasmid. No P420s were detected in the soluble fraction, but the concentration of detected P450s was similar to that detected with the pET28 plasmid. The reductase activity was higher than that detected with pET28. These results are summarised in Table 4.2.

No expression of the CYP116B3 PFOR domain could be observed in either of the fractions (Fig. 4.15, Lanes 7 and 8) and the detected reductase activity was lower than that detected using the pET28 plasmid (Table 4.2). Expression of the CYP153A6 could also not be observed in either of the fractions (Fig. 4.15, Lanes 1 and 2) and the reductase activity of the FdR-Fdx couple could not be detected. CO-difference spectra performed using whole cells indicated the presence of only P420 forms of CYP153A6. No P450 or P420 forms of the protein were detected in the soluble fraction (Table 4.2).

Expression of the CYP153A6/PFOR(CYP116B3) fusion protein was once again observed in the insoluble fraction but not in the soluble fraction (Fig. 4.15, Lanes 3 and 4). The concentration of P420 forms detected using a whole-cell CO-difference spectrum was higher than that detected with the pET28 plasmid, but no P450 forms of the protein were detected. A low P420 concentration was detected in the soluble fraction, whereas with pET28 no P420s could be detected. The reductase activity detected using the soluble fraction of the fusion was 12-fold higher than that detected with pET28 (Table 4.2).

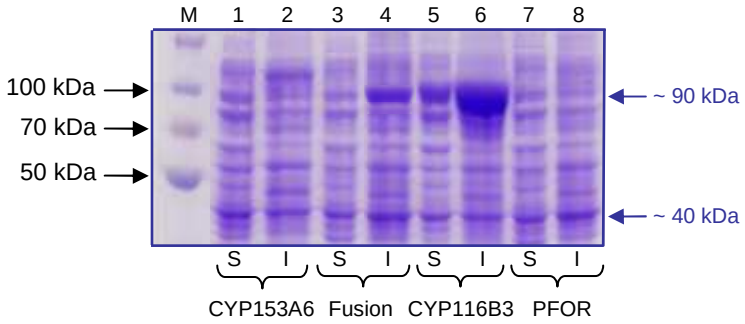


Figure 4.15 SDS-PAGE of the soluble and insoluble fractions of *E. coli* BL21(DE3)pLysE strains transformed with pET22 plasmids for expression of CYP153A6, the CYP153A6/PFOR(CYP116B3) fusion, CYP116B3 and the CYP116B3 PFOR domain. Soluble fractions are indicated by "S" and insoluble fractions are indicated by "I". Lanes 1 and 2, CYP153A6 (~ 50 kDa); Lanes 3 and 4, CYP153A6/PFOR(CYP116B3) fusion (~ 90 kDa); Lanes 5 and 6, CYP116B3 (~ 90 kDa); Lanes 7 and 8, CYP116B3 PFOR domain (~ 40 kDa). M = PageRuler^{TM} Pre-stained Protein Ladder.

Table 4.2 Summary of the results obtained from the expression of CYP116B3, the CYP116B3 PFOR domain, the CYP153A6/PFOR(CYP116B3) fusion and CYP153A6 using pET22b. The results obtained using pET28 are included for comparison purposes (highlighted in grey).

Protein	Expression [a]		Whole-cell enzyme content (μM) [b]		Soluble enzyme content (pmol.mg protein^{-1})		Soluble reductase activity (nmol.min^{-1}mg protein^{-1}) [c]
	Soluble	Insoluble	P420	P450	P420	P450	
CYP116B3 + His_6	++	++	0	0.317	0	8.82	89.95 ± 5.97
CYP116B3 – His_6	++	+++	1.470	2.563	0	7.34	125.79 ± 0.55
PFOR + His_6	+	++	N/A	N/A	N/A	N/A	138.62 ± 1.12
PFOR – His_6	-	-	N/A	N/A	N/A	N/A	63.97 ± 1.91
Fusion + His_6	-	+++	0.038	0	0	0	3.72 ± 0.41
Fusion – His_6	-	++	0.525	0	0.50	0	47.57 ± 5.27
CYP153A6 + His_6	+	+	0	0.446	0	11.03	2.93 ± 0.87
CYP153A6 – His_6	-	-	0.317	0	0	0	n.d.

Notes: a = Observed level of expression from SDS-PAGE analysis: +++ = High; ++ = Intermediate; + = Low; - = No expression

b = 18x concentrated *E. coli* culture (final OD_{600} approximately 50)

c = Average of triplicates; standard deviations are indicated

N/A = No heme domain

n.d. = Not detected

The absence of an N-terminal His-tag enhanced the expression of CYP116B3 and the reductase activities of CYP116B3 and the CYP153A6/PFOR(CYP116B3) fusion, indicating that the absence of the N-terminal His-tag may have a positive effect on the reductase activity of this fusion-type arrangement. Higher enzyme concentrations were detected, but no P450 forms of the CYP153A6/PFOR(CYP116B3) fusion were detected and no expression was observed in the

soluble fraction. The folding of the CYP153A6/PFOR(CYP116B3) fusion protein seemed relatively unaffected by the presence or the absence of an N-terminal His-tag.

Because of the negative affect of the absence of an N-terminal His-tag on the expression of CYP153A6 and the CYP116B3 PFOR domain, it was decided that pET28 would remain the plasmid of choice for the rest of the study.

4.7 Site-directed Mutagenesis of the PFOR Linker

A linker is a short fragment of DNA that links the heme and reductase domains together in both natural and artificial fusions and allows the protein domains to fold independently. By linking these domains, the velocity and efficiency of catalysed reactions can be increased (Munro *et al.*, 2007).

```
CYP116B3  1   MSASVPASA--CPVDHAALAGGCPVSTNAAAFDPFGPAYQADPAESLRWSRDEEPVFYSPELGY 62
P450RhF   1   MSASVPASAPACPVDHAALAGGCPVSANAAAFDPFGSAYQTDPAESLRWSRDEEPVFYSPELGY 64

CYP116B3  63  WVVTRYEDVKAVFRDNLVFSPAIALEKITPVSEEATATLARYDYAMARTLVNEDEPAHMPRRRA 126
P450RhF   65  WVVTRYEDVKAVFRDNILFSPAIALEKITPVSAEATATLARYDYAMARTLVNEDEPAHMPRRRA 128

CYP116B3  127 LMDPFTPKELAHHEAMVRRLTREYVDRFVESGKADLVDEMLWEVPLTVALHFLGVPEEDMATMR 190
P450RhF   129 LMDPFTPKELAHHEAMVRRLTREYVDRFVESGKADLVDEMLWEVPLTVALHFLGVPEEDMATMR 192

CYP116B3  191 KYSIAHTVNTWGRPAPEEQVAVAEAVGRFWQYAGTVLEKMRQDPSGHGWMPYGIRMQQQMPDVV 254
P450RhF   193 KYSIAHTVNTWGRPAPEEQVAVAEAVGRFWQYAGTVLEKMRQDPSGHGWMPYGIRKQREMPDVV 256

CYP116B3  255 TDSYLHSMMMAGIVAAHETTANASANAFKLLLENRPVWEEICADPSLIPNAVEECLRHSGSVAA 318
P450RhF   257 TDSYLHSMMMAGIVAAHETTANASANAFKLLLENRAVWEEICADPSLIPNAVEECLRHSGSVAA 320

CYP116B3  319 WRRVATTDTRIGDVDIPAGAKLLVVNASANHDERHFDRPDEFDIRRPNSSDHLTFGYGSHQCMG 382
P450RhF   321 WRRVATADTRIGDVDIPAGAKLLVVNASANHDERHFERPDEFDIRRPNSSDHLTFGYGSHQCMG 384

CYP116B3  383 KNLARMEMQIFLEELTTRLPHMELVPDQEFTYLPNTSFRGPDHVWVQWDPQANPERTDPAVLQR 446
P450RhF   385 KNLARMEMQIFLEELTTRLPHMELVPDQEFTYLPNTSFRGPDHVWVQWDPQANPERTDPAVLHR 448

CYP116B3  447 QHPVTIGEPSTRSVSRTVTVERLDRIVDDVLRVVLRAPAGNALPAWTPGAHIDVDLGALSRQYS 510
P450RhF   449 HQPVTIGEPAARAVSRTVTVERLDRIADDVLRLVLRDAGGKTLPTWTPGAHIDLDLGALSRQYS 512

CYP116B3  511 LCGAPDAPTYEIAVLLDPESRGGSRYVHEQLRVGGSLRIRGPRNHFALDPDAEHYVFVAGGIGI 574
P450RhF   513 LCGAPDAPSYEIAVHLDPESRGGSRYIHEQLEVGSPLRMRGPRNHFALDPGAEHYVFVAGGIGI 576

CYP116B3  575 TPVLAMADHARARGWSYELHYCGRNRSGMAYLERVAGHGDRAALHVSAEGTRVDLAALLATPVS 638
P450RhF   577 TPVLAMADHARARGWSYELHYCGRNRSGMAYLERVAGHGDRAALHVSEEGTRIDLAALLAEPAP 640

CYP116B3  639 GTQIYACGPGRLLAGLEDASRHWPDGALHVEHFTSSLTALDPDVEHAFDLDLRDSGLTVRVEPT 702
P450RhF   641 GVQIYACGPGRLLAGLEDASRNWPDGALHVEHFTSSLAALDPDVEHAFDLELRDSGLTVRVEPT 704

CYP116B3  703 QTVLDALRANNIDVPSDCEEGLCGSCEVTVLEGEVDHRDTVLTKAERAANRQMMTCCSRACGDR 766
P450RhF   705 QTVLDALRANNIDVPSDCEEGLCGSCEVAVLDGEVDHRDTVLTKAERAANRQMMTCCSRACGDR 768

CYP116B3  767 LTLRL 771
P450RhF   769 LALRL 773
```

Figure 4.16 Amino acid alignment of CYP116B3 and P450RhF. The linker region is underlined in black. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

As previously mentioned, Nodate *et al.* (2006) successfully constructed a self-sufficient CYP153A protein by fusing the heme domain of $P450_{balk}$ (CYP153A13a) to the reductase (PFOR) domain of P450RhF (CYP116B2). CYP116B3 is 93% identical to P450RhF, with the linker region differing by five amino acids out of a total of sixteen, making the linker region one of the most variable regions (Fig. 4.16). In order to determine if these five amino acids are important for the functional expression of artificial fusion proteins and therefore for the folding of the CYP153A6/PFOR(CYP116B3) fusion, site-directed mutagenesis was employed to mutate the DNA sequence encoding the CYP116B3 linker region to match that of P450RhF (Fig. 4.17), with a total of eight basepair mutations.

(a) **CYP116B3 Linker (*Rhodococcus ruber* DSM 44319)**

V	L	Q	R	Q	H	P	V	T	I	G	E	P	S	T	R
GTG	CTG	CAA	CGG	CAG	CAT	CCC	GTC	ACC	ATC	GGC	GAG	CCC	TCC	ACC	CGG

(b) **P450RhF (CYP116B2) Linker (*Rhodococcus* sp. NCIMB 9784)**

V	L	H	R	H	Q	P	V	T	I	G	E	P	A	A	R
GTG	CTG	CAC	CGG	CAT	CAA	CCG	GTC	ACC	ATC	GGA	GAA	CCC	GCC	GCC	CGG

Figure 4.17 The amino acid and nucleotide sequences of the linker region of **(a)** CYP116B3 and **(b)** P450RhF. The amino acids highlighted in blue are those that differ between the two linker sequences. The bases highlighted in yellow in **(a)** were targeted for site-directed mutagenesis to match those of P450RhF, highlighted in yellow in **(b)**.

The pET28a-CYP153A6/PFOR plasmid was amplified using the primers CYP116LF and CYP116RF which were designed according to the desired DNA sequence, with each primer starting in the middle of the linker region and running in opposite directions as indicated in Figure 4.18. An amplicon of the expected size of approximately 6 300 bp was obtained (Fig. 4.19). Following *Dpn*I treatment, phosphorylation and self-ligation, the resulting plasmid was propagated in *E. coli* XL-10 Gold.

In order to eliminate any undesired mutations which may have been introduced during the amplification of the plasmid, the DNA encoding the mutated linker and the CYP116B3 PFOR domain, which was sequenced using the T7 Terminator primer and the relevant internal primers (Table 3.3, p.38), was excised from the plasmid *via* double digestion with *Eco*RI and *Not*I. *Not*I cleaves the pET28 plasmid adjacent to *Hin*dIII. This resulted in fragments of the expected size of approximately 1 000 bp and 5 300 bp. The excised DNA fragment encoding the modified

CYP116B3 PFOR domain was used to replace the DNA encoding the original CYP116B3 PFOR domain in pET28a-CYP153A6/PFOR, resulting in construct pET28a-CYP153A6/Linker/PFOR.

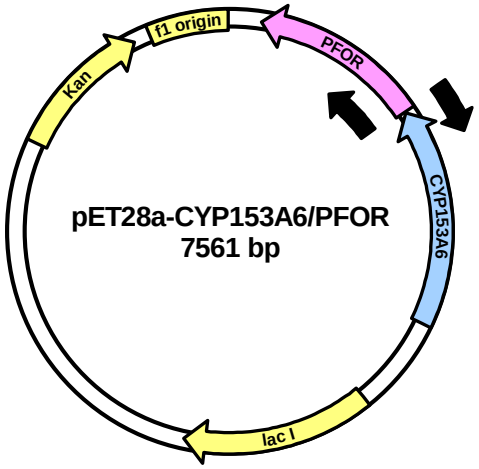


Figure 4.18 Schematic representation of the approach used for site-directed mutagenesis of the CYP116B3 PFOR linker region. Each of the primers, represented by the black arrows, included the nucleotide sequence encoding half of the desired linker sequence.

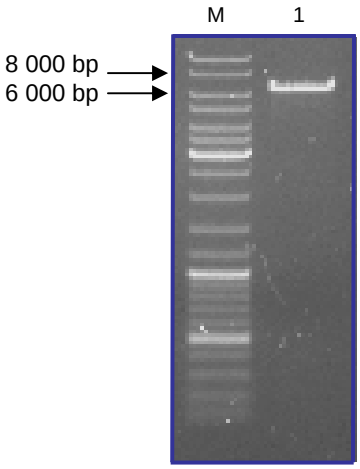


Figure 4.19 Amplification of the pET28a-CYP153A6/PFOR plasmid using primers CYP116LF and CYP116RF. 0.8% agarose gel stained with GoldView Nucleic Acid Stain showing the PCR product (Lane 1). M = GeneRuler^{TM} DNA Ladder Mix.

4.7.1 Analysis of the Expressed Protein

The expression of the protein, referred to as the modified CYP153A6/PFOR(CYP116B3) fusion, was analysed by SDS-PAGE using the soluble and insoluble fractions obtained after cell

disruption. A faint band corresponding to the expected size of approximately 90 kDa was observed in the insoluble fraction (Fig. 4.20, Lane 4), but not in the soluble fraction (Fig. 4.20, Lane 3).

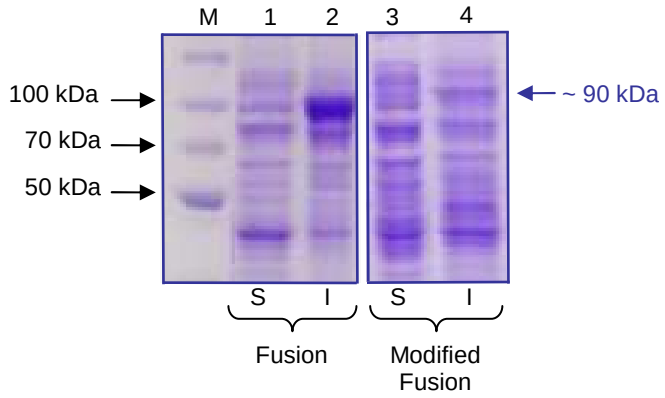


Figure 4.20 SDS-PAGE of the soluble and insoluble fractions of *E. coli* BL21(DE3)pLysE strains transformed with pET28 plasmids for the expression of the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion. Soluble fractions are indicated by "S" and insoluble fractions are indicated by "I". Lanes 1 and 2, CYP153A6/PFOR(CYP116B3) fusion (~ 90 kDa); Lanes 3 and 4, modified CYP153A6/PFOR(CYP116B3) fusion (~ 90 kDa). M = Spectra™ Multicolour Broad Range Protein Ladder.

CO-difference spectra were performed with whole cells and the soluble fraction. A low concentration of P420 forms of the modified CYP153A6/PFOR(CYP116B3) protein was detected with the whole cells, but no P450 forms were detected and no P420 or P450 forms were detected in the soluble fraction. The reductase activity of the modified CYP153A6/PFOR(CYP116B3) protein was found to be slightly higher than that obtained with the original CYP153A6/PFOR(CYP116B3) fusion expressed using pET28. These results are summarised in Table 4.3.

4.8 Analyses of Proteins Expressed using *E. coli* Rosetta-gami 2(DE3)pLysS

Rare codons are codons that require tRNAs that are rare or lacking in *E. coli*. The presence of rare codons in a gene often hampers the expression of that gene in bacteria (Ventura & Villaverde, 2006). The gene encoding the CYP153A6 heme domain contains nineteen rare codons (Fig. 4.21). CYP153A6 was, however, expressed in *E. coli* BL21(DE3)pLysE in a P450 form (see section 4.4), but in order to investigate whether these rare codons were influencing the expression and folding of this protein and therefore the expression and folding of the CYP153A6/PFOR(CYP116B3) fusions, *E. coli* Rosetta-gami 2(DE3)pLysS was used for

expression. This strain contains a pRARE plasmid encoding tRNAs for seven rare codons that enables it to be used as an expression host for genes containing rare codons.

Table 4.3 Summary of results obtained for the expression of the modified CYP153A6/PFOR(CYP116B3) fusion using pET28. The results obtained with the CYP153A6/PFOR(CYP116B3) fusion are included for comparison purposes, highlighted in grey.

Protein	Expression [a]		Whole-cell enzyme content (μM) [b]		Soluble enzyme content (pmol.mg protein^{-1})		Soluble reductase activity (nmol.min^{-1}mg protein^{-1}) [c]
	Soluble	Insoluble	P420	P450	P420	P450	
Fusion	-	+++	0.038	0	0	0	3.72 ± 0.41
Modified Fusion	-	+	0.003	0	0	0	5.14 ± 0.65

Note: a = Observed level of expression from SDS-PAGE analysis: +++ = High; + = Low; - = No expression
b = 18x concentrated *E. coli* culture (final OD_{600} approximately 50)
c = Average of triplicates; standard deviations are indicated

All of the expression up to this point was done using *E. coli* BL21(DE3)pLysE. A possibility that had to be considered was whether this strain, in combination with the pET28 expression system, was resulting in the protein being expressed too quickly, resulting in molecular crowding within the cell preventing the P450 proteins from assuming the correct conformation. *E. coli* Rosetta-gami 2(DE3)pLysS grows slower than the BL21(DE3)pLysE strain and expression occurs at a slower rate. Therefore, this strain was also used to determine if these properties could result in the proteins having enough time to fold correctly, possibly allowing them to be expressed in the soluble fraction as P450s.

Competent cells were transformed with the pET28b-CYP153A6_FdR_Fdx, pET28a-CYP153A6/PFOR and pET28a-CYP153A6/Linker/PFOR plasmids for expression. SDS-PAGE analysis was performed using the soluble and insoluble fractions obtained after cell disruption. No expression of CYP153A6 could be observed in the soluble or insoluble fractions (Fig. 4.22, Lanes 1 and 2). CO-difference spectra indicated the presence of P420 forms of the protein in both the whole cells and the soluble fraction, but no P450 forms could be detected. The reductase activity detected for the FdR-Fdx couple was half of that detected using the BL21(DE3)pLysE strain.

cat atg gaa atg acg gtg gcc gcc agc gac gcg acg aac gcg gcg tac ggg atg gcc
ttg gag gac atc gat gtg agc aat CCC gtg ctg ttc cgg gac aac acc tgg cat CCC
tat ttc aaa cgc ctg cgc gaa gag gat ccg gtt cac tac tgc aag agc agc atg ttc
ggc CCC tac tgg tcg gtg acc aag tac cgc gac atc atg gca gtc gag acc aac ccg
aag gtg ttc tcg tcg gag gct aaa agt ggc ggc atc acc atc atg gac gac aac gcc
gca gca tcg ctt CCC atg ttc atc gcc atg gat cca ccg aaa cac gat gtc cag CGA
aag acc gtg agc ccg atc gtg gcg CCC gaa aac ctt gcc acc atg gaa tcg gtg att
cgc cag cgc acc gcg gac ctc ctc gac gga ttg ccg atc aat gaa gag ttc gac tgg
gtg cat cgg gtg tcg atc gaa ttg acc acg aag atg ctg gcg acg ctg ttc gat ttt
CCC tgg gac gac cgc gcc aag ttg acg cgc tgg tcg gac gtc acc acg gcg ttg CCC
ggt ggc ggg atc atc gat tct gaa gaa cag cgc atg gcc gag ctg atg gag tgc gcg
acg tat ttc acc gag ctg tgg aac cag cgc gtg aat gcc gaa CCC aag aac gat ctc
atc tcg atg atg gcc cat tcg gag tca aca CGA cac atg gcg CCC gag gaa tat ctc
gga aac atc gtg ctg ctg atc gtc ggc ggc aac gac acc acc cgc aac tcg atg acc
ggc ggt gtg ttg gcc ctg aac gaa ttt CCC gac gaa tac cgc aaa ctg tcc gcc aac
ccg gcg ttg atc agc tct atg gtg tcg gag atc atc cgg tgg caa aca cct ctt tcg
cac atg cgt cgt acc gca ttg gaa gac atc gag ttc ggc ggc aag cac atc cgc cag
ggc gac aaa gtc gtg atg tgg tac gtg tcc ggc aac cgg gac CCC gag gcc atc gac
aat CCC gac aca ttc atc atc gat cgc gcc aag CCC cgc cag cac ttg tcc ttc ggg
ttc ggc atc cac cgc tgc gtc ggc aac AGA ccc gcc gaa CTA cag ctc aac atc ctg
tgg gaa gaa atc ctc aaa cgg tgg ccg gac cca ctg cag atc cag gtt ctt caa gaa
ccg acc cgc gtg ctc tca ccg ttc gtc aag ggc tac gaa tcg ctg CCC gtg cgc atc
aac gcc gaa ttc

Figure 4.21 Results obtained for the analysis of the *CYP153A6* gene for rare codons, using the Rare Codon Calculator (RaCC) (http://nihserver.mbi.ucla.edu/RACC/). Red codons are rare arginine codons (AGG, AGA, CGA), green codons are rare leucine codons (CTA) and orange codons are rare proline codons (CCC). The base highlighted in blue is the base that was mutated, resulting in the replacement of a leucine residue with a proline residue, generating a rare proline codon in the process.

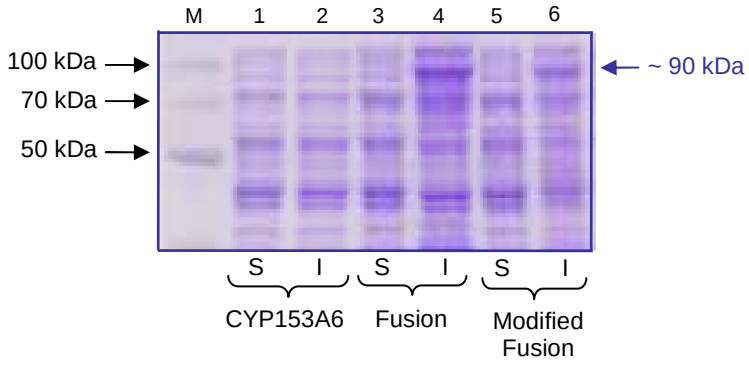


Figure 4.22 SDS-PAGE of the soluble and insoluble fractions of *E. coli* Rosetta-gami 2(DE3)pLysS strains transformed with pET28 plasmids for the expression of CYP153A6, the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion. Soluble fractions are indicated by "S" and insoluble fractions are indicated by "I". Lanes 1 and 2, CYP153A6 (~ 50 kDa); Lanes 3 and 4, CYP153A6/PFOR(CYP116B3) fusion (~ 90 kDa); Lanes 5 and 6, modified CYP153A6/PFOR(CYP116B3) fusion (~ 90 kDa). M = Spectra™ Multicolour Broad Range Protein Ladder.

With both the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion, expression was observed in the insoluble fractions (Fig. 4.22, Lanes 4 and 6, respectively), but not in the soluble fractions (Fig. 4.22, Lanes 3 and 5, respectively). P420 forms were detected in the whole cells and the soluble fractions, but no P450s were detected. The reductase activity detected for the CYP153A6/PFOR(CYP116B3) fusion was slightly higher than that detected using the BL21(DE3)pLysE strain, whereas the reductase activity of the modified CYP153A6/PFOR(CYP116B3) fusion was slightly lower. The results are summarised in Table 4.4. Overall, the performance of the Rosetta-gami strain was relatively poor and the desired results were not obtained.

Table 4.4 Summary of the results obtained for the expression of CYP153A6, the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion using the Rosetta-gami 2(DE3)pLysS strain. The results obtained using the BL21(DE3)pLysE strain are included for comparison purposes, highlighted in grey.

Protein	Strain	Expression [a]		Whole-cell enzyme content (μM) [b]		Soluble enzyme content (pmol.mg protein^{-1})		Soluble reductase activity (nmol.min^{-1}mg protein^{-1}) [c]
		Soluble	Insoluble	P420	P450	P420	P450	
CYP153A6	BL21	+	+	0	0.446	0	11.03	2.93 ± 0.87
CYP153A6	Rosetta	-	-	0.526	0	2.99	0	1.23 ± 0.07
Fusion	BL21	-	+++	0.038	0	0	0	3.72 ± 0.41
Fusion	Rosetta	-	++	0.525	0	3.45	0	4.44 ± 0.26
Modified Fusion	BL21	-	+	0.003	0	0	0	5.14 ± 0.65
Modified Fusion	Rosetta	-	+	1.864	0	0.12	0	2.35 ± 0.88

Notes: a = Observed level of expression from SDS-PAGE analysis: +++ = High; ++ = Intermediate; + = Low; - = No expression

b = 18x concentrated *E. coli* culture (final OD_{600} approximately 50)

c = Average of triplicates; standard deviations are indicated

4.9 Octane Bioconversions

4.9.1 Bioconversions using CYP153A6

Despite investigating a number of factors that could be responsible for the expression of the CYP153A6/PFOR(CYP116B3) fusion only in the insoluble fraction, no expression was observed in the soluble fraction and only P420 forms of the protein could be detected. In most cases, P420s are inactive and will not result in substrate conversion, but in order to confirm that the CYP153A6/PFOR(CYP116B3) fusion was incapable of substrate hydroxylation, a bioconversion experiment was conducted. Octane was selected as the substrate as this is the preferred substrate of CYP153A6, with terminal hydroxylation resulting in the production of 1-octanol (Nodate *et al.*, 2006; Funhoff *et al.*, 2007).

Initial octane bioconversion experiments were conducted using only CYP153A6 to firstly, confirm that the bioconversion protocol would result in substrate conversion and secondly, to optimise the bioconversion conditions. The first experiment was performed using 50 mM potassium phosphate buffer (pH 7.2), containing 0.8% glycerol, for resuspension (Nodate *et al.*, 2006). The bioconversion was carried out in vials with a volume of 12 mL, with 100 µL of *n*-octane serving as the substrate, resulting in a substrate concentration of approximately 0.56 M. After a bioconversion reaction period of 24 hours, 3.59 mM of 1-octanol was produced, with the product formation levelling off after a period of 7 to 8 hours (Fig 4.23).

Two factors were considered that could possibly be limiting, the first being cofactor regeneration and the second being the oxygen level. In order to investigate the first possibility, glucose was added to the bioconversion reaction mixture to a final concentration of 40 mM, which would allow the cells to regenerate NADH. After a reaction period of 24 hours the concentration of 1-octanol produced was 6.54 mM, almost double the concentration produced in the absence of glucose, indicating that cofactor regeneration was limiting in the absence of glucose. However, even in the presence of glucose, the product formation still levelled off after about 7 to 8 hours (Fig. 4.23).

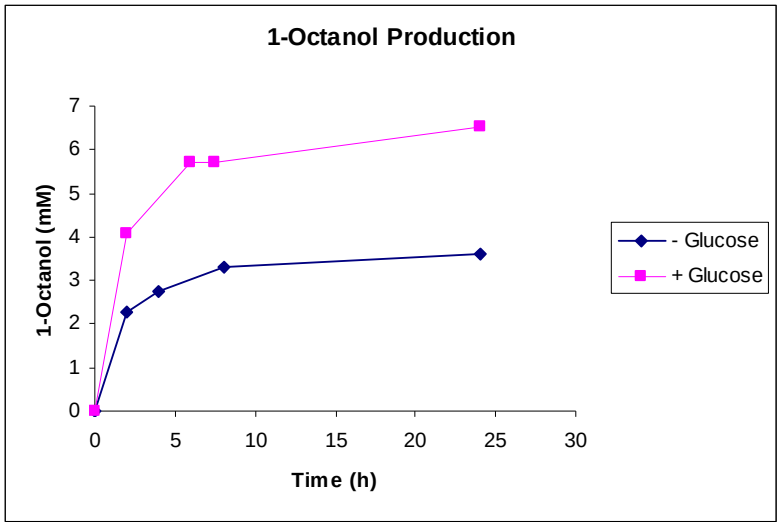


Figure 4.23 Graph showing the production of 1-octanol in the presence and absence of glucose over a 24-hour bioconversion reaction period.

In order to investigate the possibility that oxygen was the limiting factor, vials with volumes of 12 mL, 24 mL and 40 mL were used. The volume of *n*-octane added to each vial was increased to 333 µL (concentration of 1.5 M in the bioconversion reaction mixture) to match the ratio of substrate volume to culture volume used by Nodate *et al.* (2006) and Fujii *et al.* (2006) for octane bioconversions using the $P450_{balk}$/PFOR(P450RhF) fusion protein and a CYP153A from *Acinetobacter* sp. OC4, respectively (Fujii *et al.*, 2006; Nodate *et al.*, 2006). In addition, a buffer used by Fujii *et al.* (2006) was tested: 0.2 M sodium phosphate buffer (pH 7.0), containing 0.8% glycerol and 100 µg.mL^{-1} $FeSO_4$.

(a) **(b)**

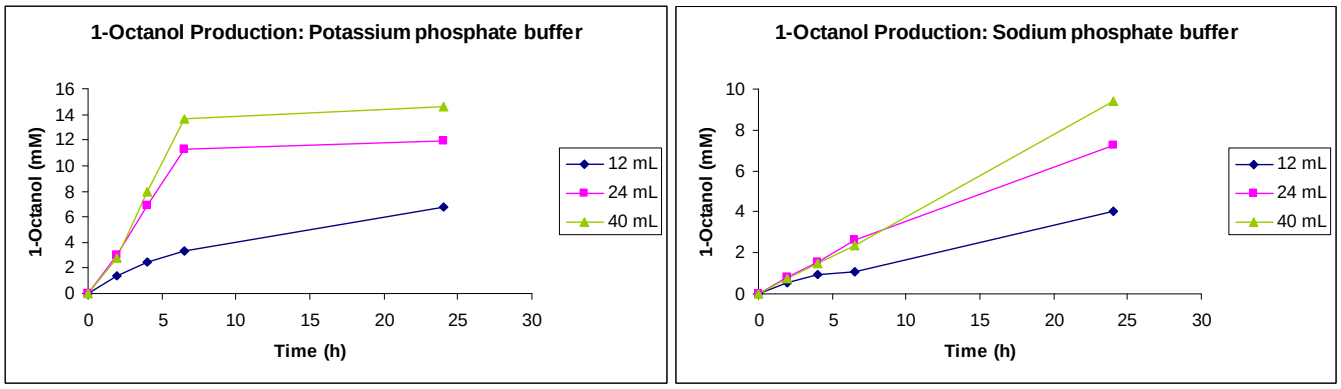


Figure 4.24 Graphs showing the production of 1-octanol over a period of 24 hours using 12-mL, 24-mL and 40-mL vials with **(a)** potassium phosphate buffer and **(b)** sodium phosphate buffer.

After a reaction period of 24 hours, it was clear that the volume of the vial, and therefore the amount of oxygen present, did play a role in the amount of product being formed. The potassium phosphate buffer reactions resulted in the production of 6.78 mM, 11.93 mM and 14.62 mM of 1-octanol with vials of 12-mL, 24-mL and 40-mL, respectively, but despite the increase in the vial volume, the formation of 1-octanol still levelled off after about 8 hours (Fig. 4.24(a)). After a 24-hour reaction period, the sodium phosphate buffer reactions resulted in the production of 4.03 mM, 7.25 mM and 9.37 mM 1-octanol using vials with volumes of 12 mL, 24 mL and 40 mL, respectively. Although the concentration of 1-octanol produced by the sodium phosphate buffer reactions was lower than that of the potassium phosphate buffer reactions, the product formation did not level off during the 24-hour bioconversion reaction period (Fig. 4.24(b)).

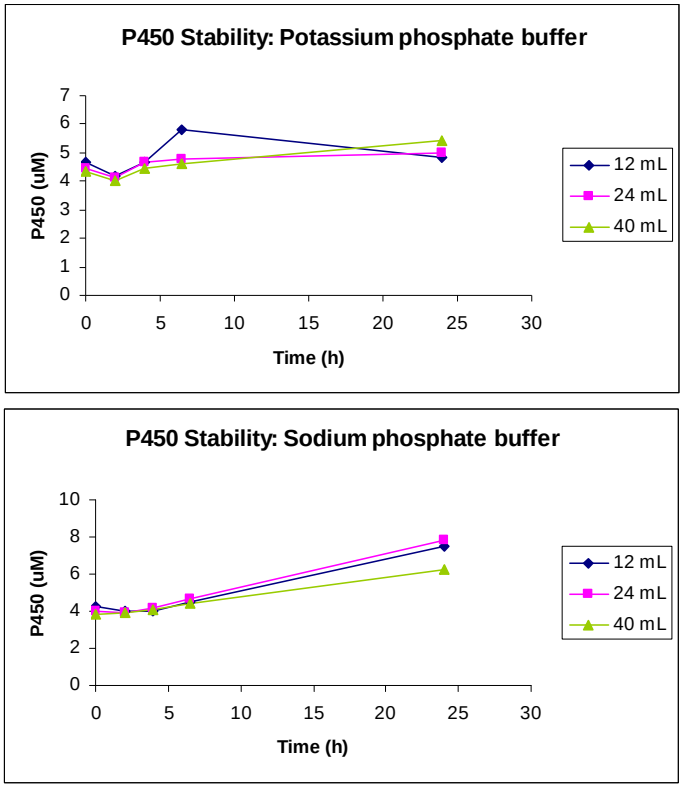


Figure 4.25 Graphs showing the P450 stability over a period of 24 hours using 12-mL, 24-mL and 40-mL vials with potassium phosphate and sodium phosphate buffers.

Whole-cell CO-difference spectra were performed using samples of the bioconversion reaction mixture in order to investigate the stability of CYP153A6 during the reaction period. These CO-difference spectra indicated that CYP153A6 was stable, with the concentration of P450 actually increasing during the 24-hour reaction period. The spectra obtained using the sodium phosphate buffer samples revealed a greater increase in P450 concentration during this period than the

increase determined from spectra obtained using the potassium phosphate buffer samples (Fig. 4.25). The product formed per 1 µM P450 after 24 hours was calculated. These results are summarised in Table 4.5.

To confirm the results obtained, the bioconversion was repeated in triplicate, using the 40-mL vials and both buffers, but the amount of 1-octanol produced did not correlate with the results obtained in the previous bioconversion experiment. The concentration of 1-octanol produced by the sodium phosphate buffer reactions after a period of 25 hours was 33.83 mM, whereas the potassium phosphate reactions resulted in the production of only 15.80 mM. Once again, the product formation of the potassium phosphate reactions levelled off after a period of 7 to 8 hours, whereas the product formation of the sodium phosphate buffer reactions did not level off during the 25-hour reaction period (Fig. 4.26(a)). As observed in the previous bioconversion experiment, the P450 concentrations increased during the reaction, with a greater increase observed with the sodium phosphate buffer samples than with the potassium phosphate buffer samples (Fig. 4.26(b)). The product formed per 1 µM P450 was calculated and these results are summarised in Table 4.6.

Table 4.5 Summary of the results obtained for the CYP153A6 bioconversion using vials with volumes of 12-mL, 24-mL and 40-mL, with potassium phosphate and sodium phosphate buffers

Vial Volume (mL)	Phosphate Buffer	P450 (µM) [a]	1-Octanol produced after 24 hours (mM)	1-Octanol produced per µM P450 after 24 hours (mM)
12	Potassium	4.84	6.78	1.40
24	Potassium	4.99	11.93	2.39
40	Potassium	5.40	14.62	2.71
12	Sodium	7.54	4.03	0.53
24	Sodium	7.87	7.25	0.92
40	Sodium	6.29	9.37	1.49

Notes: a = Cells from the bioconversion reaction mixture were 20x concentrated

Table 4.6 Summary of the results obtained for the CYP153A6 bioconversion using 40-mL vials with potassium phosphate and sodium phosphate buffers

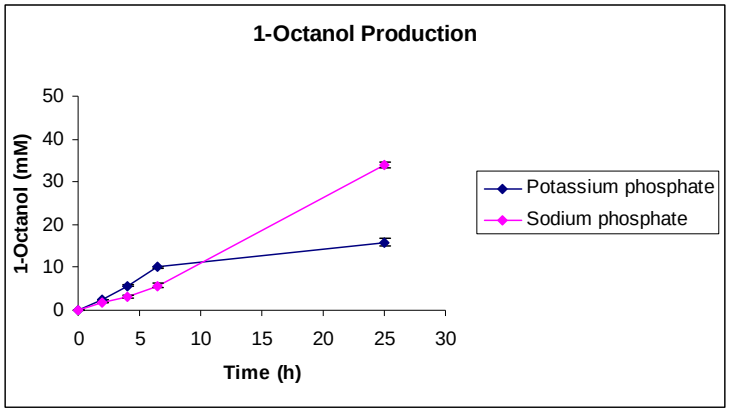
Vial Volume (mL)	Phosphate Buffer	P450 (μM) [a]	1-Octanol produced after 25 hours (mM) [b]	1-Octanol produced per μM P450 after 25 hours (mM) [c]
40	Potassium	5.75 ± 0.49	15.80 ± 0.86	2.76 ± 0.35
40	Sodium	7.11 ± 0.22	33.83 ± 0.67	4.76 ± 0.18

Notes: a = Cells from the bioconversion reaction mixture were 20x concentrated; average of triplicates; standard deviations are indicated

b = Average of triplicates; standard deviations are indicated

c = Average of triplicates; standard deviations are indicated

(a)



(b)

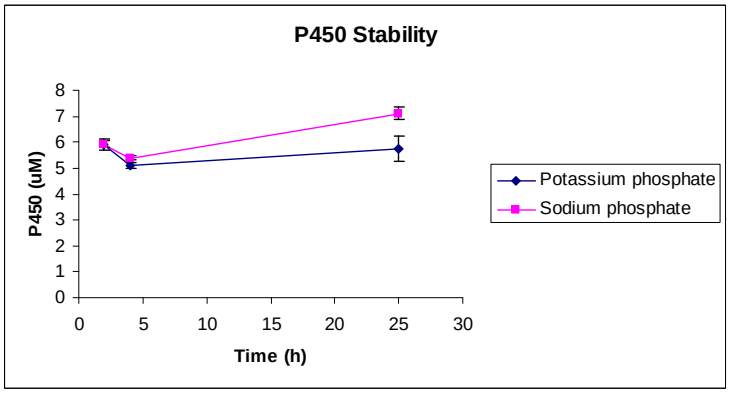


Figure 4.26 Graphs showing the **(a)** 1-octanol production and **(b)** P450 stability over a period of 25 hours using potassium phosphate and sodium phosphate buffers with 40-mL vials. Each point is the average of triplicates, with standard deviations indicated by error bars.

Because of the fact that 1-octanol was still being produced after 25 hours by the sodium phosphate buffer reactions, another bioconversion experiment was performed, in triplicate, using this buffer. Instead of monitoring the product formation for 24 or 25 hours, the reaction was followed for a period of 48 hours to determine the highest concentration of 1-octanol that could be produced under these conditions with CYP153A6. The product formation began to level off after about 30 hours (Fig. 4.27). The same trend with respect to the P450 concentration was observed as in the previous bioconversions: the P450 concentration increased during the first 24 hours. However, during the second 24-hour period the P450 concentration remained relatively constant. The average concentration of 1-octanol produced by an average P450 concentration of 4.88 ± 0.37 µM, after a reaction period of 24 hours, was 41.82 ± 3.36 mM, which translates to an average of 8.63 ± 0.33 mM of 1-octanol produced per 1 µM P450, representing the most efficient whole-cell CYP153A6 bioconversion conducted in this study. After a period of 48 hours, the average concentration of 1-octanol that was produced was 50.18 ± 3.93 mM. However, this bioconversion experiment was not reproducible as subsequent bioconversions using CYP153A6 resulted in the production of lower concentrations of 1-octanol (data not shown).

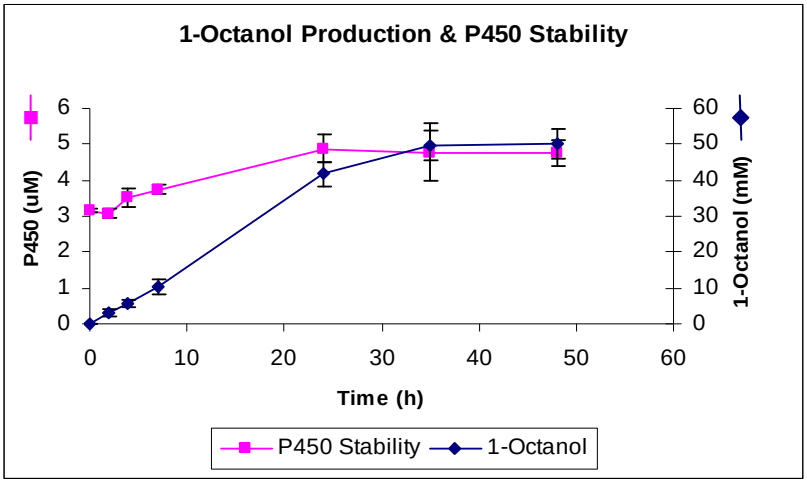


Figure 4.27 Graph showing the production of 1-octanol and the P450 concentration over a period of 48 hours using sodium phosphate buffer with 40-mL vials. Each point is the average of triplicates, with standard deviations indicated by error bars.

From the results obtained, it was clear that CYP153A6 performed better with the sodium phosphate buffer than with the potassium phosphate buffer. In order to determine whether the $FeSO_4$ present in the sodium phosphate buffer was responsible for this, a bioconversion experiment was conducted using a sodium phosphate buffer containing no $FeSO_4$, a buffer containing 100 µg.mL^{-1} $FeSO_4$ and a buffer containing 200 µg.mL^{-1} $FeSO_4$. However, similar

results were obtained with all three (data not shown), indicating that $FeSO_4$ was not responsible for the observed results.

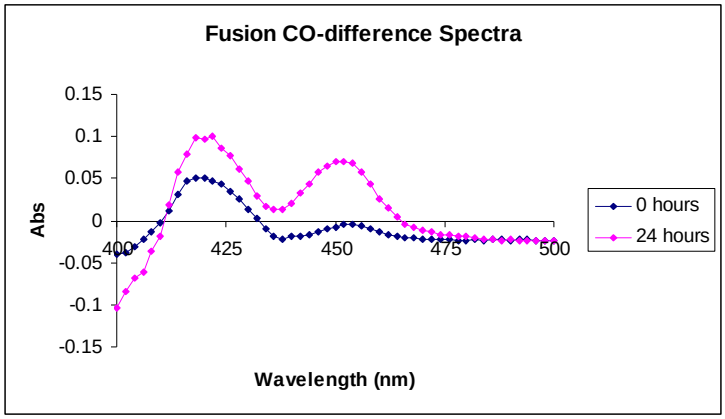
The bioconversions performed using CYP153A6 indicated that the use of 40-mL vials with sodium phosphate buffer, and the presence of glucose, gave the best results. These conditions were therefore used for the bioconversion experiments conducted during the rest of the study.

4.9.2 Bioconversions with CYP153A6/PFOR(CYP116B3) Fusions

The CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion, expressed using *E. coli* BL21(DE3)pLysE, were selected for the octane bioconversion experiment. The 0-hour and 24-hour bioconversion reaction mixtures were sampled for whole-cell CO-difference spectra to determine the enzyme content. P420 forms of both proteins were detected in the 0-hour samples, with increased concentrations being detected in the 24-hour samples. This indicated that the P420s followed the trend observed with the P450s in previous bioconversions, where there was an increase in the enzyme concentration during the 24-hour bioconversion reaction period.

An unexpected result was obtained in the form of P450s detected for both the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion in the 0-hour sample at a concentration of 0.28 µM and 0.07 µM, respectively. These concentrations also increased during the 24-hour reaction period, resulting in P450 concentrations of 1.64 ± 0.10 µM and 0.32 ± 0.22 µM for the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion, respectively (Fig. 4.28). These results are summarised in Table 4.7. Despite the presence of P450 forms, no 1-octanol formation was detected. This bioconversion experiment was repeated but the results were not reproducible as no P450 forms of the fusions could be detected in the second experiment.

(a)



(b)

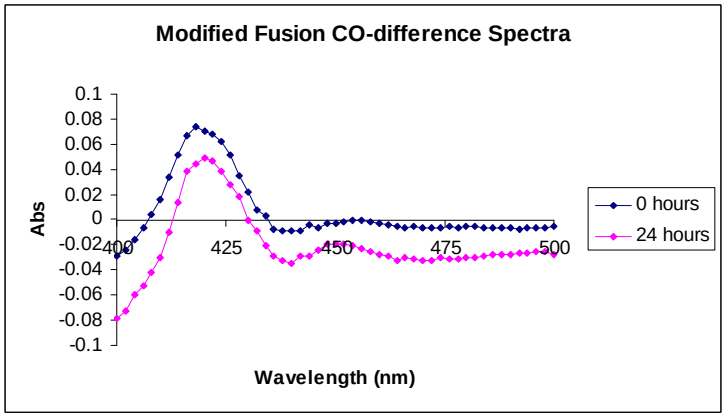


Figure 4.28 CO-difference spectra obtained with **(a)** the CYP153A6/PFOR(CYP116B3) fusion and **(b)** the modified CYP153A6/PFOR(CYP116B3) fusion after 0 hours and 24 hours.

Table 4.7 Summary of the results obtained for the bioconversion using the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion with sodium phosphate buffer and 40-mL vials

Protein	Vial Volume (mL)	Phosphate Buffer	Whole-cell enzyme content (μM) [a]			
			P420		P450	
			0 h	24 h [b]	0 h	24 h [b]
Fusion	40	Sodium	1.14	1.63 ± 0.26	0.28	1.64 ± 0.10
Modified Fusion	40	Sodium	1.19	1.41 ± 0.32	0.07	0.32 ± 0.22

Notes: a = Cells from the bioconversion reaction mixture were 20x concentrated
b = Average of duplicates; standard deviations are indicated

Following the 24-hour reaction period, 20 µL of the bioconversion reaction mixture was plated on an LB agar plate and on an LB agar plate supplemented with kanamycin, and incubated for 16 hours at 37℃. Growth occurred on both plates, indicating that the cells were still viable, despite the high octane concentration (1.5 M) in the bioconversion reaction mixtures. Some strains of *E. coli* have been shown to be resistant to organic solvents including cyclohexane, hexane and decane (Aono *et al.*, 2001). It is generally accepted that the toxicity of a solvent is inversely correlated with its log P_{ow} value, which is defined as the partition coefficient of the solvent between *n*-octanol and water layers. Octane has a higher log P_{ow} value than hexane and cyclohexane and therefore is theoretically less toxic to cells than these two organic solvents.

4.10 Amplification of Internal *CYP153A* Gene Fragments from Environmental DNA

The second aim of the study was to amplify the internal fragments of *CYP153A* genes from environmental DNA and use these fragments for cassette PCR (Kubota *et al.*, 2005). The internal fragments of the *CYP153A* genes were PCR amplified from total DNA extracted from enrichments of diesel-contaminated soil. The degenerate primers CF and P450rv3 were used for this reaction. The forward primer, CF, was designed by Kubota *et al.* (2005) according to a relatively well-conserved region near the N-terminus of the CYP153A subfamily of P450s (Fig. 4.29(a)) and was used with the reverse primer, RF, designed according to the conserved heme-binding region near the C-terminus, for the amplification of the internal fragments of *CYP153A* genes from environmental DNA. However, in this study, the use of primer RF did not result in any amplification (results not shown). Primer P450rv3 was therefore selected for use with primer CF. This primer was designed by van Beilen *et al.* (2006) for the amplification of the internal fragments of *CYP153* genes from alkane-degrading bacteria. It was designed according to the heme-binding region (Fig. 4.29(b)), overlapping the sequence used to design primer RF.

(a)

```
CYP153A26    53  YSQGNEDTGPYWSVTRYNDIMSVDTNHQVFSSE---GGITLRDQDED---FKLPMFIAMDPP 108
CYP153A2     64  YSKGDEEVGPYWSVTRYNDIMTVDTTHQVFSSDAHLGGITIRNFDED---FVLPMFIAMDQP 122
CYP153A6     51  YCKSS-MFGPYWSVTKYRDIMAVETNPKVFSSEAKSGGITIMDDNAA---ASLPMFIAMDPP 108
CYP153A11    55  YCSES-AFGPYWSITRYNDIMAVDTNHKLFSSEAKLGGIAIQDMHNDATNLELEMFIAMDQP 115
CYP153A8     50  YCRES-YVGPYWSITKFDDIMAVDTNHKVFSSEAKLGGIAIEDMHSAKSALELEMFIAMDPP 110
CYP153A29    47  YCPES-AYGPYWSVTKFNDIMQVEVNHQTFSSEAKLGGIALQDMQSGEAALELEMFIAMDPP 107
CYP153A13a   96  YQKNS-AFGPFWSVTRYEDIVFVDKSHDLFSAE---PQIILGDPPEG---LSVEMFIAMDPP 150
CYP153A1    123  YQANS-PFGAFWSVTRYDDIVYVDKNHEIFSAE---PVIAIGNTPPG---LDAEMFIAMDPP 177
CYP153A16    88  YQAHS-AFGPFWSVTRHADIVAVDKNHEVFSSE---PFIVIGSPPRF---LDIAMFIAMDPP 142
```

(b)

```
CYP153A26   349 VIDRERPRQHLSFGFGIHRCVGNRLAEMQLRIIWEEILKRWDQP--IRVLSEPQRVHSSFVK 408
CYP153A2    363 IVDRPNARRHLSFGFGIHRCVGNRLAEMQLKIVWEEILKRFPK---IEVLEEPKRVYSTFVK 421
CYP153A6    348 IIDRAKPRQHLSFGFGIHRCVGNRPAELQLNILWEEILKRWPDPLQIQVLQEPTRVLSPFVK 409
CYP153A11   355 IIDRKNPRQHLSFGYGIHRCMGNRLAELQLRIIWEEIHKRFRL---VEMVGEPERLLSNLVR 413
CYP153A8    350 IIDRKNPRHHLSFGYGIHRCMGNRLAELQLRIIWEEIHKRFAK---IEVTGEPERLFSNLVR 408
CYP153A29   347 IINRPNSRHHLSFGFGVHRCMGNRVAELQLRIIWEEILKRFSK---VEVVGAPERTLSNFIR 405
CYP153A13a  398 IIDRKDTRNHVSFGYGVHRCMGNRLAELQLRILWEELLPRFEN---IEVIGEPERVQSNFVR 456
CYP153A1    426 IIDRKGARNHLSFGFGVHRCMGNRLAEMQLRILWEELLQRFEN---IEVLGEPEIVQSNFVR 484
CYP153A16   390 IIDRANARNHISFGFGVHRCMGNRLAEMQLRILWEELLPRFEN---IEVVGEPEYVQSNFVR 448
```

Figure 4.29 Sections of the alignment of selected CYP153A amino acid sequences showing **(a)** the conserved N-terminal region according to which primer CF was designed (indicated by the blue box) and **(b)** the conserved heme-binding region, ending with the absolutely conserved heme-binding cysteine residue, according to which primer P450rv3 was designed (indicated by the black box). Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

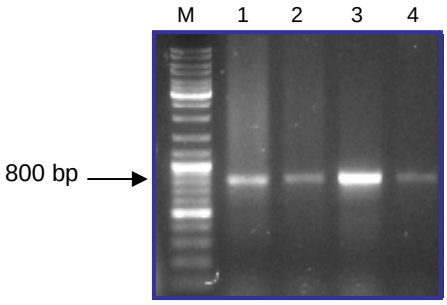


Figure 4.30 Amplification of the internal *CYP153A* gene fragments from environmental DNA using primers CF and P450rv3. 0.8% agarose gel stained with GoldView Nucleic Acid Stain showing the PCR products (Lanes 1 to 4). M = GeneRuler™ DNA Ladder Mix.

Amplicons of approximately 800 bp were expected for the amplification of the internal fragments of *CYP153A* genes from the environmental DNA and products of the appropriate size were obtained (Fig. 4.30). These products were sub-cloned using pGem-T Easy. Plasmid DNA extracted from positive transformants was digested with *Eco*RI to confirm the size of the inserts. *Eco*RI cleaves pGem-T Easy on both sides of the insert, resulting in fragments of

approximately 3 000 bp and 800 bp. Plasmids with inserts of the correct size were digested with *Hae*III, which is a four-base cutter, with the aim of obtaining different digestion profiles that would give an estimation of the number of different *CYP153A* sequences present amongst the clones. Three different digestion profiles were obtained (Fig. 4.31).

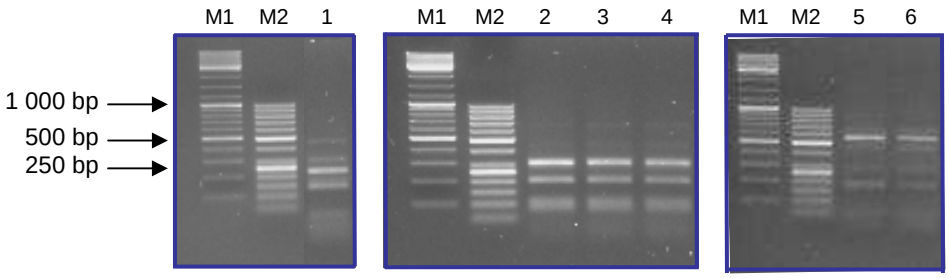


Figure 4.31 Profiles obtained from the digestion of the pGem-T Easy plasmids and their inserts with *Hae*III. The first profile is represented by Lane 1, the second by Lanes 2 to 4 and the third by Lanes 5 and 6. 2% (w/v) agarose gel stained with GoldView Nucleic Acid Stain. M1 = GeneRuler^{TM} DNA Ladder Mix; M2 = GeneRuler™ 50 bp DNA Ladder.

Plasmids representing each profile were selected for sequencing and amongst these plasmids three different insert sequences were identified. A representative of each sequence was selected. These plasmids were referred to as Aliwal03, Aliwal09 and Aliwal11. A BLAST search was performed using the inserts of these plasmids (http://www.ncbi.nlm.nih.gov/blast/BLAST.cgi) and the results obtained revealed that the translated Aliwal11 insert was 99% identical to the internal fragment of CYP153A6 (Fig. 4.32). This sequence was excluded from the rest of the study. Twenty-five percent of the plasmids contained the *CYP153A6* internal gene fragment. The translated insert of Aliwal09 was 99% identical to the internal fragment of a P450 from *Rhodococcus erythropolis* PR4 (Fig. 4.33), with this sequence representing 70.8% of the clones, indicating that this gene was probably the most abundant *CYP153A* gene present in the environmental DNA. The least abundant sequence amongst the clones was represented by the Aliwal03 insert, with the translated insert sequence sharing 91% amino acid identity with a CYP153A internal fragment from an uncultured bacterium isolated by Kubota *et al.* (2005) from petroleum-contaminated soil (Fig. 4.34). Only 4.2% of the plasmids contained this sequence, indicating that this gene was probably the least abundant *CYP153A* gene present in the environmental DNA.

```
Aliwal11  1   KHDVQRKTVSPIVAPENLATMESVIRQRTADLLDGLPINEEFDWVHRVSIDLTTKMLATLFDFP 64
CYP153A6  1   KHDVQRKTVSPIVAPENLATMESVIRQRTADLLDGLPINEEFDWVHRVSIELTTKMLATLFDFP 64

Aliwal11  65  WDDRAKLTRWSDVTTALPGGGIIDSEEQRMAELMECATYFTELWNQRVNAEPKNDLISMMAHSE 128
CYP153A6  65  WDDRAKLTRWSDVTTALPGGGIIDSEEQRMAELMECATYFTELWNQRVNAEPKNDLISMMAHSE 128

Aliwal11  129 STRHMAPEEYLGNVVLLIVGGNDTTRNSMTGGVLALNEFPDEYRKLSANPALISSMVSEIIRWQ 192
CYP153A6  129 STRHMAPEEYLGNIVLLIVGGNDTTRNSMTGGVLALNEFPDEYRKLSANPALISSMVSEIIRWQ 192

Aliwal11  193 TPLSHMRRTALEDIEFGGKHIRQGDKVVMWYVSGNRDPEAIDNPDTFIIDRAKPRQ 248
CYP153A6  193 TPLSHMRRTALEDIEFGGKHIRQGDKVVMWYVSGNRDPEAIDNPDTFIIDRAKPRQ 248
```

Figure 4.32 Alignment of the translated nucleotide sequence of the *CYP153A* insert from Aliwal11 and the amino acid sequence of the internal fragment of CYP153A6. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

```
Aliwal09     1   KHDAQRAAVQGVVAPKNLREMEALIRSRVQEVLDELPIGEPFNWVDLVSIELTARMLATLL 61
PR4 CYP153A  1   KHDAQRAAVQGVVAPKNLREMEALIRSRVQEVLDELPIGEPFNWVDLVSIELTARMLATLL 61

Aliwal09     62  DFPYDQRRKLVEWSDLATAMEQANGGPSDNDEIFRGFVDAARGLSAHWHDKAARLAVGEEP 122
PR4 CYP153A  62  DFPYDQRRKLVEWSDLATAMEQANGGPSDNDEIFRGFVDAARGLSAHWHDKTARLAAGEEP 122

Aliwal09     123 GFDLITMLQSNEDTKDLIDRPMEFLGNLILLIVGGNDTTRNSMSGGVLALNQFPEQFEKLK 183
PR4 CYP153A  123 GFDLITMLQSNEDTKDLIDRPMEFLGNLILLIVGGNDTTRNSMSGGVLALNQFPEQFEKLK 183

Aliwal09     184 ANPDLIPNMVSEVIRWQTPLAYMRRVAKKDTVLNGQFIRKGDKVVMWYASGNRDERVFERP 244
PR4 CYP153A  184 ANPDLIPNMVSEVIRWQTPLAYMRRVAKKDTVLNGQFIRKGDKVVMWYASGNRDERVFERP 244

Aliwal09     245 DELIIDRKNARN 256
PR4 CYP153A  245 DELIIDRKNARN 256
```

Figure 4.33 Alignment of the translated nucleotide sequence of the *CYP153A* insert from Aliwal09 and the amino acid sequence of the internal fragment of a P450 from *Rhodococcus erythropolis* PR4. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

```
Aliwal03    1   RHDAQRAAVQGVVAPKNLREMEGLIRSRVQEVLDNLPVDQPFDWVQNVSVELTARMLATLLD 62
Niigata004  1   RHDKQRAAVQGVVAPKNLREMEGLIRSRVQEVLDNLPVDQPFDWIQNVSIELTARMLATLLD 62

Aliwal03    63  FPYEQRRKLAYWSDLASSMEQANGGPSDNDEVFRGMRDMARGLSTLWRDKAARTAAGEEPGF 124
Niigata004  63  FPYEQRRKLVEWSDLATSMEQANGGPSDLDDTFAGMRDMARGLSEHWHDKAARRAAGEEPGF 124

Aliwal03    125 DLITLLQSNEDTKDLIDRPMEFLGNLVLLIVGGNDTTRNSMSGGVLALNRFPDQFEKLKANP 186
Niigata004  125 DLITMLQSNESTKDLIKRPMEFLGNLVLLIVGGNDTTRNSMSGGVLALNRYPDQFEKLKANP 186

Aliwal03    187 DLIPNMVSEIIRWQTPLAYMRRIAKADTVLNGQFIRKGDKVVMWYASRNRDERMFDRPDDFI 248
Niigata004  187 DLIPNMVSEVIRWQTPLAYMRRIAKADTMLNGQFIRKGDKVVMWYASGNRDERVFDRPDDLI 248

Aliwal03    249 IDRANARN 256
Niigata004  249 IDRANARN 256
```

Figure 4.34 Alignment of the translated nucleotide sequence of the *CYP153A* insert from Aliwal03 and the amino acid sequence of the internal fragment of a CYP153A isolated from petroleum-contaminated soil by Kubota *et al.* (2005). Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

The Aliwal03 and Aliwal09 plasmids and two of the plasmids constructed using *CYP153A* gene fragments obtained from the Beatrix Goldmine metagomic library, referred to as Beatrix14 and Beatrix16 were selected for cassette PCR (Abbai, 2009). The translated inserts of Beatrix14 and Beatrix16 share 76% and 82% amino acid identity, respectively, with the internal fragment of CYP153A7 from *Sphingopyxis macrogoltabida*.

4.11 Cassette PCR

4.11.1 First PCR: Amplification of the Internal *CYP153A* Gene Fragments and the 5'- and 3'-end Fragments of the *CYP153A6* Gene

Cassette PCR is a method used for isolating environmental genes without having to isolate microorganisms and consists of two PCR steps. During the first PCR step, the internal fragments of genes from environmental samples and the 5'- and 3'-ends of a selected gene are amplified, generating complementary ends which anneal during the second PCR step, allowing the generation of full-length genes.

In order to perform cassette PCR using the environmental *CYP153A* gene fragments, the plasmid inserts of Aliwal03, Aliwal09, Beatrix14 and Beatrix16 were reamplified using the degenerate primers CF and P450rv3. Amplicons of approximately 800 bp were obtained and these were purified.

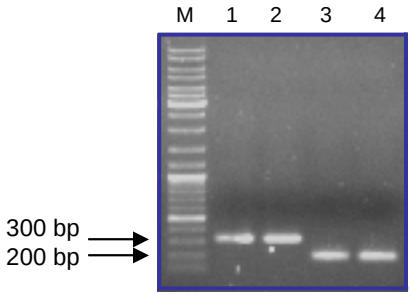


Figure 4.35 Amplification of the 5'- and 3'-ends of the *CYP153A6* gene. 0.8% agarose gel stained with GoldView Nucleic Acid Stain. Lanes 1 and 2, 5'-end PCR product amplified using primers pCom-153F and 5R; Lanes 3 and 4, 3'-end PCR product amplified using primers 3F and pCom-153R. M = GeneRuler^{TM} DNA Ladder Mix.

The 5'- and 3'-ends of the *CYP153A6* gene, referred to as "arms", were amplified from pET28b-CYP153A6_FdR_Fdx using primers pCom-153F and 5R, and 3F and pCom-153R, respectively, with the sequence of primer 5R being complementary to CF and that of 3F being complementary

to P450rv3. The expected size of the 5'-arm amplicon was approximately 300 bp while that of the 3'-arm was approximately 200 bp. Products of the appropriate sizes were obtained (Fig. 4.35) and these were purified.

4.11.2 Second PCR: Assembly of the Chimeric *CYP153A* Genes

The purified products of the first PCR step were mixed and subjected to PCR using primers pCom-153F and pCom-153R. The complementary ends of the products allowed the fragments to anneal, forming full-length chimeric genes of the expected size of approximately 1 300 bp (Fig. 4.36), which were then sub-cloned using pGem-T Easy.

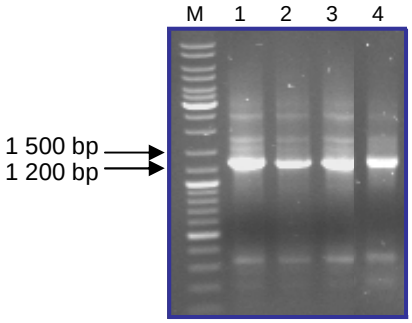


Figure 4.36 Assembly of the chimeric *CYP153A* genes using primers pCom-153F and pCom-153R (Lanes 1 to 4). M = GeneRuler^{TM} DNA Ladder Mix.

The initial plan was to apply cassette PCR to the *CYP153A6*/*PFOR(CYP116B3)* fusion gene to generate genetic diversity which would result in diverse self-sufficient CYP153A proteins. These could then serve as a starting point for directed evolution. However, the expressed fusion was not performing as expected and it was therefore decided that the functionality of the constructed chimeric genes would be investigated using the CYP153A6 operon. In order to do this the *CYP153A6* gene had to be excised from pET28b-CYP153A6_FdR_Fdx and be replaced with the chimeric genes. However, there was no restriction site at the end of the *CYP153A6* gene, so a *Pst*I site close to the end of the gene, positioned within the 3'-arm region, was selected for this purpose. A double digest was performed with *Nde*I and *Pst*I and the excised fragment was replaced with the chimeric gene fragments from the Aliwal03, Aliwal09, Beatrix14 and Beatrix16 plasmids, which were excised from the pGem-T Easy backbone using the same restriction enzymes. The resulting operons were expressed in *E. coli* BL21(DE3)pLysE.

4.12 Analyses of the Expressed Chimeras

SDS-PAGE analysis was performed using the soluble and insoluble fractions obtained after cell disruption. The expression products, also referred to as Aliwal03, Aliwal09, Beatrix14 and Beatrix16, including N-terminal His-tags, consisted of 447, 430, 440 and 439 amino acids, respectively. These products had expected molecular weights of approximately 50 kDa. A band corresponding to this molecular weight was observed in only the insoluble fraction for all four chimeras, indicating the formation of inclusion bodies (Fig. 4.37).

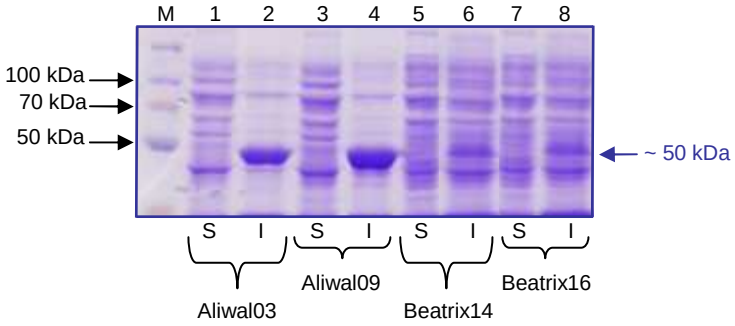


Figure 4.37 SDS-PAGE of the soluble and insoluble fractions of *E. coli* BL21(DE3)pLysE strains transformed with pET28 plasmids for the expression of the CYP153A chimeras. “S” indicates the soluble fraction and “I” indicates the insoluble fraction. Lanes 1 and 2, Aliwal03 (~ 50 kDa); Lanes 3 and 4, Aliwal09 (~ 50 kDa); Lanes 5 and 6, Beatrix14 (~ 50 kDa); Lanes 7 and 8, Beatrix16 (~ 50 kDa). M = Spectra™ Multicolour Broad Range Protein Ladder.

CO-difference spectra performed using whole cells and the soluble fractions indicated the presence of P420 forms, but no P450s. The concentration of P420s calculated in the soluble fractions did not correlate with the observed expression (Table 4.8).

The reductase activities detected for the FdR-Fdx couple of Aliwal03, Aliwal09 and Beatrix14 were approximately half of that detected for the FdR-Fdx couple of CYP153A6 expressed with the pET28 plasmid. In the case of Beatrix16 the reductase activity could not be detected. These results are summarised in Table 4.8.

Table 4.8 Summary of the results obtained for the expression of the CYP153A chimeras

Protein	Expression [a]		Whole-cell enzyme content (μM) [b]		Soluble enzyme content (pmol.mg protein^{-1})		Soluble Reductase Activity (nmol.min^{-1}mg protein^{-1}) [c]
	Soluble	Insoluble	P420	P450	P420	P450	
Aliwal03	-	+++	2.018	0	2.66	0	0.9 ± 0.12
Aliwal09	-	+++	1.451	0	4.86	0	1.40 ± 0.38
Beatrix14	-	+	1.228	0	2.10	0	0.93 ± 0.38
Beatrix16	-	+	1.182	0	3.41	0	n.d.

Notes: a = Observed level of expression from SDS-PAGE analysis: +++ = High; ++ = Intermediate; + = Low; - = No expression

b = 18x concentrated *E. coli* culture (final OD_{600} approximately 50)

c = Average of triplicates; standard deviations are indicated

n.d. = Not detected

4.13 Bioconversion using the Expressed Chimeras

Even though the chimeras were expressed only in the P420 form, in the insoluble fraction, a bioconversion experiment was conducted to confirm that these proteins could not hydroxylate octane. The concentration of P420 forms detected in the 0-hour samples, as calculated using whole-cell CO-difference spectra, amounted to approximately 1 μM and this concentration decreased during the 24-hour incubation period to approximately 0.7 μM (data not shown), but no P450 forms were detected and no 1-octanol was produced. This confirmed that these chimeras were non-functional.

Chapter 5

Discussion

Cytochrome P450 monooxygenases are heme-containing enzymes that are able to catalyse diverse reactions in a stereo- and regioselective manner. Many of these reactions are difficult to catalyse using chemical catalysts, making these enzymes potentially useful in industry, particularly for the synthesis of chemicals. One such reaction which is particularly interesting is the terminal hydroxylation of *n*-alkanes. However, P450s have a number of shortcomings which currently limit their application in industry, including low stability and low activity. Some research groups have used directed evolution to try and improve these limiting properties or to create new properties. One approach that has been used for generating the genetic diversity required for directed evolution is chimeragenesis. A number of methods that have been developed for creating chimeric proteins have successfully been used to generate chimeric P450s (Chapter 2). The goal of this study was to use existing methods to generate two types of chimeric P450s, the first being the construction of a fusion between domains from two different P450s, of which there have been many reports in literature, in some cases resulting in the solubilisation of membrane-bound P450s (Shimoji *et al.*, 1998; Gilardi *et al.*, 2002; Sukumaran *et al.*, 2002), the generation of thermostable P450s (Eiben *et al.*, 2007) or the generation of artificial self-sufficient P450s (Fisher *et al.*, 1992; Harlow & Halpert, 1996; Deeni *et al.*, 2001; Li *et al.*, 2007). The second method that was used was cassette PCR, a method which has been applied to P450s in only one report (Kubota *et al.*, 2005).

Nodate *et al.* (2006) constructed a fusion between the PFOR-type reductase domain of P450RhF (CYP116B2), a self-sufficient P450 from *Rhodococcus* sp. NCIMB 9784, and the heme domain of $P450_{balk}$ (CYP153A13a), a terminal alkane hydroxylase from *Alcanivorax borkumensis* SK2. The fusion gene was expressed in *E. coli* BL21(DE3) and the resulting $P450_{balk}$/PFOR(P450RhF) fusion protein was a soluble, self-sufficient terminal alkane hydroxylase. The heme domains of four other P450s, $P450_{cam}$ (CYP101A), $P450_{bzo}$ (CYP203A), PikC and $P450_{EryF}$, have also successfully been fused to the PFOR domain of P450RhF, resulting in self-sufficient P450s (Nodate *et al.*, 2006; Li *et al.*, 2007; Robin *et al.*, 2009). Kubota *et al.* (2005) used cassette PCR to construct 8 new catalytically active chimeric self-sufficient

CYP153As based on the $P450_{balk}$/PFOR(P450RhF) fusion described above. Because of the success that was achieved by these authors, the same strategy was followed in this study. The first aim of the study was to construct a self-sufficient terminal alkane hydroxylase by fusing the heme domain of CYP153A6, a terminal alkane hydroxylase from *Mycobacterium* sp. HXN-1500 to the PFOR domain of CYP116B3 from *Rhodococcus ruber* DSM 44319 (Liu *et al.*, 2006). This PFOR domain shares 89% amino acid identity with the P450RhF PFOR domain of *Rhodococcus* sp. NCIMB 9784, while CYP153A6 shares 51% amino acid identity with $P450_{balk}$. The second aim was to use cassette PCR to construct CYP153A chimeras based on the envisaged CYP153A6/PFOR(CYP116B3) fusion.

5.1 Aim 1: Construction of a Self-Sufficient Terminal Alkane Hydroxylase

The gene encoding the heme domain of CYP153A6 from *Mycobacterium* sp. HXN-1500 was amplified without its stop codon from the pCom8-PFR1500 plasmid. This gene was ligated upstream of the DNA encoding the PFOR reductase domain and linker region of CYP116B3 from *Rhodococcus ruber* DSM 44319, in frame with the sequence encoding the N-terminal His-tag, resulting in the *CYP153A6*/*PFOR(CYP116B3)* fusion gene. This fusion gene was expressed in *E. coli* BL21(DE3)pLysE. The only difference between this strain and the BL21(DE3) strain used by Nodate *et al.* (2006) for expression of the *$P450_{balk}$/PFOR(P450RhF)* fusion gene is the pLysE plasmid that produces T7 lysozyme, reducing basal level expression of the gene of interest. CO-difference spectra indicated the presence of only misfolded or P420 forms of the CYP153A6/PFOR(CYP116B3) fusion protein at a concentration of 0.038 µM in the whole cells. No P450s or P420s were detected in the soluble fraction and the reductase activity was very low when compared to the detected reductase activity of CYP116B3 and the CYP116B3 PFOR domain. In addition, SDS-PAGE analysis revealed that the CYP153A6/PFOR(CYP116B3) fusion was only expressed in the insoluble fraction in the form of inclusion bodies, despite the expectation that it would be expressed in the soluble fraction, based on the fact that both CYP153A6 and CYP116B3 are soluble P450s (Hannemann *et al.*, 2007). Despite the high level of expression observed in the insoluble fraction, the concentration of P420s detected in the whole cells was not very high, which indicated that the protein expressed as inclusion bodies was not detected by CO-difference spectra.

The observation that the CYP153A6/PFOR(CYP116B3) fusion was not expressed as a properly folded soluble P450 protein raised the following questions:

- Can CYP153A6 be expressed in *E. coli* and can it be expressed using a pET28 plasmid?
- Does the N-terminal His-tag affect protein folding?
- Is the sequence of the linker region affecting the folding of the fusion protein?
- Are the expression conditions optimal?
- Does the presence of rare codons affect expression?
- Is the protein being expressed too quickly to fold correctly?
- Is it possible to obtain whole-cell bioconversion of *n*-octane with strains expressing the CYP153A6 operon or the CYP153A6/PFOR(CYP116B3) fusions?

The conclusions that could be drawn from the experiments that were performed to answer these questions are discussed below.

CYP153A6 can be expressed in *E. coli* using a pET28 plasmid

When this study commenced, CYP153A6 had only been successfully expressed in *Pseudomonas putida*, using the pCom8 plasmid (van Beilen *et al.*, 2005; Funhoff *et al.*, 2006 & 2007). In order to confirm that CYP153A6 could be expressed in *E. coli* and that it could be expressed using a pET28 plasmid, which adds an N-terminal His-tag to the protein, the genes encoding the heme domain of the natural CYP153A6 and its redox partners, ferredoxin reductase and ferredoxin, were amplified from the pCom8-PFR1500 plasmid as a single amplicon. This was ligated to the pET28b(+) plasmid and expressed in *E. coli* BL21(DE3)pLysE. Bands corresponding to the expected molecular weight of CYP153A6 were observed in both the soluble and insoluble fractions, but bands corresponding to the expected molecular weights of the ferredoxin and the ferredoxin reductase could not be observed. CO-difference spectra performed with both whole cells and the soluble fraction revealed that CYP153A6 was correctly folded, with P450 concentrations of 0.446 µM and 11.03 pmol.mg protein^{-1}, respectively, being detected. There was variability between experiments that were performed on different occasions and expression experiments that were later conducted in preparation for *n*-octane bioconversions resulted in CYP153A6 concentrations as high as 4 µM with whole cells, which is similar to the concentration detected for the expression of AciA, a CYP153A from *Acinetobacter* sp. OC4 (Fujii *et al.*, 2006). This confirmed that CYP153A6 could successfully be expressed in *E. coli* using a pET28 plasmid. This is the first report of a CYP153A expressed with an N-terminal His-tag, since AciA was expressed without an N-terminal His-tag.

The effect of an N-terminal His-tag on the protein folding, expression and reductase activity is unpredictable

The fusion genes that were constructed by Nodate *et al.* (2006) between the DNA encoding the P450RhF PFOR domain and the genes encoding the heme domains of different P450s were expressed without N-terminal His-tags. All of the P450 genes that were expressed in this study using pET28 plasmids were ligated in frame with the sequence encoding the N-terminal His-tag. In order to determine whether this His-tag had any effect on the folding of the expressed CYP153A6/PFOR(CYP116B3) fusion, the *CYP153A6*/*PFOR(CYP116B3)* fusion gene was expressed using the pET22b(+) plasmid, which does not contain an N-terminal His-tag sequence. Expression of the CYP153A6/PFOR(CYP116B3) fusion protein was once again only observed in the insoluble fraction and only P420 forms of the protein were detected in the whole cells at a concentration of 0.525 μM, and in the soluble fraction at a concentration of 0.50 pmol.mg protein^{-1}. This indicated that the absence of the N-terminal His-tag enhanced the level of expression of the CYP153A6/PFOR(CYP116B3) fusion in the soluble fraction, but did not improve the formation of P450 protein. The reductase activity also increased dramatically, being almost 12-fold higher than that detected using the pET28 plasmid. Similar results were obtained with CYP116B3. The level of expression and the detected reductase activity using the pET22b(+) plasmid was enhanced and a higher P450 concentration was detected in both the whole cells and the soluble fraction.

The results obtained for the CYP116B3 PFOR domain and the CYP153A6 operon, however, did not correlate with those obtained with the CYP153A6/PFOR(CYP116B3) fusion and CYP116B3. In the absence of the N-terminal His-tag, no expression of the CYP116B3 PFOR domain or the CYP153A6 operon was observed in the soluble or insoluble fractions, and the reductase activity detected for the CYP116B3 PFOR domain was lower than that detected using the pET28 plasmid. The reductase activity of the CYP153A6 FdR-Fdx couple was undetectable. In addition, only P420 forms of CYP153A6 could be detected in the whole cells, indicating a misfolded protein. No P450 or P420 forms of CYP153A6 could be detected in the soluble fraction. Sequencing confirmed that there were no errors in these constructs.

From the results obtained it is clear that the effect of an N-terminal His-tag on the level of expression, the folding and the reductase activity of a P450 protein expressed in *E. coli* is very unpredictable. Therefore, in future work, the expression of CYP153A fusions should be carried out with and without His tags.

The P450RhF (CYP116B2) linker sequence does not enhance protein expression and folding

The linker region is a short amino acid sequence that links the heme and reductase domains together in both natural and artificial fusions, and in the latter, often originates from a P450 that is unrelated to the P450 from which the heme domain originates. The PFOR domain of P450RhF (CYP116B2) was not only successfully used for constructing the catalytically active $P450_{balk}$/PFOR(P450RhF) fusion, but was also used to construct catalytically active fusions with the heme domains of $P450_{cam}$ (CYP101A), $P450_{bzo}$ (CYP203A), PikC and $P450_{EryF}$, which share less than 40% amino acid identity with $P450_{balk}$ (Nodate *et al.*, 2006; Li *et al.*, 2007; Robin *et al.*, 2009). Robin and co-workers reported a 20-fold increase in the bioconversion of (+)-camphor to 5-*exo*-hydroxycamphor by the $P450_{cam}$/PFOR(P450RhF) fusion as a result of modifying the linker region between the $P450_{cam}$ heme domain and the P450RhF PFOR domain (Robin *et al.*, 2009). Govindaraj & Poulos (1996) found that the length of the linker region of $P450_{BM3}$, the self-sufficient P450 from *Bacillus megaterium*, was important for the activity of the enzyme. All of this indicates that the linker region is very important for the activity of a P450. One of the possibilities that had to be considered was that the linker between the CYP153A6 heme domain and the CYP116B3 PFOR domain was influencing the expression of the CYP153A6/PFOR(CYP116B3) fusion. CYP116B3 is 93% identical to P450RhF, with one of the most variable regions being the linker region, with five amino acids out of sixteen differing between the linkers of these two proteins. In order to determine whether these five amino acids are critical for the functional expression of PFOR-type fusion proteins site-directed mutagenesis was used to mutate the linker sequence of the CYP116B3 PFOR domain to match the linker sequence of the P450RhF PFOR domain. Expression of this modified CYP153A6/PFOR(CYP116B3) fusion was extremely low and only P420 forms of the protein were detected (0.003 µM in whole cells). The detected reductase activity, however, did not correlate with the very low P420 levels as it was slightly higher than that detected for the original CYP153A6/PFOR(CYP116B3) fusion protein.

Changing the sequence of the linker region of the CYP153A6/PFOR(CYP116B3) fusion to match that of P450RhF resulted in a decreased level of expression in the insoluble fraction and a slight increase in the reductase activity in the soluble fraction, but expression of the modified CYP153A6/PFOR(CYP116B3) fusion in the soluble fraction was not significantly improved and it was not expressed in the P450 form. These results, however, do not completely rule out the possibility that the changed linker sequence might have a positive effect on the folding of the protein.

The expression conditions require further optimisation

Another factor that was considered was whether the expression conditions were responsible for the formation of inclusion bodies as CYP116B3, the CYP116B3 PFOR domain and CYP153A6 were expressed in both the soluble and insoluble fractions, indicating that the CYP153A6/PFOR(CYP116B3) fusion was not the only protein being expressed in the form of inclusion bodies.

Inclusion bodies are insoluble protein aggregates which form when large amounts of proteins are rapidly produced in prokaryotic hosts and, as a result, are irregularly or incompletely folded (Ventura & Villaverde, 2006). The extent of aggregation usually depends on a combination of different factors including the composition of the growth media, growth and post-induction temperatures, the presence of chaperone proteins and the rate of protein production, with the latter being dependent on the strength of the promoter, the inducer concentration, mRNA (messenger RNA) stability and codon usage. These factors can be modified to minimise the formation of inclusion bodies, but the most successful approach has been the overexpression of chaperones along with the recombinant gene. However, in this study only the post-induction temperature and the rate of protein production were investigated in an attempt to lower the protein expression levels and thereby obtain a correctly folded CYP153A6/PFOR(CYP116B3) fusion protein.

The post-induction temperature that was used for the first expression experiments was 20℃. This temperature was used by Nodate *et al.* (2006) for the expression of the active $P450_{balk}$/PFOR(P450RhF) fusion in *E. coli* BL21(DE3). The temperature was then increased to 25℃, which was the post-induction temperature used by Nodate *et al.* (2006) for the expression of the $P450_{cam}$/PFOR(P450RhF) and $P450_{bzo}$/PFOR(P450RhF) fusions, but the expression of the CYP153A6/PFOR(CYP116B3) fusion was still mainly observed in the insoluble fraction and only a low level of P420 forms was detected in the soluble fraction, while expression of CYP116B3, the CYP116B3 PFOR domain and CYP153A6 was unchanged. The post-induction temperature was then increased to 30℃, which is 2℃ higher tha n the post-induction temperature used by Fujii *et al.* (2006) for the expression of AciA from *Acinetobacter* sp. OC4. The expression of the CYP153A6/PFOR(CYP116B3) fusion remained unchanged and expression of CYP116B3, the CYP116B3 PFOR domain and CYP153A6 was still observed in both the soluble and insoluble fractions.

The concentration of IPTG used for induction was lowered from 0.5 mM to 0.4 mM and 0.3 mM, with a post-induction temperature of 20℃, but still no expression of the CYP153A6/PFOR(CYP116B3) fusion was observed in the soluble fraction and no P450 forms were detected with whole cells or with the soluble fractions. The expression of CYP116B3, the CYP116B3 PFOR domain and CYP153A6 remained unchanged.

Modifying the post-induction temperature and decreasing the inducer concentration did not improve expression and further investigation of the expression conditions is required. One of the possibilities to investigate is the use of other methods for inducing expression, for example, the use of auto induction media (Grabski *et al.*, 2005) as opposed to IPTG induction, or perhaps the use of molecular chaperones to assist in the folding of the protein (Ventura & Villaverde, 2006).

The effect of rare codons and the protein expression rate on protein folding is still unclear

As previously mentioned, codon usage is one of the factors which may contribute to the formation of inclusion bodies in bacterial hosts (Ventura & Villaverde, 2006). Many genes contain rare codons (i.e. codons requiring tRNAs (transfer RNAs) that are rare or lacking in bacterial hosts) and this may hamper the expression of these genes in bacteria. The *CYP153A6* gene contains nineteen rare codons, but CYP153A6 could be expressed in *E. coli* BL21(DE3)pLysE in a P450 form. However, there was a possibility that improved expression and folding of CYP153A6 and the CY153A6/PFOR(CYP116B3) fusions could be obtained if *E. coli* Rosetta-gami 2(DE3)pLysS was used for expression. This strain is usually used for expressing genes containing rare codons and carries the pRARE plasmid encoding tRNAs for seven rare codons (Novagen, 2004).

Another factor which may increase the formation of inclusion bodies is the rapid production of proteins as a result of a strong promoter. The T7 promoter of the pET plasmid is a strong promoter and the expression of genes using the pET28 system, in combination with *E. coli* BL21(DE3)pLysE, results in the rapid production of proteins. This may result in molecular crowding within the cell hampering the folding of synthesised proteins. *E. coli* Rosetta-gami 2(DE3)pLysS grows slower than the BL21(DE3)pLysE strain and expression occurs at a slower rate. To determine whether these properties could perhaps result in the proteins having enough time to fold correctly and be expressed in the soluble fraction in a P450 form, the CYP153A6 operon and the CYP153A6/PFOR(CYP116B3) fusions were expressed using this strain. Expression of the CYP153A6/PFOR(CYP116B3) fusion and the modified

CYP153A6/PFOR(CYP116B3) fusion was only observed in the insoluble fraction and only P420 forms of these proteins were detected, at lower concentrations than those detected using the BL21(DE3)pLysE strain. The detected reductase activities were also lower. No expression of the CYP153A6 operon by the Rosetta-gami strain was observed in either fraction and only P420 forms of the protein were detected. The reductase activity detected for the FdR-Fdx couple was half of that detected using the BL21(DE3)pLysE strain. Overall, expression with the Rosetta-gami strain was very poor and the desired results were not obtained. The effect of rare codons and of rapid protein expression and molecular crowding on the folding of the CYP153A6/PFOR(CYP116B3) fusion protein is still unclear.

***E. coli* BL21(DE3)pLysE expressing the CYP153A6 operon gives excellent whole-cell bioconversion of *n*-octane but strains expressing the CYP153A6/PFOR(CYP116B3) fusions did not give any whole-cell bioconversion of *n*-octane.**

Despite the investigation of a number of factors which could possibly have been responsible for the expression of the CYP153A6/PFOR(CYP116B3) fusion protein in the form of inclusion bodies and in the P420 form, no expression of this protein in the P450 form was observed. It therefore seemed quite likely that the fusion was unsuccessful and that the misfolded protein was non-functional. To confirm this, a whole-cell *n*-octane bioconversion experiment was conducted.

The bioconversion protocol was first optimised by carrying out octane bioconversions using the CYP153A6 operon. The highest concentration of 1-octanol that could be produced by approximately 5 µM CYP153A6 after a period of 24 hours using sodium phosphate buffer and 40-mL vials, in the presence of glucose, was 41.82 mM. This translates to approximately 8 mM of 1-octanol produced per 1 µM P450 in 24 hours, which was higher than the concentration of 5.76 mM of 1-octanol that was produced per 1 µM of AciA from *Acinetobacter* sp. OC4 after 24 hours (Fujii *et al.*, 2006). The high whole-cell activity of CYP153A6 towards octane is quite remarkable considering the fact that the expression levels of the ferredoxin reductase and the ferredoxin were too low to detect using SDS-PAGE and that the cytochrome c reductase activity was also very low. Fujii *et al.* (2006) showed that whole-cell octane bioconversion by *E. coli* BL21 expressing only the AciA was only 4% of that of a strain expressing the complete operon from *Acinetobacter* sp. OC4 which included the ferredoxin and the ferredoxin reductase. When only the ferredoxin was expressed with AciA, the whole-cell octane bioconversion was 35% of that of the strain expressing the complete operon, indicating that the expression of the ferredoxin was critical for both the bioconversion of octane and for the stability of the P450, since very low

P450 levels of 0.8 μM were measured when the ferredoxin was not cloned as opposed to the 3 μM detected with the cloned ferredoxin and ferredoxin reductase. It is, however, noteworthy that expression of the ferredoxin and ferredoxin reductase could also not be observed with SDS-PAGE. The question remains whether enhancement of the expression of the CYP153A6 ferredoxin reductase and ferredoxin will further improve production of the P450 and of 1-octanol.

The bioconversion experiments done with the CYP153A6 operon were not very reproducible as there was some variability in the results obtained for experiments that were done on different occasions. However, one result that remained constant for all of the bioconversions was that the P450 concentration increased during the 24-hour reaction period, in some cases by up to 50%.

When a bioconversion experiment was conducted using the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion, P450 and P420 forms of both proteins were surprisingly detected in whole cells, despite previous results obtained with expression experiments when no P450 forms were detected with these fusions. These proteins also followed the observed trend, with the concentration of both the P450 and P420 forms increasing during the 24-hour reaction period. Despite the presence of P450 forms of both fusions no 1-octanol was produced. It should however be noted that with the original fusion, the P450 levels increased from 0.28 μM to 1.64 μM which was still significantly lower than the 5 μM detected in the bioconversion experiments carried out with CYP153A6. The results obtained during the octane bioconversion experiment were also not reproducible as P450 forms of the CYP153A6/PFOR(CYP116B3) fusion proteins were not detected when the experiment was repeated.

Both the lower levels of P450 produced with the CYP153A6/PFOR(CYP116B3) fusions as well as the lack of bioconversion activity of the fusions correlate with results reported by Fujita *et al.* (2009). They constructed fusions between the heme domains of AciA from *Acinetobacter* sp. OC4 and $P450_{balk}$ with the PFOR domain of P450RhF and compared the amounts of active P450 for these fusions with the corresponding three protein systems using the putidaredoxin and putidaredoxin reductase from *Pseudomonas putida*. The amount of properly folded P450 for both AciA and $P450_{balk}$ was higher for the three protein systems than for the fusions. No whole cell bioconversion activity was observed with the AciA/PFOR(P450RhF) fusion while the $P450_{balk}$/PFOR(P450RhF) fusion showed improved activity.

The increase in P450 levels during the first 24 hours of the bioconversion experiments led to the question of whether the cells were still viable after this period in the presence of 1.5 M octane.

Plating out of the bioconversion reaction mixtures confirmed that the cells were still viable despite this high octane concentration. It is generally accepted that the toxicity of a solvent is inversely correlated with its log P_{ow} value, which is defined as the partition coefficient of the solvent between *n*-octanol and water layers (Aono *et al.*, 2001). Therefore, the higher the log P_{ow} value, the less toxic the solvent should be to cells. Some *E. coli* strains have exhibited resistance to certain organic solvents: *E. coli* OST3410 and *E. coli* W3110 are able to tolerate cyclohexane (log P_{ow} 3.35), *E. coli* JA300 is *n*-hexane-resistant (log P_{ow} 3.87) and *E. coli* JA300T is decane-resistant (log P_{ow} 5.98) (Noguchi *et al.*, 1997; Aono *et al.*, 2001). Cyclohexane and hexane both have lower log P_{ow} values than *n*-octane (log P_{ow} 4.93) and should therefore be more toxic to cells than octane, making it possible that *E. coli* BL21(DE3)pLysE has some resistance to octane. Furthermore, Noguchi *et al.* (1997) reported that a high level of oxygen is required to maintain the viability of cells in the presence of organic solvents. The first octane bioconversion experiments using CYP153A6 were carried out in vials with volumes of 12 mL, 24 mL and 40 mL. The reactions in the 40-mL vials resulted in higher whole-cell bioconversion activities than with the 12-mL and 24-mL vials. One possibility is that the oxygen present in the 40-mL vial is sufficient to maintain the viability of the cells in the presence of octane for 24 hours. The cells were plated after a reaction period of 24 hours, but no cells were plated after a reaction period of 48 hours. It may be that the oxygen level after 24 hours is insufficient to maintain the viability of the cells, which could explain the levelling off in production of 1-octanol by CYP153A6 observed after 24 hours.

During all of the early expression experiments that were conducted in this study, no P450 forms of the CYP153A6/PFOR(CYP116B3) fusion were ever detected, but P450 forms were detected during one bioconversion experiment conducted at the end of this study. The only difference between expression and the bioconversions was the sodium phosphate buffer containing $FeSO_4$ and the addition of *n*-octane. From experiments that were performed to optimise the bioconversion protocol, it was concluded that the presence or absence of $FeSO_4$ did not affect the performance of CYP153A6 and therefore probably would not have a role in the folding of the proteins.

We have developed a number of theories with respect to the effect of the sodium phosphate buffer, the octane and the expressed proteins. The first theory is that the cells were still producing the desired proteins during the bioconversion experiments, made possible by the fact that the cells were still viable. This could explain the increase in P450 concentration observed during the 24-hour reaction period with both the potassium phosphate and the sodium phosphate buffers. The second theory is that the sodium phosphate buffer plays a role in the

folding of the proteins as the P450 concentration during the bioconversion reactions showed a greater increase in the presence of the sodium phosphate buffer than in the presence of the potassium phosphate buffer. The third theory is that the presence of octane facilitates the correct folding of the proteins. It is possible that during the bioconversion with the CYP153A6/PFOR(CYP116B3) fusion and the modified CYP153A6/PFOR(CYP116B3) fusion, as the protein was being produced, the presence of octane allowed the heme domain to assume the correct conformation for accepting the substrate, but that no hydroxylation could occur because the CYP153A6 and the CYP116B3 PFOR domain could not complement each other to form a functional unit. Another possibility is that the inclusion bodies, which could not be detected using CO-difference spectra, were refolded during the bioconversion reaction, forming P450s and P420s which could then be detected using CO-difference spectra.

A possible, but not very likely, explanation as to why no octane hydroxylation occurred is that there is a limiting factor, for example cofactor regeneration, which prevents the reaction from occurring. The FdR-Fdx couple of the CYP153A6 operon is NADH-dependent, whereas the PFOR domain is NADPH-dependent and it is possible that this factor negatively affects the electron transfer between the CYP116B3 PFOR domain and the CYP153A6 heme domain. However, good whole-cell bioconversions of octane were obtained with the $P450_{balk}$/PFOR(P450RhF) fusion, and the PFOR domain of this fusion is also NADPH-dependent (Nodate *et al.*, 2006; Fujita *et al.*, 2009).

There have been reports of only two fusions constructed between the heme domain of a wild-type CYP153A protein and the PFOR domain of a CYP116B protein: $P450_{balk}$/PFOR(P450RhF) and AciA/PFOR(P450RhF) (Nodate *et al.*, 2006; Fujita *et al.*, 2009). The latter fusion was not successful even though a Soret band was produced at 450 nm. The AciA/PFOR(P450RhF) fusion produced a negligible amount of 1-octanol from octane, a substrate towards which the natural form of this enzyme shows very high whole-cell activity. The protein was also unable to hydroxylate other typical AciA substrates. A possible explanation that was proposed by Fujita *et al.* (2009) was that the linker region of P450RhF was inhibiting the generation of a conformation required for the function of the heme domain, and that the heme domain and the CYP116B3 PFOR domain could therefore not form a functional complex. This is a possibility as there are amino acid differences between $P450_{balk}$ and AciA, with the most pronounced difference occurring at the N-termini of these proteins. AciA has a longer N-terminal region than $P450_{balk}$, which may result in slight structural changes (Fig. 5.1). $P450_{balk}$ and AciA share a much higher amino acid identity than $P450_{balk}$ and CYP153A6 (65% versus 51%). The most pronounced difference between $P450_{balk}$ and CYP153A6 is also at the N-termini of these proteins, but in this

case the N-terminal region of CYP153A6 is a lot shorter than both the $P450_{balk}$ and AciA N-terminal regions (Fig. 5.2). This may perhaps have implications for the interaction between the CYP153A6 heme domain and the CYP116B3 PFOR domain, which could possibly explain why the CYP153A6/PFOR(CYP116B3) fusion was unable to produce any 1-octanol, despite P450 forms of the protein being detected.

```
P450balk   1              MSTSSSTSNDIQAKIINAT----------------SKVVPMHLQIKALKNLM 36
AciA       1   MNSVAEIFEKITQTVTSTAADVATTVTDKVKSNEQFQTGKQFLHGQVTRFVPLHTQVRGIQWMQ 64

P450balk   37  KVKRKTIGTSRPQVHFVETDLPDVNDLAIEDIDTSNPFLYRQGKANAYFKRLRDEAPVHYQKNS 100
AciA       65  KAKFRVFNVQ-EFPAFIEQPIPEVATLALAEIDVSNPFLYKQKKWQSYFKRLRDEAPVHYQANS 127

P450balk   101 AFGPFWSVTRYEDIVFVDKSHDLFSAEPQIILGDPPEGLSVEMFIAMDPPKHDVQRRAVQGVVA 164
AciA       128 PFGAFWSVTRYDDIVYVDKNHEIFSAEPVIAIGNTPPGLGAEMFIAMDPPKHDVQRQAVQDVVA 191

P450balk   165 PKNLKEMEGLIRKRTGDVLDSLPLDTPFNWVPVVSKELTGRMLASLLDFPYDEREKLVGWSDRL 228
AciA       192 PKNLKELEGLIRLRVQEVLDQLPTDQPFDWVQNVSIELTARMLATLFDFPYEKRHKLVEWSDLM 255

P450balk   229 SGASSATGGEFTNEDVFFDDAADMAWAFSKLWRDKEARQKAGEEPGFDLISMLQSNEDTKDLIN 292
AciA       256 AGTAEATGGTVTNLDEIFDAAVDAAKHFAELWHRKAAQKSAGAEMGYDLISLMQSNEATKDLIY 319

P450balk   293 RPLEFIGNLALLIVGGNDTTRNSMSGGVLALNQFPEQFEKLKANPKLIPNWSLKYS-LATPLAY 355
AciA       320 RPMEFMGNLVLLIVGGNDTTRNSMTGGVYALNLFPNEFVKLKNNPSLIPNMVSEIIRWQTPLAY 383

P450balk   356 MRRVAKQDVELNGQTIKKGDRVLMWYASGNQDERKFENPEQFIIDRKDTRNHVSFGYGVHRCMG 419
AciA       384 MRRIAKQDVELNGQTIKKGDKVVMWYVSGNRDERVIERPDELIIDRKGARNHLSFGFGVHRCMG 447

P450balk   420 NRLAELQLRILWEELLPRFENIEVIGEPERVQSNFVRGYSKMMVKLTAKK 469
AciA       448 NRLAEMQLRILWEELLQRFENIEVLGEPEIVQSNFVRGYAKMMVKLTAKA 497
```

Figure 5.1 Amino acid alignment of the $P450_{balk}$ and AciA heme domains. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

$P450_{balk}$, AciA and CYP153A6 all belong to the CYP153A subfamily and share more than 50% amino acid identity. $P450_{cam}$ and $P450_{bzo}$, on the other hand, are from completely different families, sharing less than 40% amino acid identity with each other and the CYP153A P450s, but the fusion of these heme domains to the PFOR domain of P450RhF resulted in functional proteins. When it comes to fusions there are clearly no set rules and it is impossible to predict which P450s will result in functional proteins when their heme domains are fused to the PFOR-type reductase domain of a CYP116B enzyme.

```
P450balk  1   MSTSSSTSNDIQAKIINATSKVVPMHLQIKALKNLMKVKRKTIGTSRPQVHFVETDLPDVN-DL  63
CYP153A6  1                                                    HMEMTVAASDATNAAYGM  18

P450balk  64  AIEDIDTSNPFLYRQGKANAYFKRLRDEAPVHYQKNSAFGPFWSVTRYEDIVFVDKSHDLFSAE  127
CYP153A6  19  ALEDIDVSNPVLFRDNTWHPYFKRLREEDPVHYCKSSMFGPYWSVTKYRDIMAVETNPKVFSSE  82

P450balk  128 P---QIILGDPPEGLSVEMFIAMDPPKHDVQRRAVQGVVAPKNLKEMEGLIRKRTGDVLDSLPL  188
CYP153A6  83  AKSGGITIMDDNAAASLPMFIAMDPPKHDVQRKTVSPIVAPENLATMESVIRQRTADLLDGLPI  146

P450balk  189 DTPFNWVPVVSKELTGRMLASLLDFPYDEREKLVGWSDRLSGASSATGGEFTNEDVFFDDAADM  252
CYP153A6  147 NEEFDWVHRVSIELTTKMLATLFDFPWDDRAKLTRWSDVTTALP--GGGIIDSEEQRMAELMEC  208

P450balk  253 AWAFSKLWRDKEARQKAGEEPGFDLISMLQSNEDTKDLINRPLEFIGNLALLIVGGNDTTRNSM  316
CYP153A6  209 ATYFTELWN-----QRVNAEPKNDLISMMAHSESTRHMA--PEEYLGNIVLLIVGGNDTTRNSM  265

P450balk  317 SGGVLALNQFPEQFEKLKANPKLIPNWSLKYS-LATPLAYMRRVAKQDVELNGQTIKKGDRVLM  379
CYP153A6  266 TGGVLALNEFPDEYRKLSANPALISSMVSEIIRWQTPLSHMRRTALEDIEFGGKHIRQGDKVVM  329

P450balk  380 WYASGNQDERKFENPEQFIIDRKDTRNHVSFGYGVHRCMGNRLAELQLRILWEELLPRFEN---  440
CYP153A6  330 WYVSGNRDPEAIDNPDTFIIDRAKPRQHLSFGFGIHRCVGNRPAELQLNILWEEILKRWPDPLQ  393

P450balk  441 IEVIGEPERVQSNFVRGYSKMMVKLTAKK  469
CYP153A6  394 IQVLQEPTRVLSPFVKGYESLPVRINA    420
```

Figure 5.2 Amino acid alignment of the $P450_{balk}$ and CYP153A6 heme domains. Identical amino acids are highlighted in pink and similar amino acids are highlighted in green.

5.2 Aim 2: Amplification of Internal *CYP153A* Gene Fragments from Environmental DNA and Cassette PCR to Construct CYP153A Chimeras

The second aim of this study was to amplify the internal fragments of *CYP153A* genes from DNA extracted from environmental samples, and to use these fragments for cassette PCR with the goal of eventually creating diverse self-sufficient CYP153A proteins.

The internal fragments of *CYP153A* genes were amplified from environmental DNA using a set of degenerate primers designed according to conserved regions near the N- and C-termini of proteins in the CYP153A subfamily (Kubota *et al.*, 2005; van Beilen *et al.*, 2005). This environmental DNA, which was extracted from enrichments of soil sampled at a diesel-contaminated site in Aliwal-North in the Eastern Cape, yielded three different *CYP153A* sequences. One of these sequences, when translated, was 99% identical to the internal fragment of CYP153A6. This sequence was therefore excluded from the rest of the study. The plasmids containing the remaining two sequences were referred to as Aliwal03 and Aliwal09. In a different study, *CYP153A* gene fragments were amplified using the same primers from a metagenomic library constructed from samples obtained from the Beatrix Goldmine in the Free

State (Abbai, 2009). Two of these fragments were selected for use in this study and the plasmids containing these inserts were referred to as Beatrix14 and Beatrix16.

Cassette PCR was applied to the inserts of Aliwal03, Aliwal09, Beatrix14 and Beatrix16. These inserts were reamplified using the same set of degenerate primers. Two different sets of primers were then used to amplify the 5'- and 3'-ends of the gene encoding the heme domain of CYP153A6 from *Mycobacterium* sp. HXN-1500. The resulting amplicons had complementary ends which allowed them to be linked by overlap extension PCR to form full-length chimeric genes with the structure (5'-arm)-(Central *CYP153A* Fragment)-(3'-arm) (Kubota *et al.*, 2005).

The initial plan was to use cassette PCR to create diverse CYP153A6/PFOR(CYP116B3) fusions by replacing the *CYP153A6* gene with these chimeric *CYP153A* genes. These genes would ultimately serve as the parent genes for directed evolution, with the long-term goal of generating diverse self-sufficient terminal alkane hydroxylases. However, the CYP153A6/PFOR(CYP116B3) fusion could not be expressed in a P450 form in the soluble fraction and cassette PCR was therefore applied to the cloned CYP153A6 operon to test the functionality of the chimeric proteins. Unfortunately, these chimeras were mainly expressed in the insoluble fraction with low levels of P420 forms being detected. No expression of the chimeras could be observed in the soluble fraction using SDS-PAGE. The results obtained with CO-difference spectra, however, did not correlate with these results as some P420 expression in the soluble fraction was detected. The reductase activities of the FdR-Fdx couple, where detectable, were half of that detected for the FdR-Fdx couple of CYP153A6. This result was unexpected as the chimeras are not fusion proteins and the heme domain is therefore not expected to influence the activity of the redox partner proteins. However, P450s are quite sensitive to manipulation and the results may be unpredictable.

In order to confirm that these chimeras were incapable of *n*-octane hydroxylation, an octane bioconversion experiment was conducted, but no 1-octanol was produced. In addition, the trend observed with respect to the enzyme content was opposite to that observed in the bioconversions with the CYP153A6 operon and the CYP153A6/PFOR(CYP116B3) fusions as the P420 concentration decreased during the 24-hour reaction period. No P450s were detected. These results confirmed that the chimeras were non-functional.

The reason for these chimeras being non-functional can be understood when this study is compared to the work done by Kubota *et al.* (2005), where they used cassette PCR to isolate 16 new *CYP153A* gene fragments from petroleum-contaminated soil, petroleum-contaminated

groundwater and coastal seawater. $P450_{balk}$ (CYP153A13a) from *Alcanivorax borkumensis* SK2 served as the template for the amplification of the 5'- and 3'-arms. The resulting 16 chimeric *CYP153A* genes were expressed in *E. coli*. Three genes resulted in correctly folded P450s, exhibiting Soret bands at 450 nm. Five genes resulted in both P450 and P420 forms, indicated by Soret bands at both 450 nm and 420 nm. The eight remaining genes resulted in misfolded proteins, exhibiting peaks at 420 nm. Kubota and co-workers constructed a phylogenetic tree of the amino acid sequences of the CYP153A internal fragments and subsequently grouped these sequences into three clusters referred to as A, B and C. A similar phylogenetic tree is shown in Figure 5.3. The sequences that resulted in functional chimeric CYP153A proteins are indicated in Figure 5.3 by green boxes, whereas those that resulted in non-functional proteins are indicated by red boxes. The position of $P450_{balk}$ is indicated by a green arrow. Kubota and co-workers concluded that most of the sequences closely related to $P450_{balk}$ (i.e. those in clusters A and B) would result in functional chimeras, whereas sequences that are not that closely related to $P450_{balk}$ (i.e. those in cluster C) would result in non-functional chimeras.

If this reasoning is applied to the Aliwal03, Aliwal09, Beatrix14 and Beatrix16 sequences the results obtained with the chimeras are not unexpected. In Figure 5.3, the positions of the Aliwal sequences are indicated in purple and the positions of the Beatrix sequences are indicated in blue. The position of CYP153A6 in cluster C is indicated in black. The Aliwal sequences, which are part of cluster B, are not closely related to CYP153A6 and therefore will not result in functional chimeras when the gene fragments are linked to the 5'- and 3'-ends of the *CYP153A6* gene. Four of the sequences in cluster B resulted in functional chimeras with $P450_{balk}$, so there is a chance that chimeric genes constructed using the Aliwal inserts and the 5'- and 3'-ends of the $P450_{balk}$ gene would result in functional chimeras, a possibility which could be investigated in the future.

The Beatrix sequences, although in cluster C, are not that closely related to CYP153A6 and therefore result in non-functional chimeras. If the 5'- and 3'-arm fragments originate from a gene more closely related to the Beatrix inserts, there is a possibility that functional chimeras will result. However, as indicated by the red boxes in cluster B (Figure 5.3), not all chimeras generated by cassette PCR will be functional, even if the sequences are closely related. To date, there have been no reports in literature on the construction of chimeric CYP153As *via* cassette PCR, using sequences from cluster C.

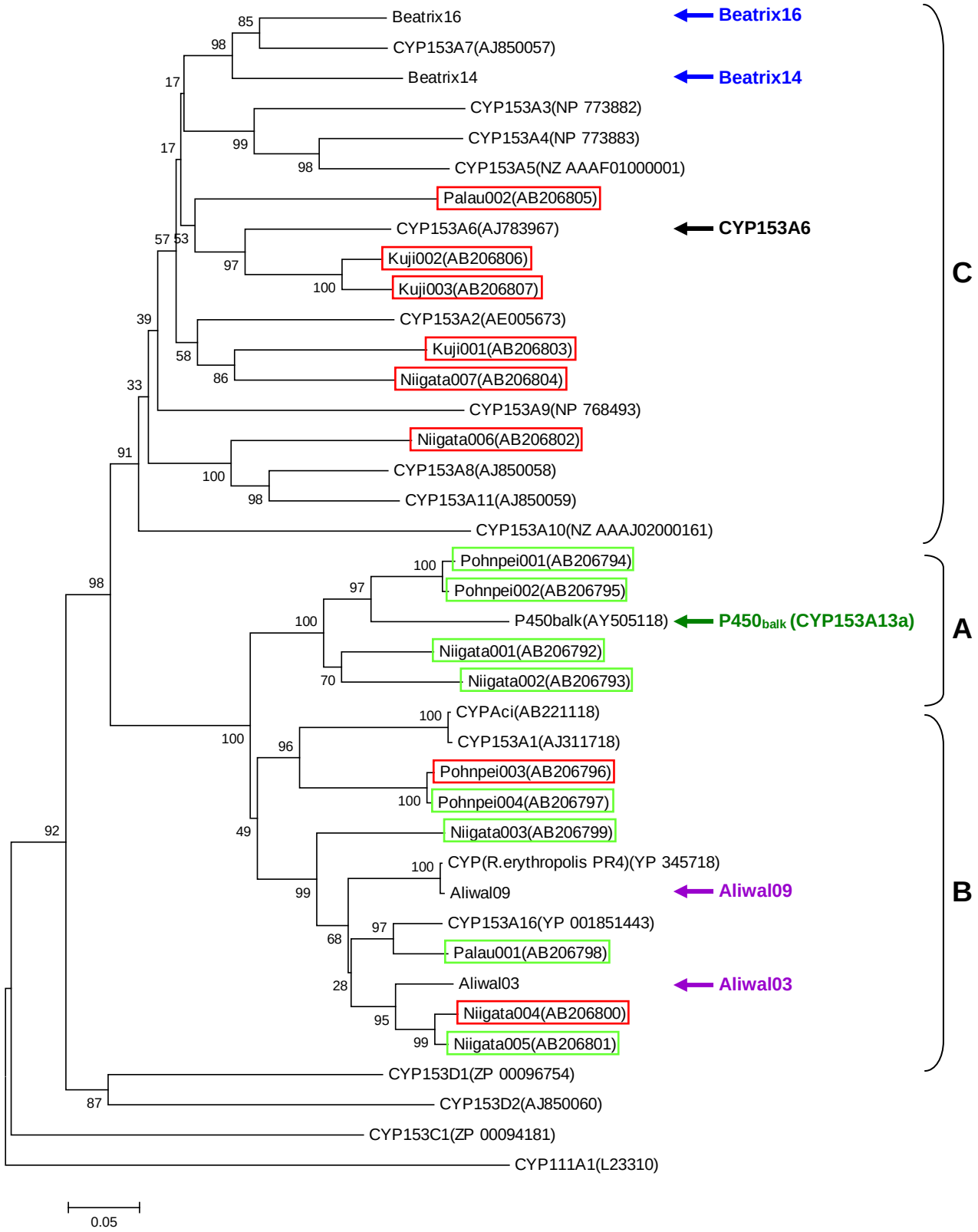


Figure 5.3 Phylogenetic analysis of the internal CYP153A fragments. A phylogenetic tree of the amino acid sequences of the internal CYP153A fragments was constructed using MEGA version 4. Neighbour-joining was performed using CYP111A1 as the outgroup. Bootstrap values are shown at the branch nodes. A, B and C represent the three clusters into which the internal CYP153A sequences were grouped by Kubota *et al.* (2005).

The *CYP153A* gene fragments that were isolated, although not resulting in functional CYP153A chimeras, can serve as the parent sequences for family shuffling if used in conjunction with functional sequences. As a result of the selection step in the family shuffling process, only sequences containing beneficial mutations will go through to the next round, whereas mutations that have a negative effect on the protein are eliminated (Zhao & Zha, 2004).

5.3 Future Research

This study provides a number of possibilities for future research. These are listed below.

- Further investigation into the effect of the expression and bioconversion conditions on the expression of the CYP153A6/PFOR(CYP116B3) fusions
- Addition of *n*-octane at an earlier stage to the cells expressing the CYP153A6/PFOR(CYP116B3) fusion, possibly shortly after induction, to investigate the effect of octane on the folding of the fusion protein
- Investigate the compatibility of different CYP153A heme domains and different CYP116B PFOR domains *via* the construction of fusions by swapping the domains of the CYP153A6/PFOR(CYP116B3) and $P450_{balk}$/PFOR(P450RhF) fusion proteins
- Expression of the CYP153A heme domain and CYP116B PFOR domains as separate proteins to investigate electron transfer between these domains in two component systems
- Construction of chimeric *CYP153A* genes using the inserts of Aliwal03 and Aliwal09 and the 5'- and 3'-ends of the gene encoding $P450_{balk}$, followed by expression and evaluation of the resulting chimeric gene products
- Construction of chimeric *CYP153A* genes using the inserts of Beatrix14 and Beatrix16 and the 5'- and 3'-ends of a gene closely related to these sequences, followed by expression and evaluation of the resulting chimeric gene products

Chapter 6

Conclusions

Cytochrome P450 monooxygenases are heme-containing enzymes that comprise one of the largest protein superfamilies known. These enzymes are able to catalyse diverse reactions, with one of the more interesting ones being the regioselective terminal hydroxylation of alkanes, the first step in alkane degradation. This is a very difficult reaction to catalyse with the use of synthetic chemical catalysts. Two P450 families are able to catalyse this reaction, one of them being the CYP153 family that is found in alkane-utilising bacteria. The industrial application of CYP153s is hampered by the fact that the electron transport system of these Class I P450s consists of three separate proteins, the heme domain and two redox partners. In order for these enzymes to function, the redox partners need to be cloned along with the heme domain, a situation which is not optimal for industrial purposes. A more optimal arrangement would be that of a self-sufficient P450, where the heme domain is fused to the redox partner(s). However, there are no naturally self-sufficient CYP153s. Chimeragenesis has, however, been used to construct a number of artificial self-sufficient P450s.

Nodate *et al*. (2006) successfully constructed an artificial self-sufficient CYP153 by fusing the heme domain of $P450_{balk}$ (CYP153A13a), a Class I terminal alkane hydroxylase from *Alcanivorax borkumensis* SK2, to the reductase (PFOR) domain of P450RhF (CYP116B2), a Class VII self-sufficient P450 from *Rhodococcus* sp. NCIMB. Kubota *et al.* (2005) used cassette PCR to replace the internal part of $P450_{balk}$ in this fusion with corresponding internal CYP153A fragments from environmental samples. The first goal of our study was to generate a fusion between the heme domain of CYP153A6 from *Mycobacterium* sp. HXN-1500, and the PFOR domain of CYP116B3 from *Rhodococcus ruber* DSM 44319, with the aim of generating a new Class VII self-sufficient terminal alkane hydroxylase. Cassette PCR would then also be applied to this fusion with the idea of generating more diverse self-sufficient terminal alkane hydroxylases, which could then serve as the starting point for directed evolution.

From experiments done with the constructed CYP153A6 chimeras and three controls, namely CYP116B3, the CYP116B3 PFOR domain and the CYP153A6 operon, we concluded that:

- **Not all fusions between Class I P450s and the reductase (PFOR) domains of CYP116B P450s result in properly folded, active self-sufficient P450s.** The CYP153A6/PFOR(CYP116B3) fusion that was constructed between the heme domain of CYP153A6 and the PFOR domain of CYP116B3 was cloned with and without an N-terminal His-tag in *E. coli* BL21(DE3)pLysE, using pET28 and pET22 vectors. However, these CYP153A6/PFOR(CYP116B3) fusions were mainly expressed in the form of inclusion bodies with only low levels of P420s being detected in the soluble fractions. Initially, no P450 forms of these proteins were detected, but one octane bioconversion experiment revealed the presence of both P450 and P420 forms of these fusions when CO-difference spectra were performed using whole cells. The concentration of P450s and P420s increased during the 24-hour bioconversion reaction period, but did not result in the bioconversion of *n*-octane. The linker region of the CYP116B3 PFOR domain was mutated to match the sequence of the linker region of the P450RhF PFOR domain. The resulting modified CYP153A6/PFOR(CYP116B3) fusion was cloned with an N-terminal His-tag, but this modified fusion gave similar results to the original fusion.

- **N-terminal His-tagged CYP153A6 can be expressed in *E. coli*.** The complete *CYP153A6* operon consisting of the genes encoding CYP153A6 with an N-terminal His-tag, ferredoxin reductase and ferredoxin was cloned in *E. coli* BL21(DE3)pLysE using the pET28b(+) vector. Expression levels of 11.03 pmol P450.mg protein^{-1} in the soluble fraction was achieved with this construct. Expression of a CYP153A with an N-terminal His-tag has not previously been reported. This His-tagged CYP153A6 catalysed the whole-cell bioconversion of *n*-octane to 1-octanol with an activity of up to 8 mM of 1-octanol produced per 1 µM P450. This represents one of the highest whole-cell activities obtained for *n*-alkane bioconversions.

- **The P450 content in whole cells increased during bioconversion experiments.** The CYP153A6 P450 concentration increased by as much as 50% during a 24-hour bioconversion reaction period. P450 forms of the CYP153A6/PFOR(CYP116B3) fusions were also detected in whole cells during one bioconversion experiment, with the P450 concentration also increasing during a 24-hour reaction period.

- **The N-terminal His-tagged CYP116B3 PFOR domain can be expressed without a heme domain and displays higher cytochrome c reductase activity than the complete CYP116B3.** Expression of the N-terminal His-tagged PFOR domain of P450RhF (CYP116B2) in *E. coli* was reported by Hunter *et al*. (2005), but the reduction of

cytochrome c by this protein was not investigated.

- **An N-terminal His-tag can be beneficial to the heterologous expression of some proteins while it can be detrimental to the expression of others.** When CYP153A6 and the CYP116B3 PFOR domain were cloned without N-terminal His-tags, expression of the PFOR domain was significantly lower, while no P450 and only very low levels of P420 could be detected in the case of CYP153A6. The expression of CYP116B3, on the other hand, was enhanced when the N-terminal His-tag was omitted.

- **Chimeras constructed using the N- and C-terminal ends of CYP153A6 did not yield correctly folded catalytically active P450s.** The internal fragments of *CYP153A* genes were amplified from environmental DNA using a set of degenerate primers. Four of these fragments were used to construct chimeric *CYP153A* genes *via* cassette PCR using the N- and C-terminal ends of the gene encoding CYP153A6. The chimeric *CYP153A* genes were expressed as part of the *CYP153A6* operon, with N-terminal His-tags. The chimeras were expressed in the form of inclusion bodies and were expressed as P420s. An octane bioconversion did not lead to the detection of P450s and the P420 concentration decreased during the 24-hour bioconversion reaction period. These chimeras did not catalyse the bioconversion of *n*-octane to 1-octanol. It should be noted that these internal CYP153A fragments shared between 55% and 71% amino acid identity with the corresponding internal fragment of CYP153A6, while the fragments that resulted in successful chimeras with $P450_{balk}$ shared between 71% and 88% amino acid identity with the internal fragment of $P450_{balk}$.

References

Abbai, N.S. (2009) Mining a South African deep mine metagenome for the discovery of novel biocatalysts. Ph.D. Thesis, University of the Free State, Bloemfontein, South Africa.

Abécassis, V.; Pompon, D. & Truan, G. (2000) High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28**: 20 e88.

Aono, R.; Tsukagoshi, N. & Miyamoto, T. (2001) Evaluation of the growth inhibition strength of hydrocarbon solvents against *Escherichia coli* and *Pseudomonas putida* grown in a two-liquid phase culture system consisting of a medium and organic solvent. *Extremophiles* **5**: 11-15.

Bernhardt, R. (2004) Optimized Chimeragenesis: Creating Diverse P450 Functions. *Chem. Biol.* **11**: 287-293.

Chefson, A. & Auclair, K. (2006) Progress towards the easier use of P450 enzymes. *Mol. BioSyst.* **2**: 462-469.

Correll, C.C.; Batie, C.J.; Ballou, D.P. & Ludwig, M.L. (1992) Phthalate Dioxygenase Reductase: A Modular Structure for Electron Transfer from Pyridine Nucleotides to [2Fe-2S]. *Science* **258**: 1604-1610.

Craft, D.L.; Madduri, K.M.; Eshoo, M. & Wilson, C.R. (2003) Identification and Characterisation of the Cyp52 Family of *Candida tropicalis* ATCC 20336, Important for the Conversion of Fatty Acids and Alkanes to α,ω-Dicarboxylic Acids. *Appl. Environ. Microbiol.* **69**: 5983-5991.

Crameri, A.; Raillard, S.-A.; Bermudez, E. & Stemmer, W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**: 288-291.

Danielson, P.B. (2002) The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. *Curr. Drug Metab.* **3**: 561-597.

Deeni, Y.Y.; Paine, M.J.I; Ayrton, A.D.; Clarke, S.E.; Chenery, R. & Wolf, C.R. (2001) Expression, Purification, and Biochemical Characterisation of a Human Cytochrome P450 CYP2D6-NADPH Cytochrome P450 Reductase Fusion Protein. *Arch. Biochem. Biophys.* **396**: 16-24.

De Mot, R. & Parret, A.H.A (2002) A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol.* **10**: 502-508.

Dilworth, F.J.; Black, S.M.; Guo, Y.-D.; Miller, W.L. & Jones, G. (1996) Construction of a P450c27 fusion enzyme: a useful tool for analysis of vitamin D_3 25-hydroxylase activity. *Biochem. J.* **320**: 267-271.

Dodhia, V.R.; Fantuzzi, A. & Gilardi, G. (2006) Engineering human cytochrome P450 enzymes into catalytically self-sufficient chimeras using molecular Lego. *J. Biol. Inorg. Chem.* **11**: 903-916.

Domanski, T.L. & Halpert, J.R. (2001) Analysis of Mammalian Cytochrome P450 Structure and Function by Site-Directed Mutagenesis. *Curr. Drug Metab.* **2**: 117-137.

Eiben, S.; Bartelmäs, H. & Urlacher, V.B. (2007) Construction of a thermostable cytochrome P450 chimera derived from self-sufficient mesophilic parents. *Appl. Microbiol. Biotechnol.* **75**: 1055-1061.

Fairbanks, G.; Steck, T.L. & Wallach, D.F.H. (1971) Electrophoresis analysis of the major polypeptides of human erythrocyte membrane. *Biochemistry* **10**: 2606-2617.

Farinas, E.T.; Bulter, T. & Arnold, F.H. (2001) Directed enzyme evolution. *Curr. Opin. Biotechnol.* **12**: 545-551.

Fisher, C.W.; Shet, M.S.; Caudle, D.L.; Martin-Wixtrom, C.A. & Estabrook, R.W. (1992) High-level expression in *Escherichia coli* of enzymatically active fusion proteins containing the domains of mammalian cytochromes P450 and NADPH-P450 reductase flavoprotein. *Proc. Natl. Acad. Sci. USA* **89**: 10817-10821.

Fujii, T.; Narikawa, T.; Sumisa, F.; Arisawa, A.; Takeda, K. & Kato, J. (2006) Production of α,ω-Alkanediols Using *Escherichia coli* Expressing a Cytochrome P450 from *Acinetobacter* sp. OC4. *Biosci. Biotechnol. Biochem.* **70**: 1379-1385.

Fujita, N.; Sumisa, F.; Shindo, K.; Kabumoto, H.; Arisawa, A.; Ikenaga, H. & Misawa, N. (2009) Comparison of Two Vectors for Functional Expression of a Bacterial Cytochrome P450 Gene in *Escherichia coli* Using *CYP153* Genes. *Biosci. Biotechnol. Biochem.* **73**: 1825-1830.

Funhoff, E.G.; Bauer, U.; García-Rubio, I.; Witholt, B. & van Beilen, J.B. (2006) CYP153A6, a Soluble P450 Oxygenase Catalysing Terminal-Alkane Hydroxylation. *J. Bacteriol.* **188**: 5220-5227.

Funhoff, E.G.; Salzmann, J.; Bauer, U.; Witholt, B. & van Beilen, J.B. (2007) Hydroxylation and epoxidation reactions catalysed by CYP153 enzymes. *Enzyme Microb. Technol.* **40**: 806-812.

Gilardi, G.; Meharenna, Y.T.; Tsotsou, G.E.; Sadeghi, S.J.; Fairhead, M. & Giannini, S. (2002) Molecular Lego: design of molecular assemblies of P450 enzymes for nanobiotechnology. *Biosensors Bioelectron.* **17**: 133-145.

Gillam, E.M.J. (2005) Exploring the Potential of Xenobiotic-Metabolising Enzymes as Biocatalysts: Evolving Designer Catalysts from Polyfunctional Cytochrome P450 Enzymes. *Clin. Exp. Pharmacol. Physiol.* **32**: 147-152.

Gillam, E.M.J. (2007) Extending the Capabilities of Nature's most versatile catalysts: Directed evolution of mammalian xenobiotic-metabolizing P450s. *Arch. Biochem. Biophys.* **464**: 176-186.

Gotoh, O. (1992) Substrate Recognition Sites in Cytochrome P450 Family 2 (CYP2) Proteins Inferred from Comparative Analyses of Amino Acid and Coding Nucleotide Sequences. *J. Biol. Chem.* **267**: 83-90.

Govindaraj, S. & Poulos, T.L. (1996) Probing the structure of the linker connecting the reductase and heme domains of cytochrome P450BM-3 using site-directed mutagenesis. *Protein Sci.* **5**: 1389-1393.

Grabski, A.; Mehler, M. & Drott, D. (2005) The Overnight Express Autoinduction System: High-density cell growth and protein expression while you sleep. *Nat. Methods* **2**: 233-235.

Graham, S.E. & Peterson, J.A. (1999) How similar are P450s and what can their differences teach us? *Arch. Biochem. Biophys.* **369**: 24-29.

Hannemann, F.; Bichet, A.; Ewen, K.M. & Bernhardt, R. (2007) Cytochrome P450 systems – biological variations of electron transport chains. *Biochim. Biophys. Acta* **1770**: 330-344.

Harikrishna, J.A.; Black, S.M.; Szklarz, G.D. & Miller, W.L. (1993) Construction and Function of Fusion Enzymes of the Human Cytochrome $P450_{scc}$ System. *DNA Cell Biol.* **12**: 371-379.

Harlow, G.R. & Halpert, J.R. (1996) Mutagenesis Study of Asp-290 in Cytochrome P450 2B11 Using a Fusion Protein with Rat NADPH-Cytochrome P450 Reductase. *Arch. Biochem. Biophys.* **326**: 85-92.

Hiraga, K. & Arnold, F.H. (2003) General Method for Sequence-Independent Site-Directed Chimeragenesis. *J. Mol. Biol.* **330**: 287-296.

Huang, W.; Johnston, W.A.; Hayes, M.A.; De Voss, J.J. & Gillam, E.M.J. (2007) A shuffled CYP2 library with a high degree of structural integrity and functional diversity. *Arch. Biochem. Biophys.* **467**: 193-205.

Hunter, D.J.B; Roberts, G.A.; Ost, T.W.B.; White, J.H.; Müller, S.; Turner, N.J.; Flitsch, S.L. & Chapman, S.K. (2005) Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Lett.* **579**: 2215-2220.

Inoue, H.; Nojima, H. & Okayama, H. (1990) High efficiency transformation of *E. coli* with plasmids. *Gene* **96**: 23-28.

Kagami, O.; Baik, S.-H. & Harayama, S. (2004) Effective DNA Shuffling Methods for Enzyme Evolution. In: Svendsen, A. (ed) *Enzyme Functionality: Design, Engineering and Screening*. Marcel Dekker, Inc., New York, pp 425-441.

Kikuchi, M.; Ohnishi, K. & Harayama, S. (1999) Novel family shuffling methods for the in vitro evolution of enzymes. *Gene* **236**: 159-167.

Klingenberg, M. (1958) Pigments of Rat Liver Microsomes. *Arch. Biochem. Biophys.* **75**: 376-386.

Kubota, M.; Nodate, M.; Yasumoto-Hirose, M.; Uchiyama, T.; Kagami, O.; Shizuri, Y. & Misawa, N. (2005) Isolation and Functional Analysis of Cytochrome P450 CYP153A Genes from Various Environments. *Biosci. Biotechnol. Biochem.* **69**: 2421-2430.

Lamb, D.C.; Lei, L.; Warrilow, A.G.; Lepesheva, G.I.; Mullins, J.G.; Waterman, M.R. & Kelly, S.L. (2009) The first virally encoded cytochrome P450. *J. Virol.* **83**: 8266-8269.

Landwehr, M.; Carbone, M.; Otey, C.R.; Li, Y. & Arnold, F.H. (2007) Diversification of Catalytic Function in a Synthetic Family of Chimeric Cytochrome P450s. *Chem. Biol.* **14**: 269-278.

Li, S.; Podust, L.M. & Sherman, D.H. (2007) Engineering and Analysis of a Self-Sufficient Biosynthetic cytochrome P450 PikC Fused to the RhFRED Reductase Domain. *J. Am. Chem. Soc.* **129**: 12940-12941.

Liu, L.; Schmid, R.D. & Urlacher, V.B. (2006) Cloning, expression and characterisation of a self-sufficient cytochrome P450 monooxygenase from *Rhodococcus ruber* DSM 44319. *Appl. Microbiol. Biotechnol.* **72**: 876-882.

Minshull, J. & Stemmer, W.P.C. (1999) Protein evolution by Molecular breeding. *Curr. Opin. Chem. Biol.* **3**: 284-290.

Munro, A.W.; Girvan, H.M. & McLean, K.J. (2007) Cytochrome P450-redox partner fusion enzymes. *Biochim. Biophys. Acta* **1770**: 345-359.

Nixon, A.E.; Ostermeier, M. & Benkovic, S.J. (1998) Hybrid enzymes: manipulating enzyme design. *Trends Biotechnol.* **16**: 258-264.

Nodate, M.; Kubota, M. & Misawa, N. (2006) Functional expression system for cytochrome P450 genes using the reductase domain of self-sufficient P450RhF from *Rhodococcus* sp. NCIMB 9784. *Appl. Microbiol. Biotechnol.* **71**: 455-462.

Noguchi, K.; Nakajima, H. & Aono, R. (1997) Effects of oxygen and nitrate on growth of *Escherichia coli* and *Pseudomonas aeruginosa* in the presence of organic solvents. *Extremophiles* **1**: 193-197.

Novagen (2004) Novagen Competent Cells User Protocol TB009 Rev. F0104, p.5.

Okuta, A.; Ohnishi, K. & Harayama, S. (1998) PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene* **212**: 221-228.

Omura, T. & Sato, R. (1964) The Carbon Monoxide-binding Pigment of Liver Microsomes. II. Solubilization, Purification, and Properties. *J. Biol. Chem.* **239**: 2379-2385.

Otey, C.R.; Landwehr, M.; Endelman, J.B.; Hiraga, K.; Bloom, J.D. & Arnold, F.H. (2006) Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS. Biol.* **4**: e112.

Reid, A.J. (2000) DNA Shuffling: Modifying the Hand that Nature Dealt. *In Vitro. Cell. Dev. Biol.-Plant* **36**: 331-337.

Roberts, G.A.; Grogan, G.; Greter, A.; Flitsch, S.L. & Turner, N.J. (2002) Identification of a new class of cytochrome P450 from a *Rhodococcus* sp. *J. Bacteriol.* **184**: 3898-3908.

Robin, A.; Roberts, G.A.; Kisch, J.; Sabbadin, F.; Grogan, G.; Bruce, N.; Turner, N.J. & Flitsch, S.L. (2009) Engineering and improvement of the efficiency of a chimeric [P450cam-RhFRed reductase domain] enzyme. *Chem. Commun.* 2478-2480.

Rosic, N.N.; Huang, W.; Johnston, W.A.; De Voss, J.J. & Gillam, E.M.J. (2007) Extending the diversity of cytochrome P450 enzymes by DNA family shuffling. *Gene* **395**: 40-48.

Sambrook, J.; Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual (2nd Edition)*. Cold Spring Harbor Laboratory Press, New York, pp. A1-B.21.

Shet, M.S.; Fisher, C.W.; Holmans, P.L. & Estabrook, R.W. (1993) Human cytochrome P450 3A4: Enzymatic properties of a purified recombinant fusion protein containing NADPH-P450 reductase. *Proc. Natl. Acad. Sci. USA* **90**: 11748-11752.

Shimoji, M.; Yin, H.; Higgins, L. & Jones, J.P. (1998) Design of a Novel P450: A Functional Bacterial-Human Cytochrome P450 Chimera. *Biochemistry* **37**: 8848-8852.

Sibbesen, O.; De Voss, J.J. & de Montellano, P.R.O. (1996) Putidaredoxin Reductase-Putidaredoxin-Cytochrome P450cam Triple Fusion Protein: Construction of a Self-sufficient *Escherichia coli* Catalytic System. *J. Biol. Chem.* **271**: 22462-22469.

Sieber, V.; Martinez, C.A. & Arnold, F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**: 456-460.

Smits, T.H.M; Seeger, M.A.; Witholt, B. & van Beilen, J.B. (2001) New Alkane-Responsive Expression Vectors for *Escherichia coli* and *Pseudomonas*. *Plasmid* **46**: 16-24.

Sono, M.; Roach, M.P.; Coulter, E.D. & Dawson, J.H. (1996) Heme-Containing Oxygenases. *Chem. Rev.* **96**: 2841-2888.

Stemmer, W.P.C. (1994a) DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**: 10747-10751.

Stemmer, W.P.C. (1994b) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**: 389-391.

Stemmer, W.P.C. (2002) Molecular breeding of genes, pathways and genomes by DNA shuffling. *J. Mol. Cat. B Enz.* **19-20**: 3-12.

Sukumaran, S.; Atkins, W.M. & Shanker, R. (2002) Engineering Cytochrome P-450s: Chimeric Enzymes. *Appl. Biochem. Biotechnol.* **102-103**: 291-302.

Van Beilen, J.B.; Lüscher, D.; Holtackers, R.; Witholt, B. & Duetz, W.A. (2005) Biocatalytic production of perillyl alcohol from limonene using a novel *Mycobacterium* sp. cytochrome P450 alkane hydroxylase expressed in *P. putida*. *Appl. Environ. Microbiol.* **71**: 1737-1744.

Van Beilen, J.B.; Funhoff, E.G.; van Loon, A.; Just, A.; Kaysser, L.; Bouza, M.; Holtackers, R.; Röthlisberger, M.; Li, Z. & Witholt, B. (2006) Cytochrome P450 Alkane Hydroxylases of the CYP153 Family Are Common in Alkane-Degrading Eubacteria Lacking Integral Membrane Alkane Hydroxylases. *Appl. Environ. Microbiol.* **72**: 59-65.

Ventura, S. & Villaverde, A. (2006) Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24**: 179-185.

Yuan, L.; Kurek, I.; English, J. & Keenan, R. (2005) Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* **69**: 373-392.

Zhao, H. & Zha, W. (2004) Evolutionary Methods for Protein Engineering. In: Svendsen, A. (ed) *Enzyme Functionality: Design, Engineering and Screening*. Marcel Dekker, Inc., New York, pp 353-373.

Websites

Cytochrome P450 Homepage (2009) http://drnelson.utmem.edu/

Fermentas (2009) http://www.fermentas.com/catalog/electrophoresis

Fermentas Double Digest (2008) http://www.fermentas.com/doubledigest/index.html

Frances H. Arnold Research Groups (2007) http://www.che.caltech.edu/groups/fha/.

NCBI BLAST (2008) http://www.ncbi.nlm.nih.gov/blast/BLAST.cgi

OligoAnalyser (2009) (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/)

Rare Codon Calculator (RaCC) (2009) (http://nihserver.mbi.ucla.edu/RACC/)

Summary

Cytochrome P450 monooxygenases are a superfamily of heme-containing enzymes that are found in all domains of life. P450s catalyse diverse reactions, many of which are difficult reactions to accomplish, even with the use of chemical catalysts. One such reaction is the terminal hydroxylation of alkanes, the first step in alkane degradation. The CYP153 family, found in alkane-utilising bacteria, is one of only two P450 families that can catalyse this reaction. One of the long-term goals of our group's research is the directed evolution of terminal alkane hydroxylases, using preferably a self-sufficient terminal alkane hydroxylase as the starting point. There are, however, no naturally self-sufficient CYP153s. Therefore, the first aim of this study was to create a self-sufficient CYP153 by fusing a CYP153 heme domain to the reductase (PFOR) domain of a self-sufficient P450.

The gene encoding the heme domain of CYP153A6 from *Mycobacterium* sp. HXN-1500 was ligated to the DNA encoding the PFOR reductase domain of CYP116B3 from *Rhodococcus ruber* DSM 44319. The fusion gene was expressed in *E. coli* using a pET28a plasmid. The resulting protein was misfolded and expressed mainly in the insoluble fraction in the form of inclusion bodies. Factors possibly responsible for this were investigated including the expression conditions, the effect of an N-terminal His-tag on protein folding, the effect of the linker region sequence on protein folding, and the possibility of rapid expression resulting in protein misfolding, but with all of these experiments only low levels of P420s were observed in the soluble fraction and no P450 forms were detected. A whole-cell octane bioconversion experiment conducted using the expressed fusion revealed the presence of P450 forms of the protein, but no 1-octanol was produced, indicating that octane possibly facilitated the correct folding of the CYP153A6 heme domain but that the heme domain and the PFOR reductase domain were unable to form a functional complex. This theory does, however, require further research.

In this study, CYP153A6 and its redox partners, ferredoxin reductase and ferredoxin were expressed in *E. coli* using the pET28b plasmid. Expression of CYP153A6 in *E. coli* using this plasmid has not previously been reported in literature. Whole-cell octane bioconversions conducted using the expressed CYP153A6 resulted in the production of 42 mM of 1-octanol

after 24 hours, with the P450 concentration increasing during this time, a trend which was also observed with the fusion.

The second aim of this study was to apply cassette PCR to the fusion to generate diverse self-sufficient terminal alkane hydroxylases, which would provide the genetic diversity required for directed evolution. Degenerate primers designed according to conserved N- and C-terminal regions of CYP153A amino acid sequences were used to amplify internal *CYP153A* gene fragments from environmental DNA extracted from enrichments of soil sampled at a diesel-contaminated site in the Eastern Cape. Three different sequences were identified, one of them being *CYP153A6*, which was excluded from the rest of the study. The two remaining sequences and two sequences originating from another project using environmental DNA from samples from the Beatrix Goldmine in the Free State were linked to the 5'- and 3'-ends of the *CYP153A6* gene, generating full-length chimeric *CYP153A* genes. Because of the fact that the expression of the fusion was unsuccessful, the functionality of these chimeric genes was tested using the above-mentioned functional CYP153A6 operon. Expression was observed in the insoluble fraction in the form of inclusion bodies, with the proteins being misfolded. A whole-cell octane bioconversion did not result in P450 forms of the proteins and no 1-octanol was produced, indicating that these chimeras were non-functional.

Keywords: P450, chimeras, CYP153A6, CYP116B3, self-sufficient, fusion, cassette PCR, terminal alkane hydroxylase, octane bioconversion, PFOR

Opsomming

Sitochroom P450 mono-oksigenases is heem-bevattende ensieme wat in alle domeine van lewe voorkom. Dié ensieme kataliseer diverse en dikwels moeilike reaksies. Een van hierdie moeilike reaksies is die terminale hidroksilasie van alkane, die eerste reaksie wat tydens die degradasie van alkane plaasvind. Die CYP153 familie, wat in bakteriëe wat alkane as koolstofbronne gebruik voorkom, is een van net twee P450 families wat hierdie reaksie kan kataliseer. Een van die langtermyn doelwitte van navorsing in ons groep is die gerigte evolusie van ‘n terminale alkaan hidroksilase met verkieslik ‘n selfonderhoudende terminale alkaan hidroksilase as beginpunt. Daar is geen natuurlike selfonderhoudende CYP153s, daarom was dit die eerste doelwit van hierdie studie om ‘n selfonderhoudende CYP153 te skep deur die heemdomein van ‘n CYP153 aan die PFOR reduktasedomein van ‘n selfonderhoudende P450 te koppel.

Die geen van die CYP153A6 heemdomein van *Mycobacterium* sp. HXN-1500 was aan die DNS van die CYP116B3 PFOR reduktasedomein van *Rhodococcus ruber* DSM 44319 geligeer. Die pET28a plasmied is gebruik om die *CYP153A6/PFOR(CYP116B3)* geen in *E. coli* uit te druk. Die proteïen was in ‘n misgevoude vorm in die onoplosbare fraksie uitgedruk. Faktore wat dalk hiervoor verantwoordelik kon wees was ondersoek, insluitent die uitdrukking kondisies, die effek van ‘n N-terminale Histidine-punt op die uitdrukking van die proteïen, die effek van die DNS wat die heem en PFOR domeine verbind op die uitdrukking van die proteïen en die moontlikheid dat vinnige uitdrukking veroorsaak dat die proteïen nie reg vou nie. Tydens hierdie eksperimente was die proteïen nie in die oplosbare fraksie uitgedruk nie en geen P450 vorms van die proteïen was waargeneem nie. Tydens ‘n oktaan biotransformasie eksperiment wat uitgevoer is met heel selle wat die proteïen uitgedruk het, is P450 vorms van die proteïen wel opgemerk, maar geen 1-oktanol is geproduseer nie. Dit dui aan dat oktaan moontlik nodig is vir die heemdomein om reg te vou, maar dat die heemdomein en PFOR reduktasedomein nie ‘n funksionele kompleks kon vorm nie. Verdere navorsing word egter benodig voor enige gevolgtrekkings gemaak kan word.

Die CYP153A6 operon is tydens hierdie studie vir die eerste keer met die pET28b plasmied in *E. coli* uitgedruk. Oktaan biotransformasies is met heel selle uitgevoer en 42 mM 1-oktanol is na 24

uur geproduseer. Die P450 konsentrasie het tydens hierdie periode toegeneem en hierdie neiging is ook met die CYP153A6/PFOR(CYP116B3) biotransformasie opgemerk.

Die tweede doelwit van hierdie studie was om kasset PKR te gebruik om meer diverse self-onderhoudende terminale alkaan hidroksilases te skep wat as die beginpunt van gerigte evolusie sou kon dien. Degenererende priemstukke is volgens gekonserveerde dele na aan die N- en C-terminale van CYP153A proteïene ontwerp en gebruik om die interne fragmente van *CYP153A* gene vanaf DNS uit grondmonsters van ’n dieselbesoedelde perseel in die Oos-Kaap te amplifiseer. Drie verskillende *CYP153A* fragmente is geïdentifiseer. Een van dié fragmente was *CYP153A6*. Die oorblywende twee fragmente en twee fragmente wat in ‘n ander projek verkry is vanaf DNS uit monsters verkry vanaf die Beatrix Goudmyn in die Vrystaat, is gekoppel aan die 5’- en 3’-dele van die *CYP153A6* geen om vollengte gene te maak. As gevolg van die onsuksesvolle uitdrukking van die *CYP153A6/PFOR(CYP116B3)* geen was dié *CYP153A* gene getoets deur gebruik te maak van die bogenoemde funksionele CYP153A6 operon. Die uitdrukkingsprodukte van dié gene was slegs in die onoplosbare fraksie waargeneem en die proteïene was misgevou. ‘n Oktaan biotransformasie eksperiment het geen P450s of 1-oktanol produksie gelewer nie, wat aangedui het dat dié proteïene onaktief was.